EEICT

STUDENT COMPETITION

BRNO UNIVERSITY OF TECHNOLOGY
FACULTY OF ELECTRICAL ENGINEERING AND COMMUNICATION

PROCEEDINGS II OF THE 27TH STUDENT EEICT 2021
ELECTRICAL ENGINEERING, INFORMATION SCIENCE, AND COMMUNICATION TECHNOLOGIES
SELECTED PAPERS

April 27, 2021
Brno, Czech Republic

BRNO FACULTY OF ELECTRICAL
UNIVERSITY ENGINEERING
OF TECHNOLOGY AND COMMUNICATION

EEICT
STUDENT COMPETITION

BRNO UNIVERSITY OF TECHNOLOGY
FACULTY OF ELECTRICAL ENGINEERING AND COMMUNICATION

EEICT
STUDENT COMPETITION

SPONSORS
Diamond tier

Honeywell

ON Semiconductor®

Thermo Fisher
SCIENTIFIC

SPONSORS

Silver tier

Honored tier

EEICT
STUDENT COMPETITION

EXPERT GUIDANCE

IEEE

EIA

MEDIA

FCC PUBLIC
ELEKTRO ON-LINE SVĚTLO

tzbinfo
www.tzb-info.cz

BRNO FACULTY OF ELECTRICAL
UNIVERSITY ENGINEERING
OF TECHNOLOGY AND COMMUNICATION

Contents

Foreword

OPENING WORD OF THE DEAN

These Proceedings contain papers presented during the **27th annual STUDENT EEICT conference**, held at the Faculty of Electrical Engineering and Communication, Brno University of Technology, on April 27, 2021. Despite the limitations due to COVID-19, the fruitful tradition of joining together creative students and seasoned science or research specialists and industry-based experts was not discontinued, providing again a valuable opportunity to exchange information and experience.

Supervised by the Electrical and Electronic Association of the Czech Republic, the EEICT involves multiple corporate partners, collaborators, and evaluators, whose intensive support is highly appreciated. Importantly, the competitive, motivating features of the conference are associated with a practical impact: In addition to encouraging students to further develop their knowledge, interests, and employability potential, the forum directly offers career opportunities through the affiliated PerFEKT JobFair, a yearly job-related workshop and exhibition complementing the actual EEICT sessions. In this context, the organizers acknowledge the long-term assistance from the Ministry of Education, Youth and Sports of the Czech Republic, which has proved essential for refining the scope and impact of the symposium.

In total, 188 peer-reviewed full papers distributed between 23 sessions were delivered interactively online, before examining boards with industry and academic specialists. The presenting authors exhibited a very high standard of knowledge and communication skills, and the best competitors received prize money and/or small gifts. These Proceedings comprise 60 award winning full papers, all selected by the conference's evaluation boards.

Our sincere thanks go to the sponsors, experts, students, and collaborators who participated in, contributed to, and made the conference a continued success.

Considering all the efforts and work invested, I hope that the **27th STUDENT EEICT (2021)** has been beneficial for all the participants.

I believe that the inspiration gathered during the event will contribute towards a further rise of open science and research, giving all the attendees a chance to freely discuss their achievements and views.

Prof. Vladimír Aubrecht
Dean of the Faculty of Electrical
Engineering and Communication

PREPARATION OF ELECTRODES OF LI-S BATTERIES USING INVERSELY VULCANIZED SULFUR

David Trochta

Bachelor Degree Programme (3), FEEC BUT

E-mail: xtroch00@stud.feec.vutbr.cz

Supervised by: Ondřej Čech

E-mail: cechondrej@feec.vutbr.cz

Abstract: This paper focuses on manufacturing of Li-S batteries, which in theory have higher energy density than the currently most popular Li-ion battery. Before putting the Li-S battery on the market, several problems which prevent its practical use have to be solved. These issues include short longevity, which could be solved by manufacturing the electrode using inversely vulcanized sulfur. This method could eliminate the immediate issue and enable the rise of the Li-S battery on the market.

Keywords: Li-S battery, inverse vulcanization, inversely vulcanized sulfur, electrochemistry

1 INTRODUCTION

We can notice a rising trend of mobile electronic assets in today's quickly developing world. Be it in phones or new models of electric cars, the demand for new types of batteries is becoming apparent, batteries which are able to satisfy and keep up with the growing demands of the current age.

The answer could be a battery composed of the combination of lithium and sulfur (Li-S). Due to its on-paper energy density, which goes up to 2600 Wh/kg, it is considerably higher than the currently most popular lithium-ion (Li-ion) batteries which only boast on-paper energy density of up to 600 Wh/kg. However, there are standing issues which cause low longevity of the Li-S battery still waiting to be solved [1].

Solution to this problem could come in the form of amorphous sulfur, which can be obtained using inverse vulcanization. The final product - inversely vulcanized sulfur (IVS) is more stable and and resilient to backwards recrystallization, which makes it a particularly fitting material for manufacturing of the Li-S battery [2].

2 Li-S BATTERY

The Li-S battery is a promising candidate for the next generation of electric energy storage systems. It has taken a lead position namely thanks to its high on-paper energy density combined with its low cost. In practice, values of up to 600 Wh/kg have been achieved so far. Its theoretically high energy density is rooted in the conjunction of lithium and sulfur. The cathode which is made of sulfur has on-paper capacity of 1675 Ah/kg. In the case of metal lithium, its on-paper capacity is exactly 3860 Ah/kg [1].

Despite the many advantageous properties of elemental sulfur, no-one has been able to create an optimal cell when it comes to its technical parameters (especially longevity) and thus compete with the currently dominant Li-ion batteries. Issues of the Li-S battery can generally be divided into three groups. First issue is the low electrical conductivity of elementar sulfur itself and its products. (i.e. Li_2S). Second is migration of higher polysulfides through the separator and its reactions with the lithium anode (the so-called sulfur shuttle effect). These polysulfides consequently reduce the effective area of the lithium anode and contribute to the decrease of conductivity. The last of the

serious issues is cracking and disintegration of the electrode due to thermal expansion. During the cyclical conversion of S_8 and Li_2S the volume of the electrode changes by 79%. Noteworthy is also the fact that carbon added into the cathodic slurry decreases the theoretical maximum energy density of the cell. For this reason, optimized sulfur/carbon ratio is needed to be achieved, as to keep the amount of sulfur as high as possible while preserving sufficient electrical conductivity.

2.1 PRINCIPE OF THE LI-S BATTERY

A traditional Li-S cell has the same internal structure as common batteries. The casing, which protects the internal components against mechanical damage and oxidation, contains the cathode (composite of sulfur and carbon), anode (metal lithium) and the separator, which separates both electrodes and prevents short-circuiting in between. The separator is filled with electrolyte - usually lithium salt dissolvable in organic dissolvents [2].

The internal process of charging and discharging itself is depicted in Image 1. It is apparent from the image that as soon as the electrons begin to flow through the outer circuit, lithium ions simultaneously start to break loose. These lithium ions then diffuse through the separator using the electrolyte and get stuck to the cathode. On the cathode, the reactions between sulfur, lithium ions and electrons are realized, consequently leading to conversions. During charging, the whole process is inverted - polysulfides convert back to sulfur and the lithium ions as well as the electrons return to the anode [2].
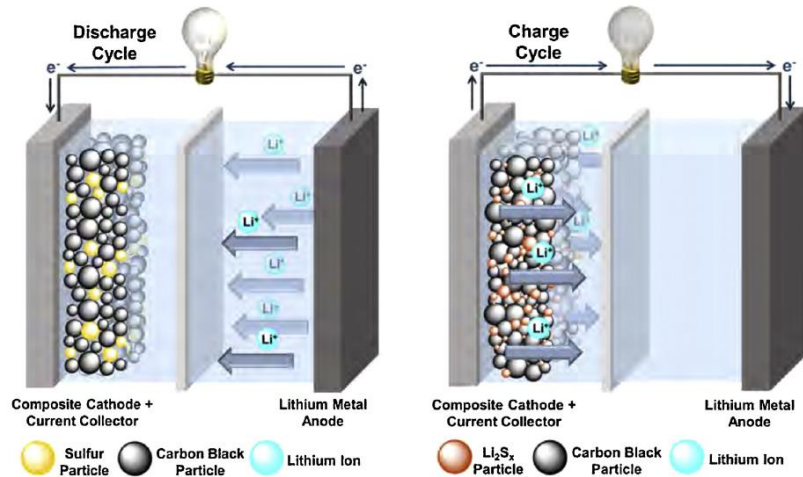
Image 1: Charging and discharging principle of the Li-S cell [2].

3 INVERSE VULCANIZATION

During traditional vulcanization, which is used to process natural rubber, sulfur creates bridges between sections of polymer chains. The situation during inverse vulcanization is similar, however the roles are different. Sulfur is the main material and a stable co-polymerous product is created using a crosslinker in this case, which is resistant to backwards depolymerization [2; 3].

Inversely vulcanized sulfur is also able to change its physical properties, including Young's modulus, tensile strength and refractive index (up to 1.8). Therefore the resulting material can additionally see value in application as a thermal insulator, infrared optic materials, or a cleaning material for water sources and natural environments.

3.1 PROPERTIES OF INVERSELY VULCANIZED SULFUR (IVS)

The final properties of IVS are affected particularly by the type and the amount of crosslinker used (sulfur and crosslinker weight percentage ratio). In general, the final form depends on the structure, as seen in Image 2. The first case contains only pure sulfur, which forms a linear polymer chain.

While in the case of the structure containing 90 wt. % of S and 10 wt. % of 1.3-diizopropenylbenzene (DIB) formation of macromolecular loops can be observed. Should more DIB be added, a so-called hyperbranched structure will be created. In the last scenario, a vinyl grouping is attached to the ends of the sulfur chains.
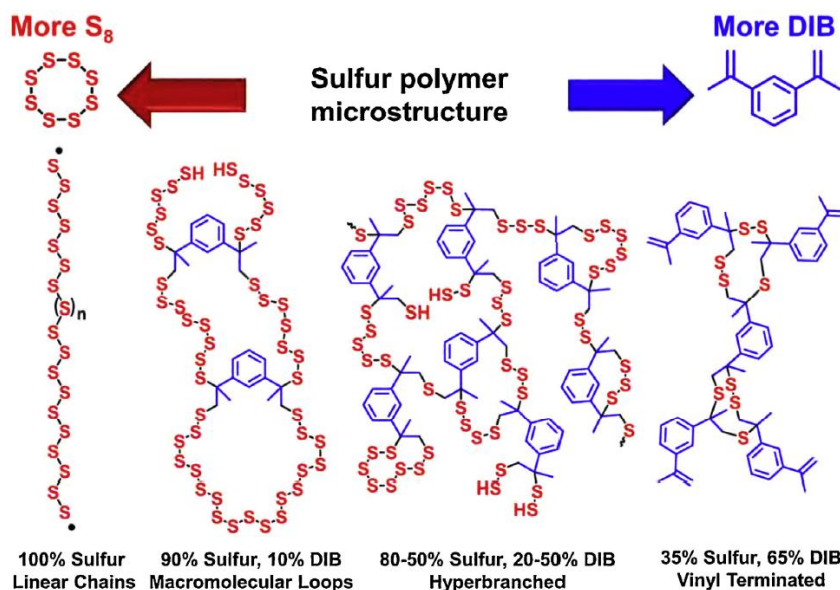
Image 2: Schematic of the IVS microstructure based on various S and DIB ratios [2].

3.2 MANUFACTURING OF INVERSELY VULCANIZED SULFUR

Manufacturing of IVS involves elemental sulfur and DIB crosslinker. The process begins by heating the elemental sulfur, which is placed in a vial along with a magnetic stirrer. The vial is then placed in an oil bath, which is heated until it reaches the temperature of 185°C. Before reaching the desired temperature however, several changes can be observed. The first change happens at the temperature of around 124°C when the elemental sulfur melts to become a yellow liquid. While rising the temperature, its color gradually changes from yellow, then light red until it becomes crimson red. As the temperature reached 185°C we added DIB into the vial and left the mixture to stir. The mixture immediately changed color back to yellow, but over the time it became crimson red again. The mixture was kept at 185°C for the following 24 hours.

Using this method we were able to prepare samples of IVS containing various ratios of input precursors. Specifically speaking S:DIB 70:30, 60:40 and 50:50 wt. %. We also attempted to create a S:DIB 80:20 wt. % sample, but this sample was crosslinked incorrectly - became plastic and sticky. IVS samples containing various ratios of input precursors do not vary in color. Subsequently we verified the internal structure using x-ray diffraction (XRD).

Image 3: IVS sample containing a ratio of S:DIB 70:30 wt. %.

4   ELECTRODE PREPARATION

To prepare the electrode slurry, it is necessary to grind down IVS into a powder. A vibration mill was used for this purpose with IVS being placed inside for 30 minutes to grind. Before the grinding itself the mill container along with IVS was put into a freezer to achieve higher brittleness of IVS and therefore higher grinding quality.

Afterwards, a mixture of IVS powder was prepared, Super P carbon and CMC (carboxymethylcellulose) binder in a ratio of 60:30:10 wt. %, with demineralized water being used as a dissolvent for CMC. The mixture was thoroughly mixed using a planetary mill by means of three 10 mm-diameter grinding balls, with a stirring speed of 350 revolutions per minute. The stirring was not continuous however, instead it included breaks during which the container cooled. Should these breaks be omitted from the mixing process, there's a risk of the mixture melting together as a result of high temperatures. The mixing duration without accounting for breaks was 8 hours. After the planetary mill procedure, a full week of additional mixing using a magnetic stirrer followed.

Upon processing the electrode slurry this way, it was coated onto an aluminum foil (current collector) using a coater and left to dry. First, the drying was done at room temperature and later proceeded in a dryer at the temperature of 50 °C. This step was necessary to prevent the cracking and peeling of the coated electrode slurry layer. The thickness of this layer was 100 μm. The drying was followed by cutting 18 mm diameter electrodes out of this coated foil and their consequent pressing using a force of 1600 kg/cm2.

5   CELL ASSEMBLY

The last step of the manufacturing process of the cell is its assembly within a glovebox containing an argon atmosphere. This argon atmosphere is necessary particularly due to the reactive nature of lithium. The assembly itself proceeds by forming a so-called electrochemical cell, which gets hermetically closed off in the box and thereafter it is able to be safely manipulated with and measured. The cell contains a lithium anode and the examined cathode. The electrodes are divided via a separator, which is filled with electrolyte. Then, the galvanostatic cycling with potential limitation (GCPL) and the electrical impedance spectroscopy (EIS) methods are used in order to test the cell.

6   CONCLUSION

In this project we presented a manufacturing process of inversely vulcanized sulfur which uses the 1.3-diizopropenylbenzene (DIB) crosslinker, carried out using various ratios of input precursors. Utilizing our method, we successfully synthesized IVS containing ratios of S:DIB 70:30, 60:40 a 50:50 wt. %. Subsequently, we described a process of electrode manufacturing optimized particularly for inversely vulcanized sulfur. Through this approach, we have proved that electrodes which use IVS as its active material can work and thus possess potential for further research.

REFERENCES

[1]   LIU, Tiefeng, Hualiang HU, Xufen DING et al. 12 years roadmap of the sulfur cathode for lithium sulfur batteries (2009–2020). *Energy Storage Materials* [online]. 2020, **30**, 346-366 [cit. 2021-03-12]. ISSN 24058297. Accessible from: doi:10.1016/j.ensm.2020.05.023

[2]   GRIEBEL, Jared, Richard GLASS, Kookheon CHAR a Jeffrey PYUN. Polymerizations with elemental sulfur: A novel route to high sulfur content polymers for sustainability, energy and defense. *Progress in Polymer Science* [online]. 2016, **58**, 90-125 [cit. 2021-03-12]. ISSN 00796700. Accessible from: doi:10.1016/j.progpolymsci.2016.04.003

[3]   CHUNG, Woo Jin, Jared J. GRIEBEL, Eui Tae KIM, et al. The use of elemental sulfur as an alternative feedstock for polymeric materials. *Nature Chemistry* [online]. 2013, **5**(6), 518-524 [cit. 2021-03-12]. ISSN 1755-4330. Accessible from: doi:10.1038/nchem.1624

BATTERY MANAGEMENT SYSTEM FOR FORMULA STUDENT

Maroš Macko

Bachelor Degree Programme (3rd year), FEEC BUT

E-mail: xmacko12@vutbr.cz

Supervised by: Pavel Šteffan

E-mail: steffan@feec.vutbr.cz

Abstract: This thesis is focused on the battery management system for Formula Student monopost. Part of the work is the design and implementation of this monitoring system's hardware and firmware. The system uses a distributed topology and is divided into a master board and slave boards. Master board design and the selection of suitable components are described. Furthermore, slave boards are designed, and the fundamentals of battery voltage and temperature measuring are explained. All systems must comply with the rules of the Formula Student.

Keywords: Battery Management System, BMS, AMS, Li-ion, Formula Student, TU Brno Racing

INTRODUCTION

A battery management system (BMS for short) is an essential part of single-cell batteries and an absolute must-have for multi-cell battery packs. This thesis is focused on the development of a monitoring and management system for electric formula racing car made by Formula Student university team TU Brno Racing. The system can monitor the voltages of each cell and their temperature. It communicates with other systems via the CAN bus. If all systems are functional and not in a fault state, it then switches battery output with relays, which can be turned off in an emergency. The management system also complies with the technical and safety requirements of the Formula Student competition, so it can pass technical inspection at the races.

1 BATTERY CONFIGURATION

Sony VTC6 18650 format Li-Ion cells were selected for the construction of the battery. They are a cylindrical battery type (cylinder with 18 mm diameter and 650 mm height) with a capacity of 3000 mAh. From those, we can build a battery with a maximum allowed voltage of 600 V, divided into independent segments in a way that none of them is over 12 kg or 6 MJ (1.6̄ kWh) of energy. For a simple build, acceptable capacity and voltage for the goals of the competitions, 96S8P (96 in series, 9 parallel) configuration was chosen, divided into 6 independent segments (each 16S8P). In this way, the battery reaches a maximum voltage of 403 V, the nominal voltage of 355 V and a capacity of 8.5 kWh.

2 TOPOLOGY

Formula student rules directly mandate segmented battery layout, so it does not make sense to think about a centralized battery management system. The distributed system has a main control unit (responsible for safety, output switching, charging and discharging rate control or communication with the main vehicle control unit) separated from monitoring parts, which are distributed in segments and are responsible for measuring the cell voltages and their temperature [1]. This distributed BMS system is divided into the main control unit (called BMS Master) and secondary units (called BMS

Slave(s)). Each segment has its own printed circuit board (PCB) for the measurements and charging circuitry controlled by the Master board.

The communication protocol between Slave and Master modules is CAN bus, which is also used in the rest of the car's systems, such as the VCU - Vehicle Control Unit, or the Inverter. For compatibility reasons, 2.0B standard was chosen, with speed configured to 1 Mbit/s. The battery has its own isolated CAN bus from the rest of the car, between BMS Master and Slave boards. This bus type has many advantages, such as automatic collision prediction and message prioritization. The Block diagram of the system is pictured in the figure 1.
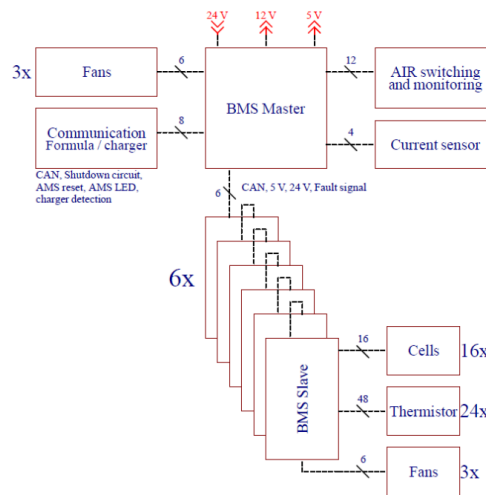
Figure 1: BMS block diagram

3 MASTER SUBSYSTEM DESCRIPTION

The BMS Master PCB communicates with all BMS Slave subsystems via CAN bus and also with the rest of the car, or when charging outside the car with the charger via the second CAN bus.

It is powered by the main converter that reduces the high traction voltage to 24 V DC for the entire formula car. Master then implements two more switching step-down converters, to 12 V DC and 5 V DC, and a linear regulator from 5 V to 3.3 V DC to power the logic circuits and the microcontroller.

The LM3150 switching integrated circuit was chosen, as it is very robustly built, implements over-temperature protection and adjustable over-current protection and handles very fast transients. Our design achieves maximum efficiency of 93% (5 V, 3 A) to 96% (12 V, 2 A). Auxiliary power supplies 5 and 12 V DC supply the BMS Slave as well as the auxiliary analog PCB. It measures high voltage, at the command of the BMS Master switches a relay (with series resistance) for pre-charging the capacitors in the inverter for motors, as well as main positive and negative relays with auxiliary contacts to determine the mechanical condition of the relay (e.g. to each other or whether they stuck properly). This information is returned to the Master PCB for processing. The BMS Master provides this module with a 12 and 5 V DC power supply.

According to the rules, the BMS system also switches the car's shutdown circuit. Switching must be implemented by a mechanical relay. This circuit runs through the whole car and in several places it has the possibility to be interrupted, for example, by emergency buttons on the side of the car or other systems. The circuit ends again in the battery and must be used as the supply voltage for the relay. The status of this safety system is monitored by the BMS Master so that the discharge of high-voltage capacity in the inverter in the event of the car being switched off can be correctly and safely implemented on the car side.

4 SLAVE SUBSYSTEM DESCRIPTION

The input voltage is 5 V DC, which must be galvanically isolated as the potential increases with each PCB Slave by the segment voltage (48 to 67 V). It is also necessary to galvanically isolate the CAN bus. The slave also controls 3 fans powered from a 24 V DC branch using PWM. This power supply is not isolated from the main DC/DC converter, so it is necessary to galvanically isolate the control PWM signal and transfer the switching on the common low voltage side.

The BMS Slave monitors voltages of all cells connected in series, as well as temperature sensors located in the segment. According to the rules, the temperatures of at least 30% of cells must be monitored evenly. One segment has 128 cells and by placing the sensor close to the cell welds, we can effectively measure 2 cells. A minimum of $128*0.3/2 = 19.2$ is required, so 20 sensors.

The advantage of using 1-Wire type digital temperature sensors would be the simplicity of the measurement, as we would not have to implement an A/D converter. However, all sensors would be in series (otherwise their meaning is lost) and thus the reliability of the system would be significantly reduced. 10 kΩ NTC thermistors, which are multiplexed in front of the A/D converter, were chosen as temperature sensors. To compensate for possible damage to the cabling or sensors, 24 NTC thermistors are used (more than necessary according to the rules).

NTC thermistors are powered by a pull-up resistor with the same value (10 kΩ) and thus form a voltage divider. If a different, stable reference voltage is used for the A/D measurement, it is necessary to measure the supply voltage, which can fluctuate in the long and short term. This ensures that we measure the precise voltage on the thermistor and its ratio to the supply voltage. Similar to measuring the voltage of individual cells, we can benefit from the properties of an external ΔΣ converter to filter the input voltage, as we are not interested in fast, but an accurate sampling of a slowly changing value.

For very precise measurements, the temperature can be approximated by the Steinhart-Hart equation (equation 1), with which we can approximate the real characteristic precise up to 0.02 °C [2]. Coefficients A B and C can be calculated from the dimensions and material parameters of the thermistor, or they can be determined by experimental measurements. However, manufacturers more often state in the technical sheet another, β (or B) constant, usable in the modified equation 3. To calculate the final temperature, we can simplify some expressions and store them as constants in the microcontroller or a look-up table (LUT) is often implemented (if there is enough free space in the memory) to directly convert a value from an A/D converter to temperature using simple binary operations.

$$\frac{1}{T} = A + B \cdot lnR + C \cdot (lnR)^3 \tag{1}$$

$$R(T) = R(T_0) * e^{\beta*(\frac{1}{T} - \frac{1}{T_0}))} \tag{2}$$

$$T = \frac{\beta}{ln\frac{R}{R(T)} + \frac{\beta}{T_0}} \tag{3}$$

By applying the equation using the parameter β and obtaining the values for the thermistor used by us (β = 3380, T_0 = 25 °C)[3] we can derive the dependence of the resistance (and thus also the measured voltage on the divider) at temperature and implement a lookup table.

During cell charging, it is necessary to balance the individual cells in series so that all cells with small differences in charging characteristics, contact resistance and different nominal capacity are charged to the same level and the others are not overcharged. Due to the requirements for the smallest possible weight and size of the printed circuit board, this BMS uses passive cell balancing.

For our 16S configuration, the LTC6813 integrated circuit (capable of up to 18S configuration) was selected for cell voltage monitoring and balancing control. It uses an A/D converter using Delta-Sigma

($\Delta\Sigma$) modulation, which is advantageous for our application compared to a successive approximation converter (SAR). The SAR converter is capable of more samples per second but also lacks any filtering. In contrast, the Delta-sigma converter does integration on the input voltage during sampling, thus providing a certain level of filtering.

CONCLUSION

The result of this thesis is the design and implementation of functional printed circuit boards and firmware for individual parts of the battery management system for the electric formula car Dragon e1. The architecture and individual components were selected after detailed market research and requirements. All parts of the design have been adapted to the needs and rules of the Formula Student competition. The system can monitor the voltages of individual cells and their temperature with sufficient accuracy. It switches the output from the battery with relays, which can be switched off in the event of an emergency and thus completely shuts down the output from the battery. It meets all the technical and safety requirements of the Formula Student competition and it can pass the technical inspection at the competitions. Firmware for the BMS Slave and BMS Master units has been implemented, along with self-testing and verification of the cells and battery segment balancing.
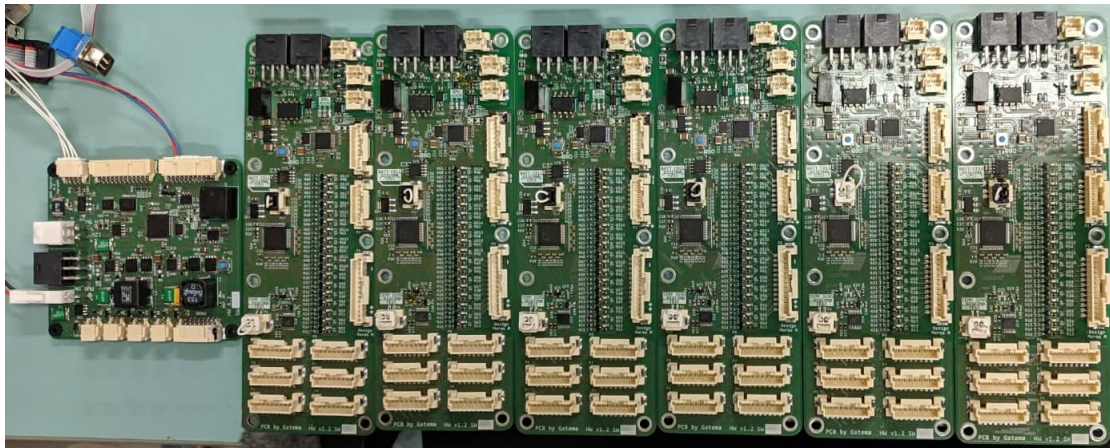
Figure 2: Completed PCBs of BMS Master and 6x BMS Slave

ACKNOWLEDGMENTS

REFERENCES

[1] Xiong, Rui.: *Battery Management Algorithm for Electric Vehicles*. Singapore: China Machine Press, 2020. ISBN 978-981-15-0247-7.

[2] Keysight Technologies: *Practical Temperature Measurements (Application Note)* [online]. USA, 2020. Available from URL: <https://www.keysight.com/zz/en/assets/7018-06789/application-notes/5965-7822.pdf>.

[3] Murata Manufacturing: *NTC Thermistors (datasheet)* [online]. 3/2020. Available from URL: <https://www.murata.com/media/webrenewal/support/library/catalog/products/thermistor/ntc/r44e.ashx>.

CELL AND SUB-CELLULAR SEGMENTATION IN QUANTITATIVE PHASE IMAGING USING U-NET

Jakub Majerčík, Michal Špaček

Bachelor Degree Programme (3), FEEC BUT

E-mail: xmajer19@stud.feec.vutbr.cz, xspace31@stud.feec.vutbr.cz

Supervised by: Tomáš Vičar, Jaromír Gumulec

E-mail: vicar@vutbr.cz, j.gumulec@med.muni.cz

Abstract: The ability to automatically segment images, especially microscopy images of cells, opens new opportunities in cancer research or other practical applications. Recent advancements in deep learning enabled for effective single-cell segmentation, however, automatic segmentation of sub-cellular regions is still challenging. This work describes an implementation of a U-net neural network for label-free segmentation of sub-cellular regions on images of adherent prostate cancer cells, specifically PC-3 and 22Rv1. Using the best performing approach, out of all that have been tested, we have managed to distinguish between objects and background with average dice coefficients of 0.83, 0.78 and 0.63 for whole cells, nuclei and nucleoli respectively.

Keywords: cell segmentation, deep learning, neural network, quantitative phase imaging, nucleus, nucleolus

1 INTRODUCTION

Automatic separation of objects from the background is often an inseparable part of computer vision-related problems and finds its use in many medical and clinical applications. Accurate segmentation of cells can speed up the whole research process of many diseases, such as cancer. By identifying different cellular compartments, we are able to analyze the biological processes taking place in these cells. In this work, we focus on the label-free segmentation of prostate cancer cells in quantitative phase refractive index microtomograms (PC-3 and 22Rv1) using machine learning methods. Our goal is to segment whole cells, nuclei, and nucleoli with reasonable accuracy. Correct segmentation of sub-cellular regions is a necessary step for automated sub-cellular cell analysis.

2 METHODS

2.1 CELL CULTIVATION

In this work, we have analysed 2 cell lines. The PC-3 human epithelial cell line was obtained from fourth-degree prostate adenocarcinoma and bone metastases. 22Rv1 is a human prostate cancer epithelial cell line obtained from a xenograft of propagated mice after castration. Both PC-3 and 22Rv1 cells were cultured at the Institute of Pathological Physiology, Faculty of Medicine, Masaryk University. PC-3 cells were cultured in Ham's F12 medium with 10% FBS, while RPMI-1640 medium with 10% FBS was used to culture 22Rv1 cells. Both media were enriched with antibiotics (penicillin 100μ/ml streptomycin 0.1mg/ml). The cells were maintained in an incubator at a constant temperature of 37°C, 60% humidity with 5% CO_2 (Sanyo, Japan). Passages of 22Rv1 cells ranged from 25 to 35, the passage of the PC-3 line from 15 to 25. Cells showing 50-60% confluence were washed with an FBS-free medium. Subsequently, a fresh medium containing FBS was added. Furthermore, variants of 22Rv1 and PC-3 cells adapted to high zinc concentrations were used in this study to increase data variability, previously prepared by Holubova et al [2].

2.2 DATA ACQUISITION AND ANNOTATION

The acquisition of microscopic images was performed using a Nanolive 3D Cell Explorer holoto-mographic microscope with a 60x/0.8 objective, which measures the quantitative refractive index. Thus, it enables the noninvasive acquisition of 3D data of almost any cell type [1]. The refractive index of both media was measured at 1.34. All four cell types (PC-3 Wild type, PC-3 Zinc resistant, 22Rv1 Wild type and 22Rv1 Zinc resistant) were cultured in separate ibidi μ-Slide Luer 0.8 chambers. Thanks to this, it was possible to easily create 4 unique data sets containing 200-220 3D images. Each image is composed of 96 slices and, upon acquisition, was identified by a label number in which the individual cell structures could be seen the best. In addition to the microscope, the acquisition also required software, specifically *Steve*, provided by Nanolive. The dimensions of the field of view were 94x94x35μm. The measurement resulted in more than 750 usable 3D images ($\sim$400 PC-3 and $\sim$350 22Rv1) containing more than 1500 cells.

The next step in creating our dataset was annotation of the measured images. Every image was manually segmented, which yielded segmentation masks of the whole cell, nucleus and nucleolus. These masks will later serve as so-called *ground truth* images in the process of training the neural network. Since the image resolution in 3rd dimension is relatively low and using all 96 slices as an input would be computationally costly, it was suitable to convert the original 512x512x96 3D image to a single-channel grey-scale 2D image: (1) by averaging through the 3rd dimension, (2) by maximum intensity projection over the 3rd dimension, (3) by standard deviation projection over the 3rd dimension and (4) by selecting the slice with the highest entropy value, thus the sharpest one. The data modified in this way contains most of the information needed for segmentation purposes.

2.3 DATA PREPROCESSING

The data had been normalized using Z-score to ensure its homogeneity. Every type of a converted 2D image was normalized separately with the corresponding values of mean and standard deviation computed through the whole dataset. Judging by the fact that segmentation of sub-cellular regions is a relatively complex task and our dataset is not large enough, the implementation of augmentation is essential to acquire the ability to train extensively and thus achieve the desired results. Our augmentation algorithm consists of: (1) addition of a random number from a predefined range, (2) multiplication by a random number from a predefined range, (3) affine transformation (random rotation, translation, shear and tilt), (4) random blurring and sharpening. All of these augmentation parameters had been set based on prior testing. After augmentation, the input image was randomly cropped to the size of 256x256 pixels.

2.4 IMPLEMENTATION

All mentioned algorithms were implemented in Python using a deep learning library PyTorch. For the purposes of this work, the neural network of choice was U-net [3], the most widely used architecture for biological image segmentation, as well as a U-net pretrained on an ImageNet database with a ResNet18 encoder [4] to compare to and possibly improve the results. Both networks have been trained on all 4 of the above-mentioned 2D images separately in addition to a stack of these 4 images. To properly segment all 3 regions (cell, nucleus and nucleolus), it was necessary to train 3 separate models, 1 for each region and input image type. This resulted in 30 trained models in total. Every model was trained with the same parameters set based on thorough testing and optimization.

The data-set was split in a 4:1 ratio; thus, 80% of the images were used for training and 20% for testing. The number of training iterations was set to 1000 with a batch size of 16. The network's parameters were optimized by the Adam optimizer. The scheduler had been reducing the initial learning rate of 0.0005 to 10% of its prior value at iterations number 500, 750 and 900. The model's perfor-

mance during training was measured by dice loss, computed against manually segmented ground truth images. The whole training algorithm ran on the remote server in the Department of Biomedical Engineering, Brno University of Technology with following computer specifications: Intel(R) Core i7 3.07 GHz CPU with an 8GB GDDR5 NVIDIA GeForce GTX 1070 GPU.

3 RESULTS

The performance of all models have been evaluated after training on testing data with no cropping applied (512x512 input image size) and by the same metric as during training - dice coefficient. Average values are shown in Table 1, from which the following conclusions can be made: (1) different 2D input image types seem to have only a negligible effect on the final performance, (2) stack of all 4 2D images as an input does not improve the performance, and (3) pretrained network allowed for a substantial improvement in nuclei and nucleoli segmentation and slight improvement when segmenting whole cells. Based on these results, the most useful approach is to use the slice with the highest entropy value as an input for the pretrained network. Utilizing mentioned approach, dice coefficients for both cell lines have been computed separately, values in Table 2. It is clear that segmentation of the 22Rv1 cell line is more problematic, with notable drop in nucleoli segmentation performance. Figure 1 shows examples of PC-3 and 22Rv1 cells with binary segmentation masks compared to ground truth masks.

Although the coefficients do not suggest an outstanding performance, we believe that the network learned to segment cellular structures reasonably well. Many nuclei and nucleoli cover only a small number of pixels. This leads to a relatively high error when the manually segmented mask does not cover these regions perfectly and the network predicts it correctly. Secondly, considering that no staining had been applied when measuring the cells (label-free), the results are reasonably high.

Table 1: Average dice coefficients for various types of input images and sub-cellular regions on testing data using both networks.

Input image type	U-net trained from scratch			Pretrained U-net		
	Whole cell	Nucleus	Nucleolus	Whole cell	Nucleus	Nucleolus
Mean	0.75	0.59	0.44	0.84	0.79	0.60
Max	0.80	0.61	0.50	0.83	0.77	0.59
Entropy	0.79	0.62	0.53	0.83	0.78	0.63
Std	0.80	0.56	0.49	0.84	0.78	0.61
Stack of all 4	0.79	0.57	0.51	0.83	0.78	0.62

Table 2: Average dice coefficients for PC-3 and 22Rv1 cell lines and their sub-cellular regions on testing data using pretrained network with the highest entropy slice as an input.

Cell line / Region	Whole cell	Nucleus	Nucleolus
PC-3	0.86	0.83	0.73
22Rv1	0.79	0.71	0.51

4 CONCLUSION

Our work was focused on the automatic segmentation of 2 prostate cancer cell lines using deep learning methods. The proposed algorithm manages to distinguish between cells, nuclei, nucleoli and background reasonably well, considering the fact no specific staining was used (label-free imaging). Our effort to achieve better results revealed that the use of a pretrained model leads to a substantial
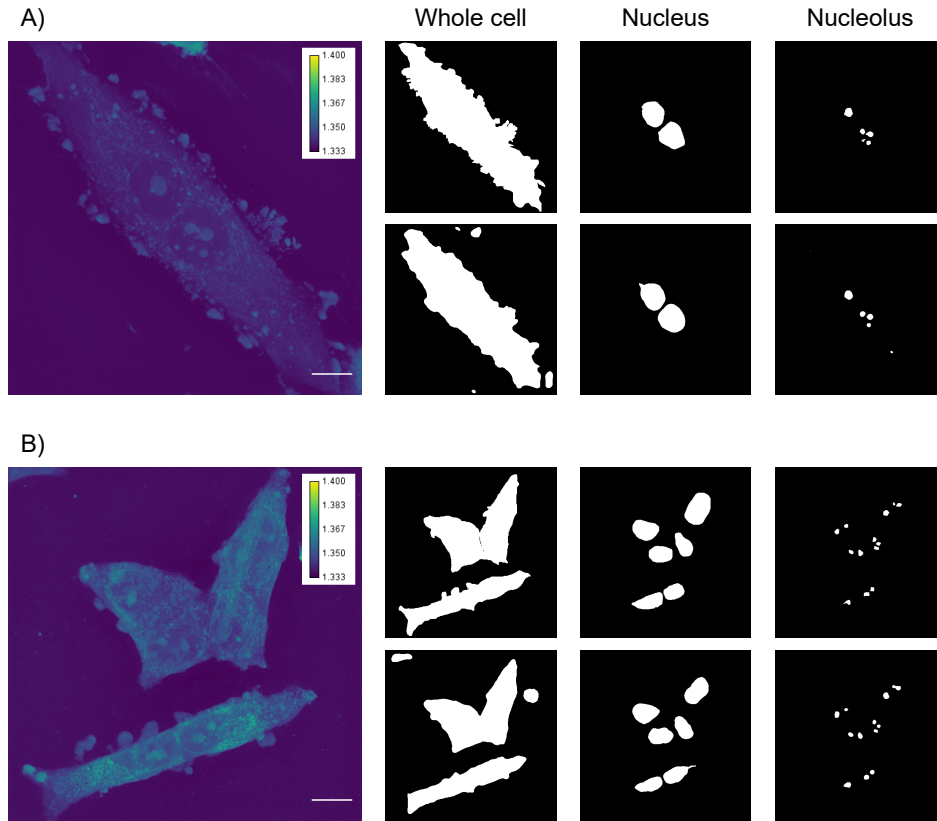
Figure 1: Examples of A) PC-3 and B) 22Rv1 cells, quantitative phase imaging, max projection, scale bar indicates 10 μm; their corresponding binary ground truth masks (top rows) compared to binary masks generated by the best performing model (bottom rows), i.e. pretrained U-net with a slice with the highest entropy value as an input.

performance improvement, especially for nuclei and nucleoli, 21% and 16% respectively. The process of segmentation is often necessary for subsequent analysis, and our trained models eliminate the need to segment these images manually. This allows for automation and potentially removing human error. Our subsequent work will utilize this algorithm and generate masks to quantitatively analyze the cellular properties and characteristics of the mentioned cell lines.

REFERENCES

[1] Live Cell Imaging Microscope: 3D Cell Explorer. Dec 2020.
URL https://www.nanolive.ch/products/3d-microscopes/cx/

[2] Holubova, M.; Axmanova, M.; Gumulec, J.; aj.: KRAS NF-κB is involved in the development of zinc resistance and reduced curability in prostate cancer. *Metallomics*, ročník 6, č. 7, 2014: s. 1240–1253.

[3] Ronneberger, O.; Fischer, P.; Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, Springer, 2015, s. 234–241.

[4] Yakubovskiy, P.: Segmentation Models Pytorch. 2020.
URL https://github.com/qubvel/segmentation_models.pytorch

ACCELEROMETER-BASED MONITORING OF STEERING WHEEL ANGLE

Tadeáš Navrátil

Bachelor Degree Programme (3.), FEEC BUT

E-mail: xnavra65@stud.feec.vutbr.cz

Supervised by: Ondrej Mihálik

E-mail: xmihal06@stud.feec.vutbr.cz

Abstract: This paper deals with the design of a device for measuring the steering wheel angle. The main goal was to design a basic concept of the device which is using data gathered from three accelerometers in the reconstruction algorithm to get the steering wheel angle. Furthermore, the paper deals with problems related to reconstruction algorithm errors and design improvements.

Keywords: Sensors, MEMS technology, Accelerometer, Gyroscope, Arduino, I2C, SPI, MATLAB, Simulink, Mathematical model, Measurement, Rotation angle, Microcontroller

1 INTRODUCTION

Observation of driver behavior is essential for the design of safety measures on roads, in cars, but also other areas of motoring research. Among the most important data for observing driver's behavior are the data of steering wheel angle.

The sensor of steering wheel angle is part of the project, that tries to monitor driver behavior a highway drive. To sense rotation angle we have chosen a slightly unconventional method based on three accelerometers, which sense the inclination to gravitation acceleration. A big advantage of this method is the portability and independence of the car model. Thanks to this method, we do not have to deal with communication with control units of different car models to obtain data of steering wheel angle. Data from the sensor of steering wheel angle will be sent to a computer, where they will be gathered and analyzed with other data from other sensors.

2 HARDWARE

The whole device consists of a microcontroller Arduino UNO and three inertial sensors GY-91 with a 3-axis accelerometer and 3-axis gyroscope. GY-91 allows communicating with a microcontroller using SPI or IIC bus. For communication between Arduino and inertial sensor, we've chosen SPI bus because SPI bus has a higher transfer speed than IIC. Also, IIC has a problem with 3 and more same devices on a single bus, because GY-91 has just two optional IIC addresses, so there would be a collision on the bus.

3 RECONSTRUCTION ALGORITHM

As I mentioned before, the whole device consists of three inertial sensors. Two of them are situated directly on the steering wheel and the third sensor is used as a reference. Outline of sensor placement is illustrated by Figure 1. For the start, we will consider a steering wheel perpendicular to the ground. The principle of measuring steering wheel angle is based on measuring the inclination of the gravitational acceleration vector relative to the coordinate system of each accelerometer on the steering wheel. Unfortunately, both accelerometers are not affected only by gravitational force, but also by
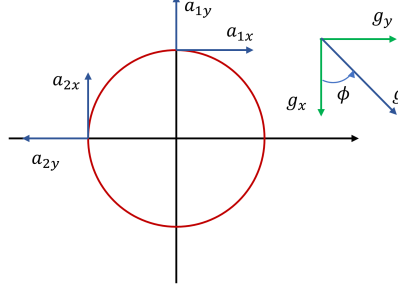
Figure 1: 2D model of steering wheel with accelerometers

centrifugal and inertial force while turning the steering wheel. If we take a closer look at the Figure 1, we can realize, that centrifugal and inertial force, which affect both sensors on the steering wheel has the same magnitude. The effect of gravitational force differs. To get rid of the effect caused by centrifugal and inertial force, we subtract corresponding force components from each sensor. After this, we get Equations (1) and (2)

$$\vec{a_{1g}} = (-g\sin(\theta-\phi), -g\cos(\theta-\phi)) \tag{1}$$
$$\vec{a_{2g}} = (-g\cos(\theta-\phi), \ g\sin(\theta-\phi)) \tag{2}$$

where g is the magnitude of acceleration caused by the outer world, ϕ is the angle of its rotation (see Figure 1) and θ is steering wheel angle. If θ is positive, the steering wheel is turned in the positive sense of rotation. To ensure that components of centrifugal and inertial force cancel out, we have to synchronize the sampling of both accelerometers. To ensure that the data from both accelerometers are synchronized, we used FIFOs of the sensors. We also have to fasten inertial sensors on the steering wheel as precisely as possible. During a car drive, we have to consider disturbances. For example car shocks, or forces affecting the car during turning, etc. These effects can be compensated using a third accelerometer, which detects the resultant of all forces affecting the car instead of just the gravitational force mentioned earlier. By adjusting the equations (1) and (2) we can get (3), by means of which we can determine the angle of the steering wheel.

$$\theta = \arcsin\left(\frac{a_{1x}-a_{2x}+a_{1y}-a_{2y}}{-2\sqrt{g_x^2+g_y^2}}\right) + \arctan\frac{g_y}{g_x} \tag{3}$$

Using rotation matrix, we can generalise this equation to an arbitrarily inclined steering wheel. We arrive at two equivalents of Equation (3), which use either arccos or arcsin. In the final algorithm we dynamically combine these two functions in order to minimize effect of noise and maximize accuracy. For that reason we divide measured interval $\langle-\pi,\pi\rangle$ to five smaller intervals.

4 MOUNTING OF SENSORS ON THE STEERING WHEEL

Inertial sensors attached to the steering wheel have to meet some requirements. They have to be fastened strongly enough, so they do not move during steering wheel rotation. Concurrently they have to be placed perpendicular to each other, at the same distance from the axis of rotation, for the algorithm to work properly. Furthermore, the sensor position cannot limit driver in-vehicle control or block certain safety or control features (e.g., airbag or horn).

Therefore we have created a system very similar to a wristwatch. It consists of a belt from raw material and a buckle from a 3D printer with an inertial sensor case attached to it. When we want to

place the sensor, we just turn around the steering wheel frame belt through the buckle and tighten it up. Sensors are placed on the steering wheel according to Figure 1, so they do not disturb the driver, because recommended position of the hand on the wheel is (if you imagine dial) on 2 and 10 hours.

5   IMPROVEMENTS OF THE RECONSTRUCTION ALGORITHM

In Section 3, a basic algorithm calculating steering wheel angle was described. In this algorithm, we neglected some facts like the tilt of the steering wheel or inaccuracy in inertial sensor position.

5.1   COMPENSATION OF SENSOR ROTATION AROUND X AXIS

The way the sensors are attached ensures his stability on the y- and z-axis. Unfortunately, sensors can be rotated around the x-axis. To determine the sensor rotation angle, we devised a calibration procedure. During the procedure user oscillate with the steering wheel. Based on data from the gyroscope we can determine this angle. Then we compensate this deviation using a rotation matrix.

5.2   DETERMINATION OF STEERING WHEEL TILT

In Section 3 we assumed, that the steering wheel is perpendicular to the ground. Normally the steering wheel tilt is adjustable. Therefore before every measurement, we have to determine the steering wheel tilt and then compensate it with a rotation matrix. Using accelerometers and the angle that their axes make with the gravitational acceleration vector, we determine the tilt angle of the steering wheel.

5.3   DETERMINATION OF MULTIPLE ROTATIONS

As mentioned above, the interval $\langle -\pi, \pi \rangle$ in which we are calculating rotation angle is divided into five smaller intervals. The program remembers the last interval the steering wheel was in. If there is a transition between the extreme intervals, we add or subtract one from the state variable k. The resulting rotation angle will be given by the sum of the rotation angle based on the equation 3 and the k-multiple of 2π rad.

5.4   USAGE OF COMPLEMENTARY FILTER

For further enhancement of the basic algorithm, we combined the data from accelerometer system with the measurement using gyroscopes. To measure rotation using gyroscopes is very simple. If sensors are perfectly attached, then rotation angle corresponds to the integral of z-axis signal of both accelerometers on the steering wheel. Gyroscopes, unfortunately, measure signal with an offset, which is present even then if the gyroscope is not rotating. Therefore gyroscopes have problems with noise on low frequencies. On the other hand, accelerometers have problems with noise on high frequencies.

A complementary filter is designed in a frequency domain by taking the high-frequency part of the spectrum from the data we obtain from gyroscopes. We complement this data with the data obtained from accelerometers so that we again obtain the entire frequency spectrum. The complementary filter describes the Equation (4), where ω is an optional parameter chosen from the interval $(0,1)$. [3]

$$\theta \quad = \quad \omega\, \theta_{gyro} + (1-\omega)\theta_{akc} \tag{4}$$

6   PRACTICAL IMPLEMENTATION OF DEVICE PROTOTYPE

For practical implementation of the device prototype and verification of its functionality, we originally intended to use a driving simulator in the laboratories of BUT FEEC. Due to the current situation,

we were forced to test this device prototype on a gaming steering wheel *Thrustmaster Ferrari Red Legend Edition*. This steering wheel has some limitations including limited rotation range, which was approximately $\left\langle -\frac{\pi}{2}, \frac{\pi}{2} \right\rangle$. Therefore we could not test the system's capability to detect multiple rotations.

Because this prototype was tested at home, we did not have to comply with some requirements for the device, which are necessary for its actual use. For example, we placed the control microcontroller in the center of the steering wheel, which must be free in the car due to the presence of an airbag. The microcontroller is connected to the sensors via a non-soldering field. The final version of the device prototype can be seen in the Figure 2.
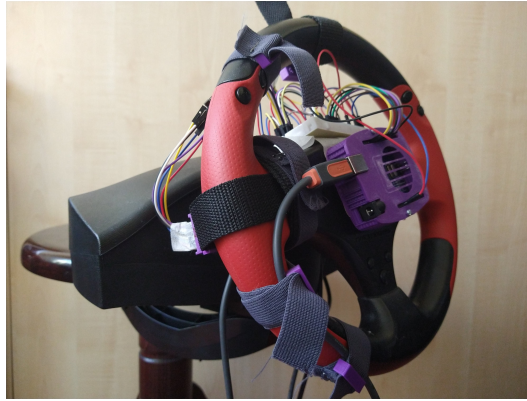
Figure 2:   Final version of device prototype

The final device should work so that each inertial sensor will have its own microcontroller (e.g., Arduino Nano), which will send wirelessly directly to the data gatherer. Wireless communication is not part of this paper.

7   CONCLUSION

The paper aimed to create a device for measuring the rotation of steering wheel angle. It was important for the device to be portable and independent of the type of vehicle. The steering wheel angle is obtained from the data of three inertial sensors. After building the basic algorithm, we extended it to detect and compensate for the rotation of the sensors and the steering wheel itself compared to the ideal position. Because the steering wheel in classic cars can make more turns on one side, we have extended the interval in which the device can measure. Finally, we used a complementary filter to improve the measurement results by combining data from accelerometers and gyroscopes. We first simulated the device in the MATLAB Simulink environment and then we assembled its prototype, attached it to the game steering wheel and tested it. The results of the sensor are very good and correspond to the actual angle of rotation.

REFERENCES

[1] FRADEN, Jacob. *Handbook of modern sensors: physics, designs, and applications.* 4th ed. New York: Springer, c2010. ISBN 9781441964663.

[2] Invensense.tdk.com. 2021. *MPU-9250 Product Specification Revision 1.1.* [online] Available from URL: shorturl.at/mMPR5 [Cited March 8, 2021].

[3] de Maele, P., 2021. *V Reading a IMU Without Kalman: The Complementary Filter*[online] Available from URL: https://www.pieter-jan.com/node/11 [Cited March 8, 2021].

DEEP NEURAL NETWORK FOR AUTONOMOUS UAV NAVIGATION

Jan Klouda

Bachelor Programe (3), FEEC BUT

E-mail: xkloud04@stud.feec.vutbr.cz

Supervised by: Jiří Janoušek

E-mail: xjanou09@stud.feec.vutbr.cz

Abstract: The project deals with autonomous drone control. A neural network is used to create autonomous control for object recognition. This recognition is performed with a ground station, where the computer evaluates the position of the drone and autonomously controls the flight of the drone through the detection of objects.

Keywords: UAV, drone, Python, autonomous navigation, object recognation

1   INTRODUCTION

The article deals with the creation of a fully autonomous drone. The basis of control of this aircraft is the ground station, which receives telemetry data and images from the aircraft. It processes this data and creates an autonomous drone through communication with the flight unit. The ground station contains a neural network that recognizes dozens of objects from the camera image on the drone in real time, this creating an overview of the position of objects. This makes it easy to control a drone and make a various tasks, from tracking a car or a human, through a very precise landing, to reconnaissance or rescue missions. I wrote this whole program in Python and successfully tested it in various situations. Wi-Fi telemetry or a 433 MHz antenna is used for communication between the ground station and the flight unit. I decided to choose a personal computer with a graphics card as the ground station, because the program requires higher performance for smooth running and immediate detection.

2   CONNECTION OF ARTIFICIAL NEURONAL NETWORK WITH FLIGHT UNIT

The connection of the whole system is described in Picture 1 as a block diagram, where the UAV (Unmanned Aerial Vehicle) is controlled by a ground station. The UAV consists of several basic blocks. The flight controller unit here functions as an aircraft control element, calculating data from accelerometer, gyroscope, barometer and GPS coordinates. The flight controller also receives commands from the ground station (personal computer) using wifi telemetry module.
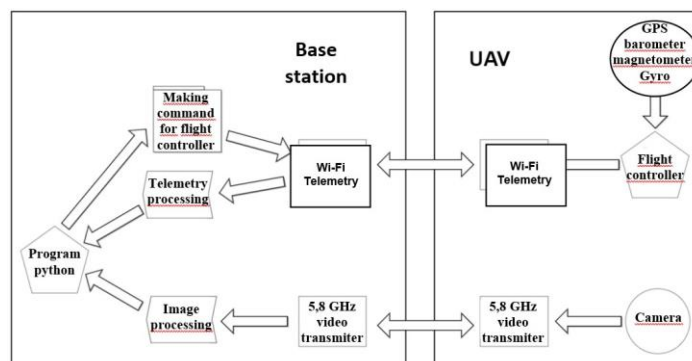
Picture 1: Block diagram of autonomous flight

To make the connection, i use a Wi-Fi module connected to the flight controller. I set the flight controller to send flight information (telemetry) and receive control signals via this module. As the Ground Station, on which the computer code runs, i choose a personal computer. To start communication, it is necessary to connect the computer to the created Wi-Fi network, and create MAVLink communication via the UDP (User Datagram Protocol) port. My program first creates this communication channel with the flight controller, and then communicates via MAVLink protocols. MAVlink is a protocol used between drones to transmit control signals, aircraft status information and other configuration messages between the drones and the ground control station. The information is published in a hybrid way, namely publish-subscribe (for sending data streams) and point-to-point (for configuration protocols, mission protocol, or parameters). The so-called dialects, ie message sets supported by a specific MAVlink system, are used to transmit information. Picture 2 describes the structure of the used MAVLink protocol.
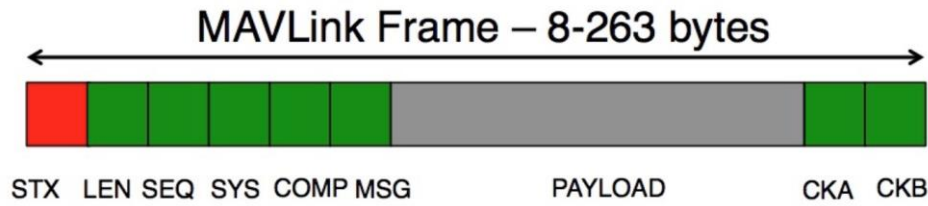
Picture 2: Structure MAVLink packet protocol [1]

2.1 YoLo NEURAL NETWORK

The use of neural networks for object detection eliminates the disadvantages of cascade detection systems. The difference is that the neural network learns itself, and therefore the monitored object can be detected from different angles. Thus, the accuracy of using neural networks has increased. Thus, a learned neural network is able to detect a person, even if only his hand is visible in the picture. Therefore, i use matrix scales to learn such a neural network, and i gradually adjust them for a given object. Many calculations are required to learn this network and start detection. For example, the Darknet library (used for easier operation of neural networks on a computer) can use the computer's graphics processor to calculate neural networks using the CUDA toolbox, increasing the speed of object detection [2].

The YOLO (You look only once) recognition system works on the same principle as most neural networks, but uses a different approach. It applies a single neural network to the entire image, which divides the image into many frames delimiting objects, at different scales. These frames are created based on the probability for each area. Areas of an image with a high number of frames are thus considered as a very probable area with the occurrence of the searched object. YOLO is a great example of a single-stage detector, and approaches object recognition by considering detection as a regressive problem, taking an input image while learning the coordinates and probability of a bounded frame of a marked class. This means that one convolutional network predicts bounding boxes and probability classes for these boxes simultaneously [3].

Picture 3 illustrates how a program detect a person. This is a direct image of the camera on the drone during the test. The picture shows the marked object, as well as its position in image coordinates, as well as commands sent to the flight unit as "forward" and "right". based on this information, calculate the relative position of the drone to the monitored object (person). Thus, if a person approaches the drone to a defined distance, the ground station decides that the detected person is too large in the image and sends a command to the flight unit to fly backwards. The same control of the distance between the drone and the object to be observed occurs when the object in the field of view of the camera is too small, and the ground station sends a command to fly forward. If the position of the object leaves the center of the camera's field of view on the drone, ie the person moves to any side, the program adjusts the position and rotation of the drone so that the monitored object is always in the middle of this field of view. Thus, human tracking is ensured. A similar pro-

cedure occurs during landing. The drone begins to descend on the landing surface, and optically monitors where it is located, in this way i created an accurate landing. All these calculations and controls are performed 20x to 30x per second, so that the control is sufficiently fine and thus responds immediately to a change in the position of the object. The biggest slowdown in code is the detection of neural networks, and when using a more powerful graphics processor, this detection becomes faster and control can theoretically speed up.

Picture 3: Tracking a moving person

2.2 CODE SAMPLE

I wrote the whole program in Python. Most often i used the OpenCv libraries (for image preprocessing and image field creation), the Dronekit library (for creating MAVLink packets and thus ensuring communication between the flight unit and the ground station) and the Darknet library, which uses YoLo neural networks and using this libraries detect objects. Source code 1 describes the detection of an object in each camera frame using a neural network, and then calculates the coordinates of the object in the field of view. Source code 2 describes the creation of the MAVLink protocol for sending control commands directly to the drone flight unit [4].

```
if frame_resized is not None:
    image=darknet.draw_boxes(detections,frame_resized,class_colors)
    image = cv2.cvtColor(image, cv2.COLOR_BGR2RGB)
    ax=dx=dy=en=w=0
    directionx=directiony=""
    dy,dx,en,w = darknet.middle(detections,dy,dx)
    if en==1:
        dx=(width/2)-dx
        dy=(height/2)-dy

    if w<150:
        ax=gnd_speed
        directiony="vpred"
    if w>250:
        ax=-gnd_speed
        directiony="vzad"
```
Source code 1: Source code for calculating the position of a detected object

```
def set_velocity_body(vehicle, vx, vy, vz):
    msg = vehicle.message_factory.set_position_target_local_ned_encode(
            0,
            0, 0,
            mavutil.mavlink.MAV_FRAME_BODY_NED,
            0b0000111111000111,
            0, 0, 0,
            vx, vy, vz,
            0, 0, 0,
            0, 0)
    vehicle.send_mavlink(msg)
    vehicle.flush()
```
Source code 2: Source code for sending control signals to the flight unit

3   CONCLUSION

In making this project, i used multiple neural networks and different types of object detection. In the project, i solved the use of neural networks for autonomous control and created an autonomous UAV that is controlled based on object recognition. I connected the neural networks to the flight unit and created a program for autonomous control. YoLo neural networks were the most accurate for my use and detected multiple objects at once, so i could create subconscious places where the drone is located, and then create a program that uses this data to evaluate where the drone should move. The whole system is autonomous, and during the flight the drone is controlled only by computer code commands. The tests are documented on camera. Because communication with the ground unit is required when using the YoLo neural network, i have created a program that uses a similar detection method but is not hardware intensive. This code modification is also suitable for Raspberry Pi microcomputers. The microcomputer is mounted on the drone and the system thus becomes fully independent. I have also successfully tested this modification. I also use this project in my bachelor's thesis, where i evaluate the results of detection, various modifications of my code, and evaluate the results of tests.

REFERENCE

[1]   MAVLink [online].   [cit.   2021-03-02].   from: https://erlerobotics.gitbooks.io/erlerobot/content/en/mavlink/mavlink.html

[2]   YOLO object detection with OpenCV [online]. 2018 [date: 2020-10-18]. from: https://www.pyimagesearch.com/2018/11/12/yolo-object-detection-with-opencv/

[3]   Haar Cascades vs Mask R-CNN [online]. 2018 [date: 2020-10-18]. from: https://www.reddit.com/r/learnmachinelearning/comments/8d0puz/haar_cascades_vs_mask_rcnn/

[4]   Darknet.   GitHub   [online].   2020   [date: 2021-5-25].   from: https://github.com/AlexeyAB/darknet

WEB SERVER SECURITY ANALYZATION TOOL

Samuel Kopecký

Bachelor Degree Programme (3), FEEC BUT

E-mail: xkopec52@stud.feec.vutbr.cz

Supervised by: David Smékal

E-mail: smekald@feec.vutbr.cz

Abstract: Nowadays it is common practice to have a securely hosted web server. To have a secure web server it is required to ensure that CIA (Confidentiality, Integrity, Availability) triad is in place. CIA triad is fulfilled by using different mechanisms like symmetric cryptography and public key infrastructure. These mechanisms on the other side use a set of algorithms and parameters. This paper focuses on the automation of analyzing the security of these algorithms, parameters and also vulnerability testing.

Keywords: SSL, TLS, Web Server, Automation, Security, Analyze

1 INTRODUCTION

In order to use a websites safely, it is necessary to secure confidentiality(protection against unauthorized access to the communication), integrity(protection against the modification of sent information) and authenticity(identity verification of the communicating entities) of the data being transmitted. The HTTP (Hypertext Transfer Protocol) protocol itself, which is used to transfer the data between a client and a server, doesn't include any mechanisms to ensure these requirements. That's why the protocol HTTPS (HTTP secure) was created, to ensure the communication is secured via a "tunnel" created by the protocol SSL/TLS(Secure Socket Layer/Transport Layer Securiry) [1].

The protocol TLS contains many mechanisms, which it uses to secure confidentiality, integrity and authenticity [2]. In order to use these mechanism, the client and server need to agree on specific parameters and algorithms that will be used during the transmission of messages.

However, before the client and server exchange the required information, every TLS connection needs to start with an exchange of messages using the handshake protocol. If the client wants to establish a secure connection with the server, he first sends a message notifying him that he wants to create a connection. The server then responds to this message. It is during the handshake protocol that the communicating entities exchange the necessary algorithms and parameters to establish a secure connection. [2]

2 PROGRAM FUNCTIONALITY

The program is written in the programming language Python and it is able to analyze the parameters and algorithms that the client and server exchange during the communication establishment. More precisely it analyzes the supported TLS protocol versions, implementations of a web server, server certificates, used cryptographic parameters and algorithms (cipher suite) and a chosen set of vulnerabilities, to which the server may be vulnerable. For the analyzation of the cipher suite, supported versions of the TLS protocol and vulnerabilities a custom implementation of the analyzation algorithms is used. Certificate analyzation is done by using the python library cryptography [3]. Scanning of the web server implementation is performed using the Nmap tool and by using the information inside the HTTP response headers. The collected data is then printed to the user of the program.

Functionality of the program is divided into several logical parts, that are as a whole responsible for the analyzation of the web server and its parameters. The following chapters describe these logical parts of the program. These logical parts are also illustrated by the figure number 1.
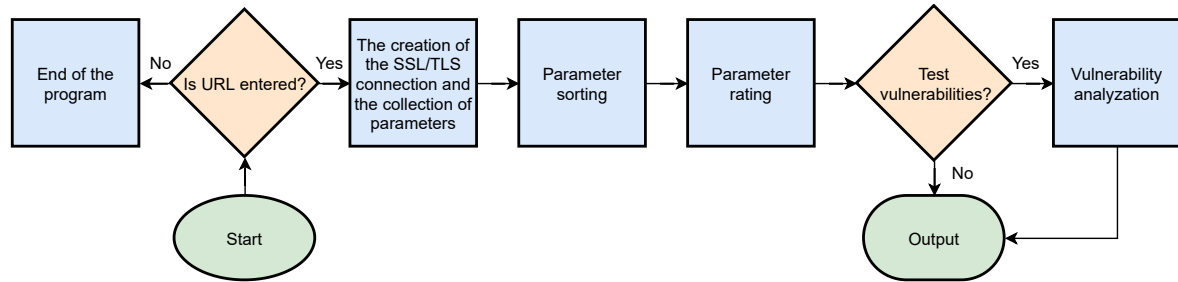
Figure 1: Program flowchart

2.1 COLLECTION OF THE WEB SERVER PARAMETERS

One of the first things that the program does is finding out the cipher suite. The cipher suite defines the algorithms that ensure confidentiality, integrity and parameters that determine the key lengths for these algorithms. The values in the cipher suite are separated with the character "_". Cipher suite is acquired by the program immediately after the creation of the connection with the server and is saved for future processing. Cipher suite example: TLS_ECDHE_RSA_WITH_AES_128_GCM_SHA256.

The next object that the program loads is the certificate used for the verification of the web server identity. This certificate contains the server public key with which the client is able to verify information that is digitally signed with the servers private key. The certificate also contains other information such as the algorithm which is used to create the digital signature. The program first loads the certificate in DER binary format and then converts it into a python object. To find out which versions of the TLS protocol the server supports, it creates a connection on the version of the protocol it is currently testing. If the server doesn't establish a connection, the server doesn't support that version. If a connection is established, the version is added to a python list and then is sent to the program output.

2.2 PARAMETER SORTING

The next action the program takes is parameter sorting or parameter extraction from the collected objects. Parameters from the certificate can be extracted easily, since the certificate is saved in a simple python object where parameters are saved as properties of the object. Unlike the certificate, the cipher suite is collected in a text format and the parameters positions are not strictly given. For example in one cipher suite the symmetric encryption algorithm is present on the 4th position (position is defined as parts divided with the character "_"). But in a different cipher suite it can be present on the 3rd position, if the cipher suite doesn't define a public key algorithm. That's why the program sorts the parameters by dividing them into a list with the "_" character. Then it loops through the list and compares the values with a predefined json file which contains another list of all possible parameter values for each parameter category. When the correct category of the parameter is found the value is saved together with the category into a python object.

2.3 PARAMETER RATING

When all the parameters are sorted and saved in python objects, the program rates them by their security. Of course only the parameters, that can be rated are rated, for example a certificate version

can't be rated. To rate the parameters, another json file is used, which contains 4 levels of security for each parameter category. Every level contains a list of parameter values. When the program rates a parameter it loads it and loops through all of the levels, until it finds a match with value of the rated parameter and one of the security levels. The parameter values of the security levels are created with the inspiration from the NIST recommendations, more precisely NIST SP 800-52 [4] and NIST SP 800-131A [5]. The number 1 denotes the highest security and number 4 the lowest.

Numeric parameters such as the key lengths, are rated similarly. They too have 4 security levels defined in the json file. Since these parameters can't be rated alone and depend on the algorithm for which they determine the length of the key, each numeric parameter has an assigned algorithm. Because of that, the security levels in the json file are defined differently. One value is defined for each key size and algorithm pair, which consists of the algorithm name, a sign of comparison and the key size. For example the value for the algorithm RSA with a key size of 2048 bites would like this: "1: RSA,>=2048". This means, that the algorithm RSA with a key size bigger or equal to 2048 has the security level of 1. After the acquisition of the security level of the parameter the program saves this value into a dictionary, where the key is the key size and value is the security level.

2.4 VULNERABILITY ANALYZATION

Finally the program analyzes the vulnerabilities of the web server. The implementation of the vulnerability tests depend on the nature of the vulnerability. Program supports several tests which are performed using custom implementations. Every test tests the the presence of the vulnerability. If more vulnerability tests are ran in one instance, every test is ran in one thread, to speed up the testing process. Vulnerability tests that the program supports are: Heartbleed, ZombiePOODLE/GOLDENDOODLE, attacks using renegotiation and CCS (Change Cipher Spec) Injection. These tests were chosen because their exploitation results in the voidance of TLS protocol security. Heartbleed allows for the collection of the web server private keys. POODLE attacks and CCS injection void the confidentiality ensured by the symmetric cryptography. Renegotiation attack compromises the integrity and in some cases even the confidentiality of the messages being sent [2].

3 TOOL USAGE

The program is ran as a python script and defines some optional arguments. It requires one mandatory argument, that is -u, which is used to enter the url of the web server to scan. Vulnerability tests can be triggered using the the test number and the -t argument. Some other arguments and test numbers can be displayed using the -h argument. In the standard mode the program output is specified as a text output (stdout) to the environment in which the program is running. However, if the -j argument is specified together with output file name, the program outputs all information about the web server to the given file.

In picture number 2 a shortened program output is shown. The output consists of 4 sections and each part has its own name and its elements are intended. The first sections holds the cryptographic parameters that are rated using numbers, which denote the security level. Parameters of the first section are extracted from the cipher suite and from the the certificate. The final rating of the parameters is at the end of this section also.

The next section lists the TLS protocols that the web server supports and their ratings same as in the first section. The third section consists of the un-ratable certificate parameters, for example the serial number. In the last section the specific implementations of the web servers that are used by the server are listed. The program will also list the the results of the vulnerability tests after they are implemented.

```
Cryptographic parameters:                              Certificate information:
        Key exchange algorithm: RSA->2                         Valid from: 2020-04-21
        Symmetric encryption algorithm: AES->1                 Valid until: 2021-04-26
        Symmetric encryption algorithm length: 128->1          Subject:
        Symmetric encryption algorithm block mode: CBC->2              localityName=Brno
        Hash function: SHA->2                                         commonName=www.vutbr.cz
        Public key algorithm: RSA->1                           Issuer:
        Public key length: 2048->1                                   commonName=TERENA SSL CA 3
        rating: 2                                              Alternative names:
Protocol support:                                                    vut.cz
        TLSv1->2                                        Web server versions:
        TLSv1.2->1                                              http_header: Apache
        rating: 2                                              nmap: Apache httpd
```

Figure 2:   Shortened program output

4 CONCLUSION

During the development of the program, its functionality was tested on Linux distributions such as Linux, Mint and Debian using Docker containers. The program was working correctly in these environments. Next the program was tested on WSL (Windows Subsystem for Linux) but during the analyzation of the web server vutbr.cz an error occurred. The web server vutbr.cz supports TLS versions from 1.0 to 1.2, but the program listed only version TLS 1.2. This error happened because the implementation of the OpenSSL library was compiled only with the support for protocol TLS 1.2 and higher. Since the Python interpreter, which interprets the program uses the same instance of the OpenSSL library, it was unable to create a connection even if the server supported it. However, the program works correctly because if the operating system of the user has the OpenSSL library compiled such that it doesn't support versions 1.0 and 1.1 the attacker is unable to force the user to use a lower version protocol. There are many other TLS scanning tools freely available on the internet, such as the website ssllabs.com. However, these tools run on 3rd party servers. This means if many users utilize this service the scanning may be slowed down. The advantage of this tool is that the user can run the program on his own computer or run multiple instances of the program to scan multiple servers parallelly using his own computing power.

REFERENCES

[1] GOURLEY, David, Brian TOTTY, Marjorie SAYER, Anshu AGGARWAL and Sailu REDDY. *HTTP: The Definite Guide: Understanding Web Internals*. O'Reilly Media, 2002, 658 s. ISBN 9781565925090.

[2] RISTIĆ, Ivan. *Bulletproof SSL and TLS: Understanding and Deploying SSL/TLS and PKI to Secure Servers and Web Applications Ivan Ristic*. 6 Acantha Court, Montpelier Road, London W5 2QP, United Kingdom: Feisty Duck, 2014. ISBN 978-1-907117-04-6.

[3] Cryptography: GitHub.   *GitHub*   [online].   [cit.   2021-7-3].   Avaliable   at: https://github.com/pyca/cryptography

[4] MCKAY, Kerry A. and David A. COOPER. NIST SP 800-52. *Guidelines for the Selection, Configuration, and Use of Transport Layer Security TLS) Implementations*. 2nd ed. NIST, 2019, 72 s. Avaliable at: https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.800-52r2.pdf

[5] BARKER, Elaine and Allen ROGINSKY. NIST SP 800-131A. *Transitioning the Use of Cryptographic Algorithms and Key Lengths*. NIST Pubs, 2019, 33 s. Avaliable at: https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.800-131Ar2.pdf

UNIVERSAL SECURITY SOLUTION FOR IOT DATA COLLECTION SYSTEMS

Petr Šťovíček

Bachelor Degree Programme FEEC BUT

E-mail: xstovi04@stud.feec.vutbr.cz

Supervised by: Petr Dzurenda

E-mail: dzurenda@vutbr.cz

Abstract: The article presents our data collection solution based on RIOT operating system and Message Queuing Telemetry Transport (MQTT) technology. Our solution is secure, universal, and easy to integrate to already existing and often insecure IoT networks. Furthermore, we integrated our solution into the temperature and humidity monitoring system. This system detects the potentially dangerous level of these physical quantities and warns users by smartphone notification about relevant threats such as the impact on the infectivity of COVID-19.

Keywords: IoT, Internet of Things, Cryptography, Constrained Devices, Microcontrollers, MQTT, RIOT, OpenHAB, COVID-19

1 INTRODUCTION

The current growth of Internet of Things (IoT) systems and Industry 4.0 is closely associated with connecting a large amount of computationally and memory-constrained devices to the Internet. These devices usually consist of microcontrollers with very different hardware and software resources. Common cryptographic schemes are often impossible to implement because of their computational and memory requirements. This leaves IoT systems susceptible to all kinds of attacks.

One of many application scenarios for these devices is data collection systems such as air quality evaluation systems. Recent research [1] on air quality suggests that temperature and humidity have a direct impact on the viability and transmissibility of COVID-19. Specifically cold and dry environment increases spread of this virus. Therefore, monitoring and maintaining the optimum temperature and humidity in the room is desirable, thus slow down the spread of viruses.

In this article, we proposed a system that allows a user to monitor and react to the current temperature and humidity in the rooms where transmission may occur. Collected data is then analyzed, evaluated, and interpreted to the user. The proposed system is based on MQTT [2] IoT messaging protocol and RIOT [3] operating system. RIOT allows the creation of a program that can be executed on devices from various manufacturers and architectures hence providing hardware independence. That can lead to significant cost saving during project creation.

2 SYSTEM ARCHITECTURE

The proposed data collection system works in a client-server model, where all communication takes place only between the client and the server. Clients do not communicate with each other but only with the server. Clients are represented as various microcontrollers combined with sensors. The data collection server is in the form of a microcomputer. Widely spread WiFi technology is used for data transfer between system elements, this enables system deployment even in places where routing cables is not possible.

The MQTT protocol ensures the transmission of messages between clients and the server. This protocol is simple and lightweight. The MQTT server is called `broker`, its primary role is to route messages between clients, but it can also provide client authentication and access control. There are two types of clients. A `publisher` is a client that sends data to the broker. A `subscriber` is a client that receives data from the broker. MQTT messages are sent to so-called topics. The publisher sends data to the topic via message `publish`. If the subscriber is subscribed to this topic, the message will be forwarded to him by the broker. A Client can be a subscriber and a publisher to many topics at the same time. MQTT alone does not offer a sufficient level of security. For this reason integrity, authenticity, and confidentiality of the transmitted data are additionally ensured with available cryptographic mechanisms.

Security of transmitted data is ensured by the Advanced Encryption Standard (AES) cipher in the authenticated mode Counter with Cipher block chaining Message authentication code (CCM). Each topic has its own encryption key which is assigned to the client by the central trusted authority Key Management Service (KMS) which provides authentication and authorization of clients. KMS maintains a database of clients and their rights to the topics.

Figure 1 depicts a sketch of the proposed system architecture. In step one, the client sends data over an insecure channel to the broker. Data in the message are encrypted with the topic key. In step two, the broker forwards the message to the MKS. KMS performs client authentication and authorization. If the process is successful, KMS decrypts data and inserts it into the new message, which is sent in step three to the broker. The broker distinguishes between two types of traffic. If the message comes from an insecure channel, it is considered as untrustworthy. If the message comes from KMS, it is considered as trustworthy. Only trustworthy messages can be written to the subtopics of KMS (e.g. `kms/topic`). In step four, the message is forwarded to the OpenHAB for storage. OpenHAB is an open-source automation tool that allows to store and further work with measured data [4]. Step five shows data presentation to the user which can take place via mobile application or web interface.
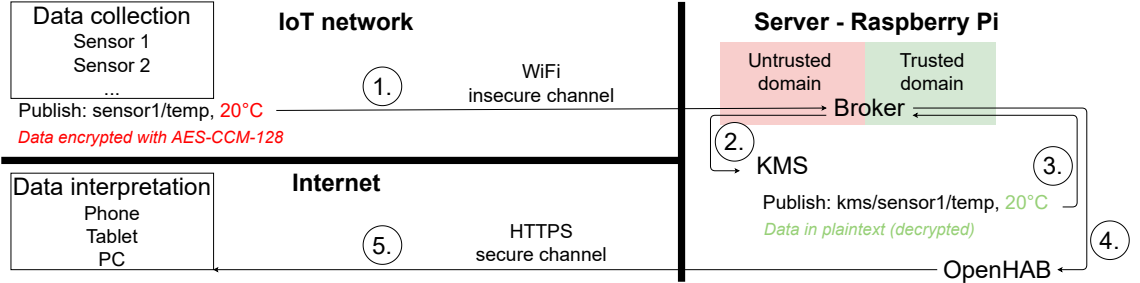
Figure 1: Simplified system architecture.

Figure 1 does not specify a key distribution. Before the client can encrypt any data he must first obtain an encryption key for a given topic (`topic_key`) by requesting it from KMS. Key requests are encrypted with session key (`temp_key`). Negotiation of the session key is achieved by using the Schnorr signature scheme combined with Diffie–Hellman (DH) protocol.

Whole process of session key (`temp_key`) creation, negotiation of the topic key (`topic_key`) and sending data to the topic is depicted in Figure 2. The process begins with a random value `nonce`, which is requested in step one and sent to the client in step two. The client uses this value to compute the session key that is then used to encrypt the name of the topic requested by the client. The ciphertext, signature (`e, z`) and client identifier `ClientID` are sent to the KMS in step three. KMS then verifies the validity of the signature, authenticates the client, and calculates the session key according to Equation 1.

$$temp\_key = A^{SK_{KMS}} = (g^r)^{SK_{KMS}} = g^{r \cdot SK_{KMS}} = (g^{SK_{KMS}})^r = PK_{KMS}^r \qquad (1)$$

In step four, KMS sends the topic key to the client along with the `counter_max` value, which indicates how many times the key can be used for encryption. This message is then decrypted by the client with the session key. Topic key negotiation is now complete, the client can use the obtained key to encrypt data in step five. KMS authenticates and decrypts received data and in step six it redirects it in plaintext to the trustworthy topic, which is OpenHAB subscribed to.
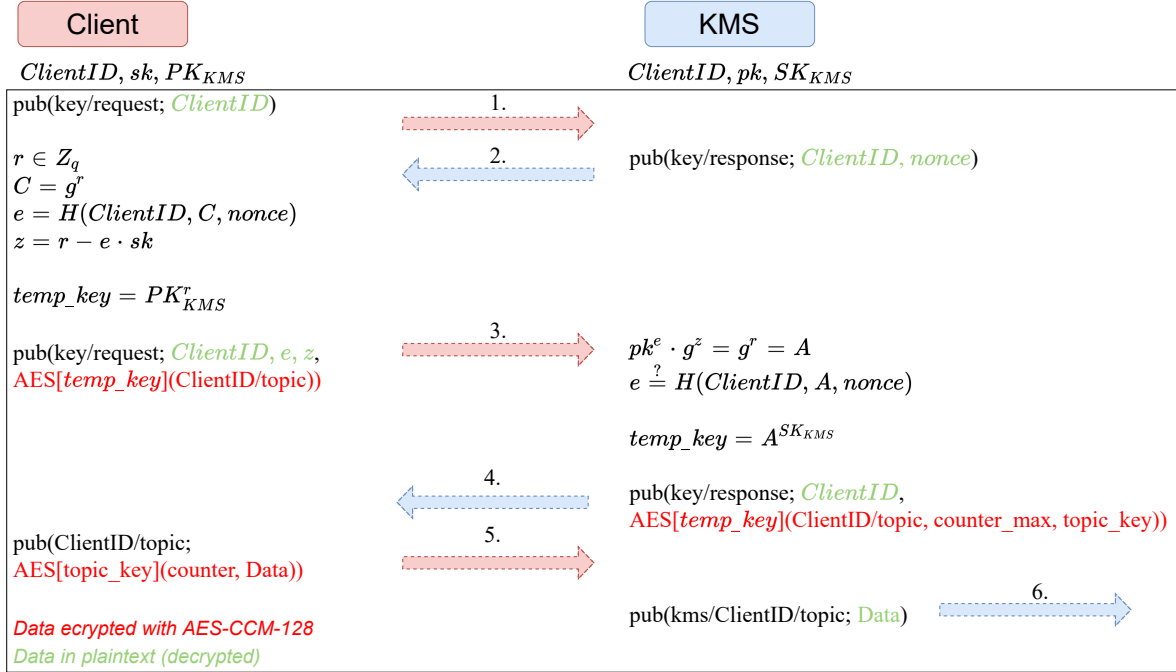
Figure 2: The cryptographic core of the proposed system.

3 SYSTEM IMPLEMENTATION

The proposed system utilizes Raspberry Pi 4 Model B as a server, which hosts MQTT broker Mosquitto, KMS (python script), and OpenHAB. ESP32 and ESP8266 microcontrollers paired with BME280 temperature and humidity sensors represent clients. Clients run RIOT OS, their code is written in C programming language. The cryptographic core is implemented over elliptic curves. Elliptic curve operations and modular arithmetic are provided by the micro-ecc [5] library, while the AES-CCM-128 cipher is provided by the crypto [6] library. Figure 3 shows the performance of both mentioned RIOT compatible libraries for the most computationally demanding operations. The micro-ecc library was selected over the Relic library due to smaller memory requirements.
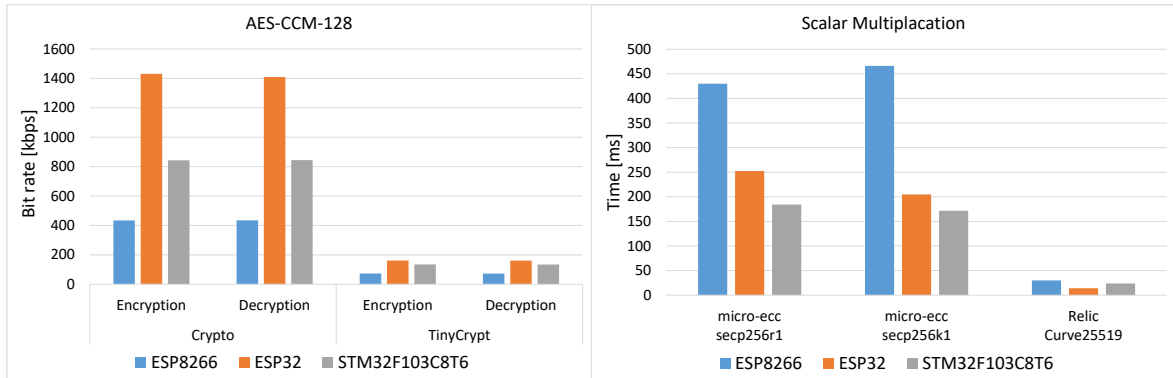
Figure 3: Performance tests: AES-CCM-128 (left), Scalar Multiplication (right).

Figure 4 shows the mobile application interface that is available for iOS and Android. The first from the left is the home screen, which shows an evaluation of the current states in the monitored rooms. When the set limit is exceeded, the background color changes from green to red, the warning message is displayed and the notification is sent. The remaining three images show graphs that appear when the user clicks on the current humidity/temperature on the home screen. Graphs are customizable (display min/max, change the length of the monitored period, combine graphs, etc.).

Figure 4: Screenshots from the mobile application.

4 CONCLUSION

In this article, we introduced a data collection system that ensures confidentiality, authenticity, and integrity of transmitted data. We use the WiFi technology and MQTT protocol to exchange application messages. Code is RIOT OS compatible, so it is portable to other supported devices. Measured data are stored and interpreted by OpenHAB service. That allows the creation of graphical interfaces, graphs, setting limits, sending notifications, etc. This particular implementation of the proposed system allows the user to monitor temperature and humidity in rooms. The system then warns the user when the limits, at which the spread of the disease can accelerate, are exceeded. The future goal of the project is the incorporation other devices into the system such as a humidifier or heating control.

REFERENCES

[1] MECENAS, BASTOS, VALLINOTO a NORMANDO. Effects of temperature and humidity on the spread of COVID-19: A systematic review [online]. 2020 [cit. 2021-03-07]. Dostupné z: https://doi.org/10.1371/journal.pone.0238339

[2] MQTT: The Standard for IoT Messaging [online]. [cit. 2021-03-23]. Dostupné z: https://mqtt.org/

[3] RIOT OS [online]. Berlin, 2008 [cit. 2021-03-07]. Dostupné z: https://www.riot-os.org/

[4] OpenHAB [online]. [cit. 2021-03-23]. Dostupné z: https://www.openhab.org/

[5] Kmackay / micro-ecc [online]. 2017 [cit. 2021-03-07]. Dostupné z: https://github.com/kmackay/micro-ecc

[6] Crypto [online]. [cit. 2021-03-07]. Dostupné z: http://api.riot-os.org/group__sys__crypto.html

IMPACT OF IMPLEMENTATION OF RES AND ELECTRO-MOBILITY ON ELECTRICAL PARAMETERS IN LV DISTRIBUTION NETWORK

Jan Vojtech

Bachelor Degree Programme (3), FEEC BUT

E-mail: xvojte11@stud.feec.vutbr.cz

Supervised by: Michal Ptacek

E-mail: ptacekm@vut.cz

Abstract: This paper is focused on the evaluation of electrical parameters in low voltage distribution network for proposed scenarios based on the future strategic plans of European Union and some Czech National Action Plans. The results of the analysis demonstrate the influence of implementation of renewable sources and electromobility to voltage, voltage drops, current loads etc.

Keywords: distribution network, smart grids, electromobility, renewable sources, charging stations

1 INTRODUCTION

European Union (EU) has long term target in decreasing greenhouse gasses emissions with increasing effort to reach climatic neutrality before 2050. To reach the goal, innovations must be done in the whole electric energy process including generation, transfer, and consumption. The biggest priority is dramatical decreasing emissions in traffic, because it is the biggest source of air pollution in cities. Development of electric vehicles (EV) and decentralised sources (DECE) is tightly connected with demands on projecting and controlling of distribution network (DN).

2 NATIONAL ACTION PLANS OF CZECH REPUBLIC

National action plans (NAP) are established with reference to EU plans and strategic planning of Czech Republic (CZE). These NAP are focused on suggested plans in detail to reach targets set in CZE and are regularly updated to correspond with new and more strict demands of EU. The most important measures for DN are in renewable sources (RES) and electromobility sectors because of their potential impact on reliable and safe management of DN.

2.1 RENEWABLE ENERGY

National Renewable Energy Action Plan [1] contains plans until 2020 without future. So, it is better to look on plans in this sector from The National Energy and Climate Plan of the Czech Republic [2]. This document is expecting 22 % share of RES on gross final electrical energy consumption, which is rise of 9 percentage points compared to 2020.

2.2 CLEAN MOBILITY

National Action Plan for Clean Mobility (NAPCM) [3] was updated in April 2020 and contains Green Deal targets from 2019 [4], which means reaching climatic neutrality before 2050. NAPCM sets EV as the main tool for reaching clean mobility, even though it is remarked that it is not the only and universal usable solution. Compared to other technologies are EV usable nowadays which is their main advantage. However, EV development is limited by sparse charging station networks.

But higher EV sales will probably be the cause of new and more strict emission limits. It is expected that there will be from 220 000 to 500 000 EV in CZE in 2030 which means 3-7 % of all cars in CZE.

2.3 SMART GRIDS

National Action Plan for Smart Grids (NAPSG) (updated 2019) puts a huge importance on development of smart grids (SG). Reaching full smart grids concept can help distribution network operators (DNO) to lower costs of network development. SG can be characterized in short as safe, reliable, and automated network with real-time control ability. However, RES should be developed mostly by DECE with real-time measures and remote-control ability according to NAPSG. [5]

According to a study mentioned in NAPSG [6], EV should be charged by slow chargers from low voltage (LV) network at small consumers, i. e. home small consumer (HSC) and workplace small consumer (WSC). Private charging station networks should consist of 95 % chargers with output up to 11 kW and 5 % chargers with output of up to 22 kW.

3 DNO TERMS

DNO could determine conditions for connecting new sources and consumption points in its DN area. These conditions partly correspondent with Energy Regulatory Office conditions. Nowadays new conditions of EG.D, a.s. company are discussed (but not public yet). Latest version of conditions contains these conditions:

1. EV charger with power input up to 100 kW must be capable of 1 level power regulation (on/off), EV charger which power input up to 250 kW must be capable of 4 level power regulation and EV charger which power input above 250 kW must be equipped with voltage measuring and active and reactive power measuring.

2. DECE with power output up to 100 kW must be capable of active power regulation by relay (on/off), DECE with power output up to 630 kVA must be capable of continuous output regulation of active and reactive power and must be equipped with voltage measuring, active and reactive measuring and selected DECE must be equipped also with other measuring like temperature, wind speed and sun exposure.

4 DN DESCRIPTION, SCENARIO INTRODUCTION AND ANALYSIS

4.1 DN DESCRIPTION

For demonstration of the impact of RES and EV chargers implementation into LV DN was chosen DN in Jinačovice village with 779 residents in Brno-venkov district. Basic typology with analysis is on the Fig. 1. Distribution lines consists of cable lines only. Most of cable lines were built in the 1980s and 1990s. The most frequent types of cables in DN are AYKY 3x185+95 and AYKY 3x120+70. DN is powered by 6 distribution transformer stations (DTS) with nominal power 4x 630 kVA and 2x 400 kVA. DS 338 consumers, therefrom 294 HSC and 44 WSC. DN also includes 13 photovoltaic systems (PVS) in power range from 2,2 kW to 20 kW. No EV chargers with power input 3,7 kW or more are recorded.

4.2 SCENARIO INTRODUCTION

Chosen scenarios represent different implementation levels of DECE and EV chargers. Target of chosen scenarios is to analyse the impact of the implementation level on electrical parameters of DN including recommendations for DNO. The following scenarios were analysed:

A. Current state 1 – dominant consumption of HSC, low WSC electricity consumption, PVS are not considered, EV chargers are not considered

B. Current state 2 – low consumption of HSC, dominant WSC electricity consumption, maximum power output of present PVS, EV chargers are not present

C. PVS – DN model contains evenly distributed new 7 kW PVS (long-term average according to DNO private statistics) covering 22 % of village consumption – low HSC electricity consumption, dominant WSC electricity consumption, maximum power output of present and new PVS, EV chargers are not considered

D. EV – DN model contains new 11 kW consumption points (average power of chargers on market), number of new consumption points is equal to 7 % of vehicles in the village – dominant HSC electricity consumption and EV chargers, low WSC electricity consumption, PVS are not considered

E. Concentrated PVS – PVS from scenario C power one (DTS)

F. Concentrated EV – all EV from scenario D are powered from one DTS

DAISY Bizon Projektant was chosen as software analysis tool for demonstration of the impact of implementation level of PVS and electromobility. This software analysis tool is commonly used for basic steady state calculations of DN by DNO (EG.D, a.s. company).

4.3   ANALYSIS AND RESULTS

This paper includes scenario A analysis. The left part of Fig. 1 displays voltage levels in DN, the lowest voltage levels are represented by dark green colour. The right part of Fig. 1 displays current load (CL) related to nominal current load, the highest RCL is represented by green colour. Lowest CL in this analysis is 87 % and lowest voltage level in DN is 360 V. Related power load of DTS varies from 29 % to 68 %.
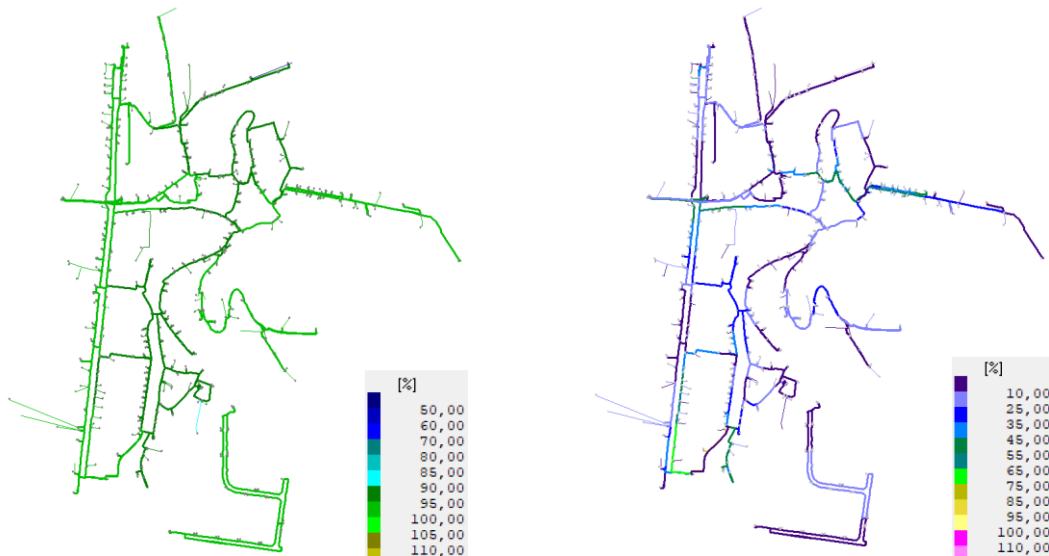
Fig. 1: Scenario A analysis

The calculations of current state show, that during extreme load of DN voltage level can drop to nominal voltage tolerance limit $\pm$ 10 %. This state can be improved by replacing current cable lines with new cables with larger diameter in parts of DN with highest RCL and by adding new cable to make cable line from DTS to consumer with low voltage level shorter. Nominal power of current DTS is sufficient.

5   CONCLUSION

Current state of DN was analysed and will be used as default state for future calculations of DN. Results shown it is necessary to improve DN before the implementation of new EV chargers. Oth-

erwise, voltage level in some parts of DN could drop under voltage level tolerance limit. Further analysis will be focused on proving DN in extreme conditions. In connection with future results new and larger measures will be recommended.

ACKNOWLEDGEMENT

REFERENCES

[1] National Renewable Energy Action Plan [Národní akční plan pro energii z obnovitelných zdrojů] [online]. Prague: Ministry of Industry and Trade, 2015 [cit. 2020-11-09]. Available from: https://www.mpo.cz/assets/dokumenty/54909/62718/649151/priloha001.pdf

[2] The National Energy and Climate Plan of the Czech Republic [Vnitrostátní plán České republiky v oblasti energetiky a klimatu] [online]. Prague: Ministry of Industry and Trade, 2019 [cit. 2020-12-16]. Available from: https://www.mpo.cz/cz/energetika/strategicke-a-koncepcni-dokumenty/vnitrostatni-plan-ceske-republiky-v-oblasti-energetiky-a-klimatu--252016/

[3] Updated National Action Plan for Clean Mobility [Aktualizace Národního akčního plánu pro čistou mobilitu] [online]. Prague: Ministry of Industry and Trade, 2019 [cit. 2020-11-16]. Available from: https://www.mdcr.cz/getattachment/Media/Media-a-tiskove-zpravy/Vlada-schvalila-aktualizovany-Narodni-akcni-plan-c/Aktualizace-NAP-CM.pdf.aspx

[4] A European Green Deal [online]. Brussels: European Commission, 2019 [cit. 2021-03-08]. Available from: https://eur-lex.europa.eu/resource.html?uri=cellar:b828d165-1c22-11ea-8c1f-01aa75ed71a1.0002.02/DOC_1&format=PDF

[5] National Action Plan for Smart Grids 2019 – 2030 [Národní akční plán pro chytré sítě 2019 – 2030] [online]. Prague: Ministry of Industry and Trade, 2019 [cit. 2020-12-16]. Available from: https://www.mpo.cz/assets/cz/energetika/strategicke-a-koncepcni-dokumenty/narodni-akcni-plan-pro-chytre-site/2019/9/Aktualizace_NAP_SG_2019-2030.pdf

[6] Partial Study for Work Team A25 – Prediction of Electromobility Development [Dílčí studie pro pracovní tým A25 - Predikce vývoje elektromobility v ČR] [online]. Prague: Euroenergy, spol. s r.o. [cit. 2020-12-30]. Available from: https://www.mpo.cz/assets/cz/energetika/strategicke-a-koncepcni-dokumenty/narodni-akcni-plan-pro-chytre-site/2019/10/Studie-NAP-SG-A25_Elektromobilita.pdf

SIGNAL PROCESSING FOR VOCAL RECOGNITION OF STURNUS VULGARIS

Jana Lázničková

Bachelor Degree Programme (1), FEKT VUT

E-mail: xlazni09@stud.feec.vutbr.cz

Supervised by: Petr Marcoň

E-mail: marcon@feec.vutbr.cz

Abstract: This paper describes the issue of sturnus vulgaris detection in the vineyards in order to scare these animals more effectively. The analysis and classification of bird singing is difficult because many problems can appear. One of the problems is background noise e.g. sounds of cars, trees, and also the singing of various bird species at once. Another problem is different types of bird songs. For example, an alarm melody, search for food, and also communication between the birds during a flight. This article presents a solution to one of these problems in case when only audio recordings are available.

Keywords: sturnus vulgaris, raspberry, MFCC, python, spectrogram

1 INTRODUCTION

Starlings in the vineyards cause significant damage every year. Winegrowers use several scaring devices to reduce the invasions of starlings. However, devices on the market often disturb surroundings with excessive noise. There is an idea to scare starlings before they land in the vineyards, therefore, their presence would have to be detected by vision (camera) or noise (microphone).

This paper is focused on the creation of a *Mel-spectrogram*, which facilitates the subsequent use of CNN (convolutional neural network) to detect the singing of starlings. Many papers using robust methods of deep learning to identify the sounds of birds have been published. To increase the sensitivity of the detection, the recorded audio must be pre-processed.

Humans perceive only a narrow range of frequencies and have a poor perception of the linear scale. Moreover, people perceive differences at lower frequencies more effectively than at higher frequencies. To better visualize differences at higher frequencies, the conversion into the *Mel-scale* is made according to the literature [1].

$$f_{mel} = Mel(f) = 2595\log_{10}(1 + \frac{f}{700}) \tag{1}$$

2 ALGORITHM FOR DATA PREPARATION FOR CNN

The algorithm using freely available libraries of the python programming language is used for sound detection. The algorithm has been modified in order to filter out the interfering components to facilitate the detection of the starling sound. This algorithm is implemented in the Raspberry PI4 (RPi4) microcomputer. This microcomputer was chosen because of the required computing power and also because it is easy to mount, to transfer, and it supports real-time data transfer well.

The recorded birdsong is sampled and then transferred from the time domain to the frequency domain using the Fourier transform (FT). The result of the Fourier transform is the spectrum of the signal (spectrum = expression of amplitudes and phases of individual harmonic components in the frequency domain). A Short-time Fourier transform (STFT) is applied to this spectrum and the result

of this transformation is a spectrogram. The Matplotlib library is used to plot the spectrogram of the sampled signal. STFT is provided by Librosa library.

The spectrogram is depicted for 10 seconds, so as much useful information as possible from the sound recordings is retained. After that, a *Mel-spectrogram* is created from this spectrogram. The *Mel-spectrogram* is a spectrogram where the y-frequency axis is converted to a *Mel-scale*. A high-pass filter was used to remove low-frequency noise [3] because birds sing at high frequencies. The required frequencies from 1400 Hz is transmitted.

Mel-Freqeuncy Cepstral Coefficients (MFCC) [1] are coefficients that make up the *Mel-Frequency Cepstrum* MFC. These coefficients are acquired by applying discrete cosine transform (DCT) to the *Mel spectrogram.* The plotting of the *Mel-spectrogram* and the calculation of MFCC coefficients allow creating the CNN more effectively. The transfer of the data from a *Mel spectrogram* to CNN requires all of them to be at the same size. The input signal can be received by RPi4 in real-time, and also with a directional microphone. Similar solutions are known, e.g. for the detection of other species of birds [4] or bats [5].

The entire process of audio recording processing can be described by one block diagram:
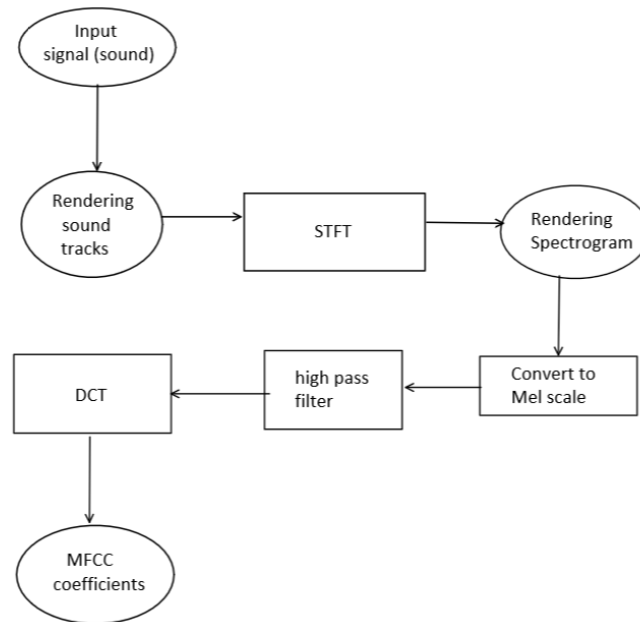
Figure 1:   A block diagram of audio recording

3   ASSESSMENT OF MEASURED SIGNALS

Figures 2 to 4 show the results of the analysis of the starling song. Figure 2 shows only the monophony of a starling. The linear frequency spectrogram of its sound is vizualized in Figure 3. Figure 4 shows the Mel spectrogram, which is based on a linear frequency spectrogram. This spectrogram is converted to a Mel-scale, and then a high-pass filter is applied to it. Figure 5 shows the spectrogram of a golden eagle for comparison with a starling. The difference between spectrograms 4 and 5 is obvious.
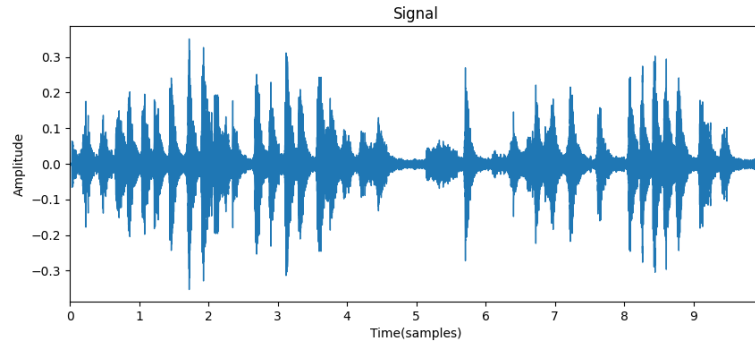
Figure 2: The audio recording of sturnus vulgaris.
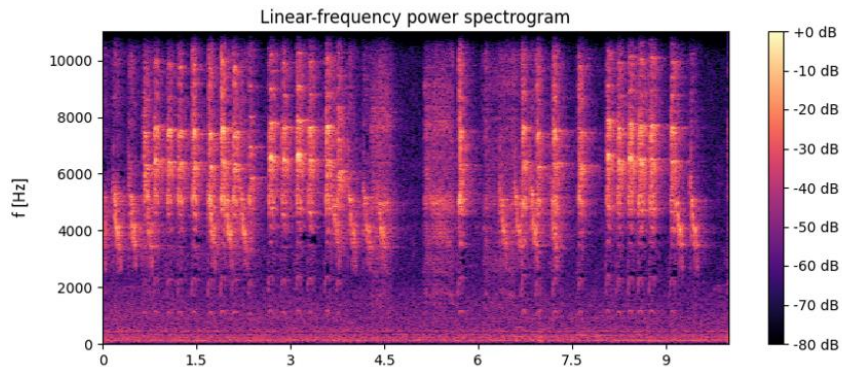
Figure 3: The spectrogram of sturnus vulgaris (linear scale).
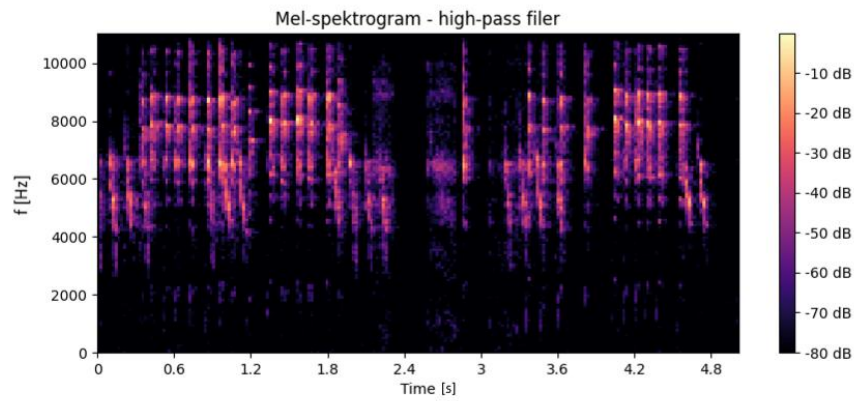
Figure 4: The *Mel-spektrogram* of sturnus vulgaris.
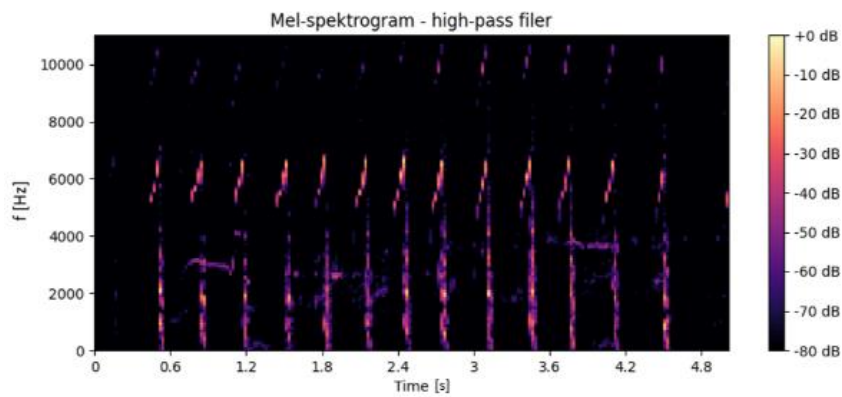
Figure 5: The *Mel-spektrogram* of golden eagle.

4 CONCLUSION

The paper describes processing of birds sound recordings to increase the sensitivity of detection algorithms. A filter filtering out the low-frequency noise was introduced in the process. The conversion of a linear spectrogram to a *Mel spectrogram* and the subsequent calculation of MFCC coefficients is necessary to smoothly follow the CNN [6] to detect the sound of starlings in the vineyards. More work on the project will follow. In terms of software, CNN will be used and in terms of hardware, RPi4 will be attached to the rotating mechanism, which will be controlled by a microcomputer and according to the directional microphones will identify not only the sound but also the position of the source of the sound.

ACKNOWLEDGEMENT

REFERENCES

[1]   Leland, R.: Understanding the Mel Spectrogram. In: Analytics Vidhya [online]. [cit. 2021-02-23]. Available from: https://medium.com/analytics-vidhya/understanding-the-mel-spectrogram-fca2afa2ce53.

[2]   Pollák, P.: Kepstrální analýza řečového signálu: Detekce řečové aktivity. 2016. Přednáška. ČVUT.   Available from: https://moodle.fel.cvut.cz/course/view.php?id=618

[3]   Svozil, M: Statistické zpracování řečových parametrů. Brno, 2015. Diplomová práce. Vysoké učení technické v Brně. Vedoucí práce Ing. Miroslav Staněk.

[4]   Kortas, M.: Sound-based bird classification [online]. In: Towards data science, 2020 [cit. 2021-02-25].   Available   from: https://towardsdatascience.com/sound-based-bird-classification-965d0ecacb2b.

[5]   Meertens, R.: Detecting bats by recognising their sound with Tensorflow [online]. In Pinch of Intelligence (Blogging artificial intelligence, 2. srpna 2017 [cit.2021-01-25]. Available from: http://www.pinchofintelligence.com/detecting-bats-recognising-sound-tensorflow/).

[6]   Janousek, J, Marcon, P., Pokorny J., Mikulka J.: Detection and Tracking of Moving UAVs. In: PIERS [online]. IEEE, 2019, s. 2759-2763 [cit. 2021-02-02]. ISBN 978-1-7281-3403-1.

FORMULA STUDENT CONTROL UNIT AND POWER SUPPLY SYSTEM

Dominik Klement

Bachelor degree Programme (3), FEEC BUT

E-mail: xkleme13@stud.feec.vutbr.cz

Supervised by: Miloslav Steinbauer

E-mail: steinbau@feec.vutbr.cz

Abstract: This article describes the development of an electric vehicle control unit (VCU) used in a Formula Student competition. The VCU processes data from sensors located on the vehicle, evaluates this data, and then controls peripherals. The power supply system, integrated in the VCU, regulates the input voltage to the branches that supply other units on the vehicle.

Keywords: Vehicle Control Unit, Formula Student Electric, EV, Electric Vehicle, TU Brno Racing, ECU

1 INTRODUCTION

This article describes the development of an electric vehicle control unit (VCU) used in a Formula Student competition. The VCU processes data from sensors located on the vehicle, evaluates this data, and then controls peripherals. The power supply system, integrated in the VCU, regulates the input voltage to the branches that supply other units on the vehicle.

Formula Student (FS) is a European branch of the originally American Formula SAE competition. It is a prestigious competitions of university teams composed of students seeking bachelor's and master's degree. The beginnings of the competition dates back to 1981. It got to Europe 17 years later. However, races are not only held on these continents. In addition to selected US and European countries, the competition is also held in Brazil, Japan, India and Australia. In total, over 800 teams from all over the world take part in the races [1].

Figure 1: Formula Student car Dragon 9 from team TU Brno Racing

2 CONTROL UNIT REQUIREMENTS

The control unit has the task of creating power buses for all peripherals in the car, sensors for measuring variables such as pressure and temperature required for proper operation of the cooling

system, communicating with an external inertial unit, executing ESP algorithm, and controlling tractive system.

2.1 POWER SUPPLY SYSTEM

On the first electric race car from this team, the control unit was designed as a modular system. Each unit in the car regulated the necessary voltage for its operation internally. Therefore, it was more difficult to diagnose a possible error and it led to a multiplication of incompatible components. The aim of this article is to gradually unify as many components as possible, thereby reducing weight and simplifying the production of the wiring harness. Each unit will therefore only contain an identical input protection module, thus preventing the multiplication of unique components.
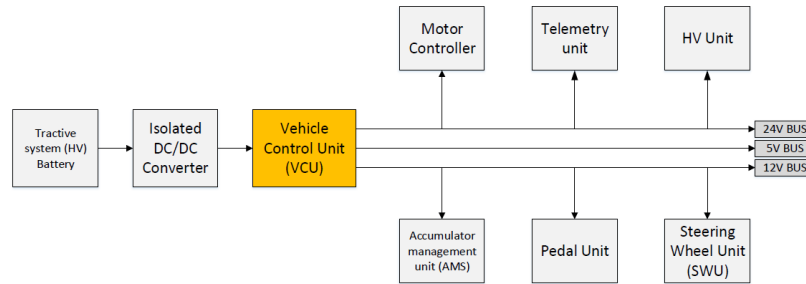
Figure 2: Block diagram of the proposed solution

2.2 RELIABILITY

The control unit must withstand undervoltage, overvoltage, electrostatic discharge, polarity reversal of the power input, short-circuit of any output to ground during a fault situation, it must withstand a wide temperature range, from negative temperatures during winter testing to temperatures reaching +70 °C during races taking place in peak summer.

2.3 INPUTS AND OUTPUTS

The unit oversees control of individual parts of the car, reading sensors, communication with other devices on three CAN buses and a pair of RS232 ports, regulating the voltage for the entire car and recording data to internal storage for further vehicle analysis. These requirements are summarized in Table 1.

Category	Type	Amount	Requirements
Input	Wheel speed sensor	4	12 V compatible
Input	Digital	4	24 V compatible
Input	Analog	6	12 V compatible
Output	Power	3	24 V
Output	Power	2	12 V
Output	Power	1	5 V
Output	Digital	8	24 V compatible
Output	Analog	2	0–10 V
Communication	CAN 2.0 B	3	1 Mbit/s
Communication	RS 232	2	1 Mbit/s

Table 1: Inputs and outputs

The Vehicle Control Unit is controlled by a dual-core microcontroller from the STM32H7 series. The more powerful ARM Cortex M7 core clocked at 480 MHz is used to run the traction control algorithm and optimization of torque of the individual tractive motors. The ARM Cortex M4 core

clocked at 240 MHz controls peripherals, reads sensors and records data. The block diagram in Figure 3 shows the internal wiring of the subsystems in the unit [2].
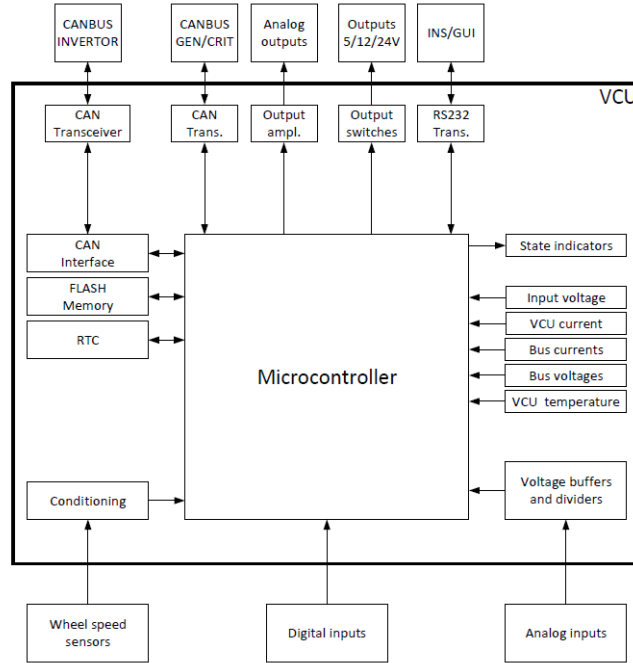
Figure 3: VCU block diagram

3   DESIGN

Altium Designer 20 was used to design the schematics and printed circuit boards. This program allows hierarchical grouping of schematics and thus streamlining the design, the aim was to create a diagram that will be easy to expand upon in the future with additional inputs and outputs as needed.

The final design of the printed circuit board is systematically divided into two regions, the power, where the necessary voltages for the operation of the entire vehicle are regulated, and the signal, where the microcontroller processes signals and controls peripherals. The printed circuit board shown in Figure 4 has a 6-layer stack up, connected to it is a daughter board facilitating interface connectors.
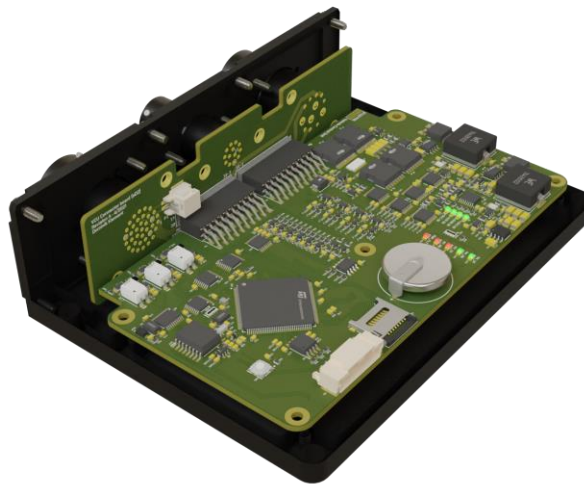
Figure 4: 3D render of the assembly

An enclosure has been designed for the printed circuit board to protect it from damage and to provide cooling to the power components. This cover will be milled from EN AW-6061 alloy. Figure 5 shows a rendered model designed in CAD software Solidworks.

Figure 5: Vehicle Control Unit enclosure

4 CONCLUSION

The result of this article is a block diagram of a power supply of Formula Student Electric vehicle, design of hierarchical schematic of the vehicle control unit including connection of individual inputs, outputs, communication interfaces, input protection, design of power supply system, design of 6 layer PCB and aluminium enclosure.

ACKNOWLEDGMENT

REFERENCES

[1] TU Brno Racing. Formule Student. *TU Brno Racing.* [Online] https://tubrnoracing.cz/en/1746-2/formula-student//.

[2] STMicroelectronics. High-performance and DSP with DP-FPU. *ST life.augmented.* [Online] https://www.st.com/en/microcontrollers-microprocessors/stm32h745zi.html.

METABOLITE GENOME-WIDE ASSOCIATION STUDIES OF ARABIDOPSIS THALIANA

Jana Schwarzerová

Master Degree Programme (2), FEEC BUT

E-mail: xschwa16@stud.feec.vutbr.cz

Supervised by: Wolfram Weckwerth, Benjamin Ramberger, Karel Sedlář

E-mail: wolfram.weckwerth@univie.ac.at, benjamin.ramberger@technikum-wien.at, sedlar@feec.vutbr.cz

Abstract: Current research based on the edge of bioinformatics and ecology engineering has huge potential due to combination of laboratory analyses and advanced bioinformatics algorithms. The paper deals with a combination of GC-MS and LC-MS based metabolomic analysis for identification and quantitation of metabolites in environmental perturbations with advanced bioinformatics approach of metabolite genome-wide association studies (mGWAS). This complex view is applied to *Arabidopsis thaliana*. The main goal is to obtain genetic predictions focused on *A. thaliana* under different environmental conditions. Currently, important ecological issues such as climate change, pollution etc. have impact on the change of environment. It has a great effect on plants which serves as producers of oxygen or food. While simple observation reveals only a phenotype change, changes in genotypes of organisms can be captured using mGWAS and further utilized in industrial ecology and biotechnology.

Keywords: Metabolomics, Systems biology, Ecology, Single nucleotide polymorphism, Genetic prediction

1 INTRODUCTION

The next generation genetic prediction [1] is based on genome-wide association studies (mGWAS). It is a bioinformatics method for observing variant of genes in whole genomes. The mGWAS approach investigates the relationship between genetic factors and metabolome. Although some tools for genomic prediction exist [2][3][4], a huge part of applications rrBLUP for data analysis of ecology ecologically important producers is missing.

In general, metabolites represent the ultimate response of biological systems. Thus, metabolomics is considered to be the link between genotypes and phenotypes. So far, research of mGWAS is focused on genetics of the human metabolome but not on the most important current issues connected with ecological problems [5] such as climate change, pollution, environmental degradation as deforestation of rainforests, etc. All these problems impact the change of environment in the near future. As the environment is reflected in phenotypes, mGWAS can uncover origin of observed changes on molecular level, i.e., in genotypes of various organisms in industrial ecology and biotechnology. It includes plants as a food source or their role of helping to maintain the stability of the environment. Thanks to the mutations captured within computational analysis by the means of bioinformatics and systems biology, we could predict upcoming genotype changes and using synthetic biology suppress or completely avoid adverse effects or support changes leading to desirable phenotypes.

2 MATERIALS AND METHODS

The mGWAS approach has been used to investigate the relationship between genetic factors and metabolome for *A. thaliana* depending on two different environmental conditions during cultivation. The paper presents methodology which merges two main branches in historical development

of systems biology [6]. These branches include wet lab experiments studying metabolomic regulations and *in-silico* analysis for description, prediction, and simulation of metabolic networks.

2.1 PLANT MATERIAL AND HARVEST

Dataset of *A. thaliana* were cultivated under two different temperature-related conditions, 6°C and 16°C. In study by Weiszmann et. al. [7], natural variation of growth rates of *A. thaliana* was monitored together with dynamics of primary metabolites under moderate (16°C) and low (6°C) temperature. Chemicals, plant material and harvest were described in the study by Doerfler et. al. [8]. Samples of *A. thaliana* plants Col-0 (wild type) was cultivated under controlled conditions. Dataset were obtained by Gas chromatography coupled to mass spectrometry (GC-MS) and liquid chromatography coupled to mass spectrometry (LC-MS). GC-MS analysis protocol was used according to Weckwerth et. al. [9]. LC-MS analysis is described in the study by Doerfler et al. [8].

2.2 RIDGE REGRESSION AND OTHER KERNELS FOR GENOMIC SELECTION (RRBLUP)

Currently, one of the best methods for genomic prediction of breeding values is based on ridge regression (RR). RR is equivalent to the best linear unbiased prediction (BLUP) if the genetic covariance among lines to observations is proportional to their similarity in genotype space. To facilitate the use of RR and non-additive kernels in plant breeding, a new software package for R called rrBLUP has been developed [4].

The basic rrBLUP model [4] is

$$y = WGu + \varepsilon, \tag{1}$$

where $u \sim N(0, I\sigma^2)$ represents normal distribution in a vector of marker effects, where mean is zeroes and variance is σ^2, G is genotype matrix for biallelic single nucleotide polymorphisms (SNPs) and W is the design matrix relating lines to observations y.

In this study the basic mGWAS function from package R/ rrBLUP [4] was used for a genome-wide association analysis. Calculates maximum-likelihood solutions for mixed models of the form:

$$y = X\beta + Zg + S\tau + \varepsilon, \tag{2}$$

where X, Z and S are incidence matrices. X is of $n \times m$ size of with unphased genotypes for n lines and m biallelic markers, coded as $\{-1,0,1\}$. Z is the matrix relating observations to lines in the training set. β is a vector of fixed effects that can model both environmental factors and population structure. The variable g models the genetic background of each line as a random effect. The variable τ models the additive SNP effect as a fixed effect.

3 RESULTS

The first analysis represents pre-processing, in which the metabolomic dataset was checked. In the next step, mGWAS was applied. The dataset was visualized by Principal Component Analysis (PCA), see Figure 1 A. The presumption was to create clusters of samples gathered under the same conditions. This assumption was met. According to Kolmogorov-Smirnov test, the null hypothesis is rejected and there is an evidence that the data tested are not normally distributed. For further analysis the dataset was normalized using log-transformation as shown in Figure 1 B.

In the following analysis mGWAS was performed based on the mixed model. The algorithm which was used is the best linear unbiased prediction. The most frequent metabolite which had statistical signification in 16°C condition was Galactinol, see Figure 2.
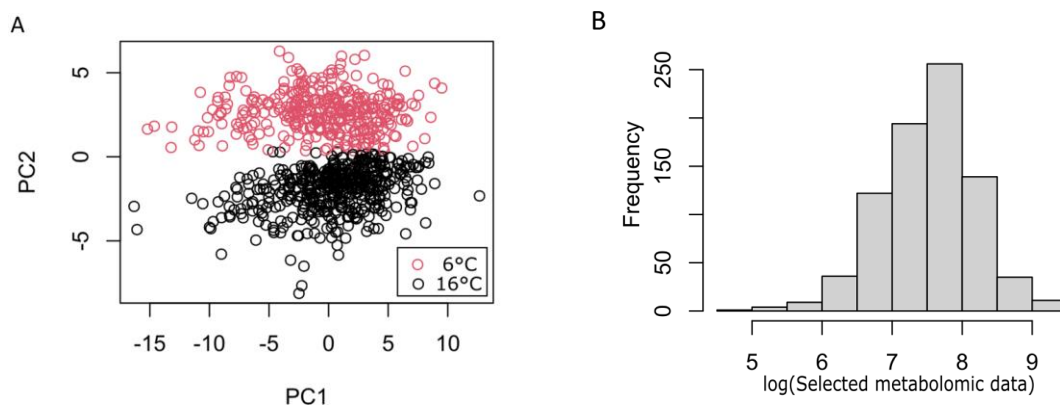
Figure 1: Pre-processing of the data. The left panel (A) shows the PCA of the dataset and the right panel (B) shows a histogram of the normalized dataset using log-transformation
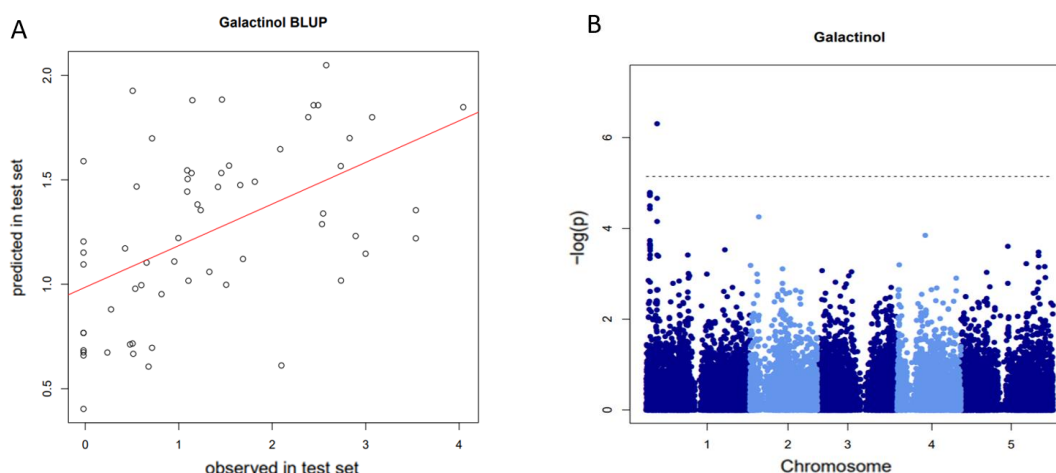
Figure 2: The Galactinol BLUP plot represents the predicted model (A) using rrBLUP methods and Manhattan plot (B) is the final mGWAS based on the mixed model (B) where is shown a significant change on the first chromosome. Each dot represents a SNP. The light blue is even number of chromosome, dark blue is odd number of chromosome.

Thanks to mGWAS of *A. thaliana,* the metabolites, which cause the changes genome according to different temperature conditions, were identified. The three most of significant metabolites are Butanoic, Glutamic acid and Putrescine in 6°C and the three most of significant metabolites are Asparagine, Fumaric acid and Galactinol in 16°C. All these metabolites have a key role during plant life.

4   DISSCUSION AND CONCLUSIONS

In recent years, an increasing number of SNPs arrays and DNA re-sequencing clarified the majority of the genotypic space for a number of organisms, including human, maize, rice, and *A. thaliana* [10]. mGWAS presents a powerful tool to reconnect this trait back to its underlying genetics prediction based on metabolite. mGWAS can offer a valuable first insight into trait architecture or help finding candidate loci for subsequent validation. Once such genetic markers are identified, they can be used to understand how genes contribute to properties of organisms, for example as growth behavior in plants. Now it has huge potential for prediction in the near future which will

be affected by climatic changes. At the core of the rrBLUP package is the function mixed.solve, which can be used to solve the marker- based versions of the genomic prediction problem. We used this pipeline and applied mGWAS based on the mixed model to a dataset of *A. thaliana* cultivated according two different temperature-related condition, 6°C and 16°C. These conditions can simulate climate change and global warming. It is known that photosynthesis needs to be tightly linked to carbohydrates and primary metabolism in order to sustain growth and development, it remains unclear how natural variation of primary metabolism relates to growth rates [7]. Thanks to these lab experiments and mGWAS, the metabolites causing the changes of genome were identified. In 6°C, the three most of significantly metabolites are Butanoic, Glutamic acid and Putrescine but in 16°C the three most of significantly metabolites are Asparagine, Fumaric acid and Galactinol.

ACKNOWLEDGEMENT

REFERENCES

[1]   Montemayor, D., Sharma, K. mGWAS: next generation genetic prediction in kidney disease. Nat Rev Nephrol 16, 255–256 (2020). https://doi.org/10.1038/s41581-020-0270-0

[2]   GOGARTEN, Stephanie et al. GWASTools: an R/Bioconductor package for quality control and analysis of genome-wide association studies. 2012 doi:https://doi.org/10.1093/bioinformatics/bts610

[3]   ARABFARD, Masoud et al. Genome-wide prediction and prioritization of human aging genes by data fusion: a machine learning approach. 2019 doi:https://doi.org/10.1186/s12864-019-6140-0

[4]   JEFFREY, Endelman. Ridge Regression and Other Kernels for Genomic Selection: Package 'rrBLUP' [online]. 2019 Available from: doi:10.3835/plantgenome2011.08.0024

[5]   LEVIN, Simon A. The Problem of Pattern and Scale in Ecology: The Robert H. MacArthur Award Lecture [online]. 1992 Available from: https://doi.org/10.2307/1941447

[6]   IDEKER, Trey, Timothy GALITSKI a Leroy HOOD. A NEW APPROACH TO DECODING LIFE: Systems Biology. 2001 Available from: doi:10.1146/annurev.genom.2.1.343

[7]   WEISZMANN, Jakob et al. Plasticity of the primary metabolome in 241 cold grown Arabidopsis thaliana accessions and its relation to natural habitat temperature. doi:https://doi.org/10.1101/2020.09.24.311092

[8]   DOERFLER, Hannes, et. al. Granger causality in integrated GC–MS and LC–MS metabolomics data reveals the interface of primary and secondary metabolism. 2013 doi:10.1007/s11306-012-0470-0

[9]   WECKWERTH, Wolfram et. al. Process for the integrated extraction, identification and quantification of metabolites, proteins and RNA to reveal their co-regulation in biochemical networks. 2004 doi:10.1002/pmic.200200500

[10]   KORTE, Arthur a Ashley FARLOW. The advantages and limitations of trait analysis with GWAS: a review. doi: https://doi.org/10.1186/1746- 48-9-29

MRI OF RAT'S HEART AND T1 QUANTIFICATION

Jiri Vitous

Master Degree Programme (5), FEEC BUT

E-mail: xvitou03@stud.feec.vutbr.cz

Supervised by: Radovan Jirik

E-mail: jirik@vutbr.cz

Abstract: This paper is focused on developing a retrospective gating algorithm for rat heart MRI and T1 quantification. The first goal is to develop an algorithm binning acquired echo signals according to the corresponding phases of the cardiac and respiratory cycles. The echo signals are acquired using an inversion-recovery 2D gradient-echo sequence with radial encoding and golden-angle increment of the k-space trajectory in the read-phase directions. The next goal is compressed-sensing reconstruction of images. The third goal is fitting a T1-relaxation model to the reconstructed images to obtain quantitative T1 maps.

Keywords: rat, heart, cine, MRI

1   INTRODUCTION

Magnetic resonance (MR) cardiac imaging is quite challenging due to cardiac and respiratory motion. To obtain clean imaging some form of synchronization/gating is required. The standard clinical procedure usually incorporates prospective gating of acquisition with respect to a measured electrocardiography (ECG) signal. With some sequences, this is not easily done due to the induction of electric current in the ECG leads, which can make successful ECG gating harder or near impossible, especially with small-animal heart imaging and high-field MR systems. Because of these limitations, methods for retrospective self-gated acquisition have been proposed.[1] [2]

These methods are usually based on some form of navigator acquisition, either from within the imaged slice or from a separate navigator slice. This work is focused on MRI methods for quantification of myocardium T1 in small animals (mice, rats). For small-animal MRI, dedicated methods have to be used because of much faster cardiac and respiratory activity, compared to humans, and because of the impossibility of breath-hold acquisition typical for clinical cardiac MRI.

The images can be further processed into T1 relaxation maps, which can serve as an early marker of several diseases. We can use for example two T1 maps measured before and after contrast-agent administration to estimate the ECV (Extracellular fractional volume), which serves as a marker of fibrosis. [4]

For T1 quantification, 2 main methods are usually used: so-called VFA (Variable flip angle)[2] and IR (Inversion recovery)[1]. The VFA mapping is very sensitive to inhomogeneities of the B1 field (present especially in high-field MRI), because of this we focus on IR sequences.

To the author's knowledge, the only IR method for small-animal cardiac quantification of T1 maps with retrospective gating is 2D inversion recovery acquisition in [1], which cannot detect respiratory and cardiac phases close to the point when the magnetization crosses zero during its recovery towards the steady-state following the IR pulse. Furthermore, the method needs long waiting periods between the end of the pulse train and following inversion pulse, due to the usage of simple T1 quantification model and external respiratory synchronization to start the inversion cycle at the same respiratory phase which elongates the acquisition time.

The method proposed in this paper solves these two problems. It provides reliable cardiac and respiratory synchronization also around the zero-crossing points without the need for external sensors and since we use a more complex model, we can eliminate the waiting periods.

2 MATERIALS AND METHODS

The proposed method uses a 2D gradient-echo acquisition scheme with radial readout as in [1]. A navigator signal consisting of the first point of every FID (Free induction decay) signal is extracted from acquired data. Since we use an inversion recovery sequence the navigator depends not only on cardiac and respiratory motion but also on the recovery of the magnetization of blood and tissues to their steady state after being inverted by an IR pulse. This multi-exponential recovery trend has to be eliminated from the navigator signal first. To achieve this, the data points are reordered with respect to their position in the IR-IR interval (an interval between consecutive IR pulses) and fitted with a polynomial of the 8th degree, such a high degree is mandatory in order to cover the fast recovery at the beginning of the inversion period. Usually, no regularization is needed as this fit is performed on all data at once (roughly 80000 points for a 10-minute long acquisition). A set of simple bandpass FIR (Finite impulse response) filters is then applied to extract the base frequency bands of the cardiac and respiratory activities. The resulting cardiac and respiratory navigator signals are then fed through a one-way zero-cross detector to form timestamps marking each period. As mentioned in the introduction, the timestamps around the magnetization's zero-crossing points cannot be reliably estimated. To do this the mean cardiac and respiratory periods between acceptable timestamps are computed and the missing data are then interpolated assuming constant cardiac or respiratory rates. For respiratory navigation this serves as a starting point from which every interpolated point is then shifted to the largest local maximum of the original navigator signal, to compensate for irregularities in the breathing cycles and mark the inspiration phase. The interpolated heart timestamps are kept at the interpolated values since, the heart rate usually does not change much within the inversion cycle.

After this interpolation, the timestamps are used to bin the acquired echo signals according to their cardiac and respiratory phases. Binning the data with respect to the current length of the heart or breath period ensures that even if the heart rate changes, the same fraction of the period i.e. the same image is always reconstructed, which further reduces artifacts in the image. The binned echo signals are then used in compressed-sensing iterative image-sequence reconstruction, including regridding of the measured data points from the polar to the Cartesian system (required for radial sampling acquisition methods) and Total variation regularization applied both in the spatial as well as temporal (Inversion recovery) dimensions.[5]

Quantification of T1 is based on the Inversion recovery Look-Locker model, where one inversion pulse is followed by a train of readout pulses. This model includes the effect of the excitation pulses of the sequence.[6]

To validate the proposed methods a preliminary study on rats with induced diffuse myocardial fibrosis has been conducted. Four rats were examined in 2 time points, one before any treatment and one following 14 days of the fibrosis-inducing treatment. Two of the rats served as controls and two rats were induced myocardial fibrosis according to [3]. The animal model was implemented at the Department of Physiology of Faculty of Medicine at Masaryk University. The MRI examination took place at the Institute of scientific instruments of the Czech Academy of Science in Brno using the Bruker Biospec 94/30 9.4T NMR spectrometer. The procedure consisted a of series of scans focused on anatomical features, mainly the ejection fraction estimation and then the precontrast and postcontrast measurement of T1 relaxation maps using the proposed radial sequence. Contrast agent (gadopentetate dimeglumine [Magnevist]; Bayer-Schering AG, Berlin, Germany) was administered. The dose was computed according to the examined rat weight (0.2 mmol/kg) and was intravenously administrated as a bolus between the precontrast and postcontrast acquisitions. The pre- and post-contrast T1

measurements were done in a single short-axis slice with the following parameters: TR/TE= 7.5/1.7 ms, IR-IR interval: 11.25 s, matrix 128x128, adiabatic IR pulse. The resulting T1 maps were then used to calculate the ECV according to [4].

3   RESULTS

The proposed methodology was evaluated with respect to synchronization, perceived quality of the reconstructed T1 maps and the consistency of the resulting ECV values. The quality of synchronization can be supported by the analysis of the gating results. This analysis consisted of manual annotation of four 10-minutes long acquisitions acquired from healthy rats and comparison with the automatic detection of respiration. The result was an F1 score of 99.5 %. A similar evaluation of the cardiac gating was not possible because the cardiac navigator shape is not used for synchronization (only its base frequency), therefore a timestamp is guaranteed in every cardiac cycle, furthermore an ECG acquired during the acquisition would be too noisy, as expected for small animals and a high-field MR scanner.

The resulting T1 maps (example in Figure 1) showed a good spatial consistency with clear outlines of the cardiac structures and homogeneous myocardium areas. The pre- and post-contrast T1 values evaluated within hand-drawn myocardium regions of interest and for each examination of each animal were converted to ECV values. The boxplots in Figure 2 show the calculated ECV values for F (Fibrotic) and C (Control) subjects at the beginning of the experiment and 2 weeks after the treatment.

For fibrotic myocardium, the ECV is known to be higher than for a normal myocardium tissue. The significance of ECV increase was evaluated by paired single-sided t-test with a null hypothesis that the ECV values did not change and an alternative hypothesis that the ECV values with the fibrotic specimen have increased. The resulting p-value of the fibrotic group of $2.48 \cdot 10^{-120}$ strongly supports the alternative hypothesis, indicating that the proposed methodology can measure elevated ECV in fibrotic myocardium.

For a more sound conclusion, hematocrit (HCT) should be measured at each MR examination, as the ECV values, as calculated from the T1 maps, are weighted by a factor including HCT. Since the ECV values decrease with the healthy group, it is probable that the fibrosis must even strongly counteract this trend caused probably by animal growth and possibly HCT fluctuation.
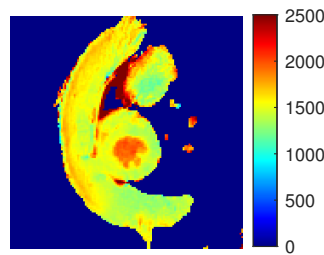
Figure 1: Example of one short axis slice of precontrast T1 map

4   CONCLUSION

We have proposed a set of acquisition and processing methods that might give reliable estimates of myocardial T1 and ECV which are important biomarkers characterizing fibrosis. The conducted preliminary study has shown significant changes in ECV in the fibrotic group ($p < 0.05$). More conclusive results will be obtained when more animals are included in the study, which is the aim of our follow-up work, together with including measurements of HCT.
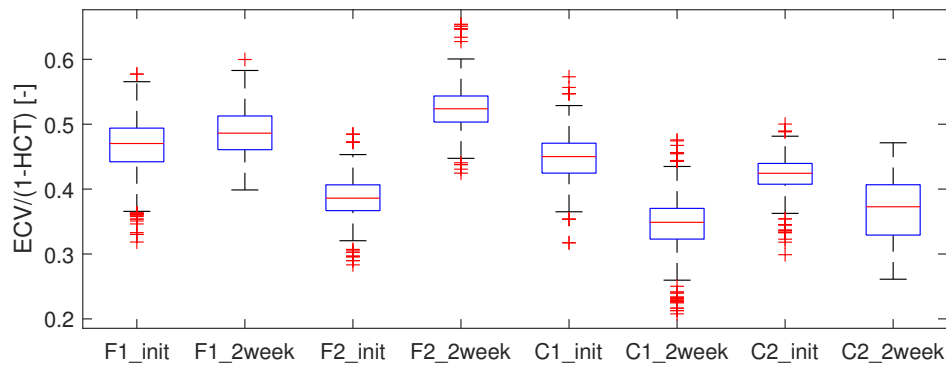
Figure 2: Resulting ECV/(1-HCT) values. F (Fibrotic), C(Control)

5 ACKNOWLEDGEMENT

REFERENCES

[1] WINTER, Patrick, Thomas KAMPF, Xavier HELLUY, Fabian T. GUTJAHR, Cord B. MEYER, Wolfgang R. BAUER, Peter M. JAKOB a Volker HEROLD. Self-navigation under non-steady-state conditions: Cardiac and respiratory self-gating of inversion recovery snapshot FLASH acquisitions in mice. Magnetic Resonance in Medicine [online]. 2016, 76(6), 1887-1894 [cit. 2021-03-09]. ISSN 07403194. Dostupné z: doi:10.1002/mrm.2606

[2] ZHANG, Xiaoyong, Bensheng QIU, Zijun WEI, et al. Three-dimensional self-gated cardiac MR imaging for the evaluation of myocardial infarction in mouse model on a 3T clinical MR system. PLOS ONE [online]. 2017, 12(12) [cit. 2021-03-10]. ISSN 1932-6203. Dostupné z: doi:10.1371/journal.pone.0189286

[3] FUJISAWA, Genro, Rodney DILLEY, Meryl J. FULLERTON a John W. FUNDER. Experimental Cardiac Fibrosis: Differential Time Course of Responses to Mineralocorticoid-Salt Administration. Endocrinology [online]. 2001, 142(8), 3625-3631 [cit. 2021-03-12]. ISSN 0013-7227. Dostupné z: doi:10.1210/endo.142.8.8339

[4] MEWTON, Nathan, Chia Ying LIU, Pierre CROISILLE, David BLUEMKE a Joăo A.C. LIMA. Assessment of Myocardial Fibrosis With Cardiovascular Magnetic Resonance. Journal of the American College of Cardiology [online]. 2011, 57(8), 891-903 [cit. 2021-03-12]. ISSN 07351097. Dostupné z: doi:10.1016/j.jacc.2010.11.013

[5] ADLURU, Ganesh, Chris MCGANN, Peter SPEIER, Eugene G. KHOLMOVSKI, Akram SHAABAN a Edward V.R. DIBELLA. Acquisition and reconstruction of undersampled radial data for myocardial perfusion magnetic resonance imaging. Journal of Magnetic Resonance Imaging [online]. 2009, 29(2), 466-473 [cit. 2021-03-12]. ISSN 10531807. Dostupné z: doi:10.1002/jmri.21585

[6] HENDERSON, Elizabeth, Graeme MCKINNON, Ting-Yim LEE a Brian K. RUTT. A fast 3D Look-Locker method for volumetric T1 mapping. Magnetic Resonance Imaging [online]. 1999, 17(8), 1163-1171 [cit. 2021-03-13]. ISSN 0730725X. Dostupné z: doi:10.1016/S0730-725X(99)00025-9

METHODOLOGY OF TIME-DEVELOPMENT ANALYSIS OF VERTEBRAL TUMORS IN CT DATA

Michal Nohel

Master Degree Programme (2), FEEC BUT
E-mail: xnohel04@stud.feec.vutbr.cz

Supervised by: Roman Jakubicek
E-mail: jakubicek@vutbr.cz

Abstract: This paper presents the methodology of time-development analysis of vertebral tumors in CT data. It including an overview of suitable features which can relevantly characterize the shape of tumor tissue. We proposed two different analysis methodologies: for compact tumors and the whole vertebral body. The test database of five lytic compact tumors containing five follow ups was created. The initial result of time-development for statistical features for compact tumors on created database and whole body vertebra were shown.

Keywords: time-development, CT, lesion, spine

1 INTRODUCTION

Bones are the third most commonplace of metastasis right after lungs and liver. Seventy percent of skeletal metastases cases originate from breast and prostate cancer. These metastases are the main cause of morbidity characterized by strong pain, worsened mobility, pathological fractures, hypercalcemia, and compression of the spinal cord. There are two types of skeletal metastases: osteolytic and osteoblastic (see Fig. 1). [1]

Time-development analysis generally means monitoring given features over time. In oncology, the analysis of the development of treated lesions using this method is still at the beginning of the research process. However, in clinical practice, it might be beneficial for evaluating how lesions respond to a particular treatment. If relevant features characteristic for the development of lesion over time were found, it would be possible to assess the effectiveness of that treatment earlier.

Figure 1:   An axial image example of osteoblastic lesion (left) and osteolytic lesions (right) in spinal CT data.

2   DESIGN OF EVALUATION APPROACH

2.1   SHAPE CHARACTERIZATION OF TISSUE

Tumor analysis is a very difficult task because it is first necessary to segment the tumors themselves. It can be segmented manually, which is very challenging or samiautomatically with using some method like region growing. Obtaining the annotation of the compact tumor might be a challenging task as tumors might grow and change their shape and characteristics over time (see Fig. 2). Often there is more than one tumor in one vertebra, separated tumors from the first scan might connect during the following scans, or one tumor might separate into many smaller ones (see the second row of Fig. 2). Due to this fact, the analysis of these tumors might be unrealistic and another method has to be used. Tumors (metastases) according to their shape we suggested dividing into compact tumors and complex tumors which affect the whole body of the vertebra. Thus, it is necessary to divide the realization of time-development of spinal tumors according to the size of the damage on the vertebra into the analysis of compact tumors and the analysis of the characteristics of the whole body of the vertebra.
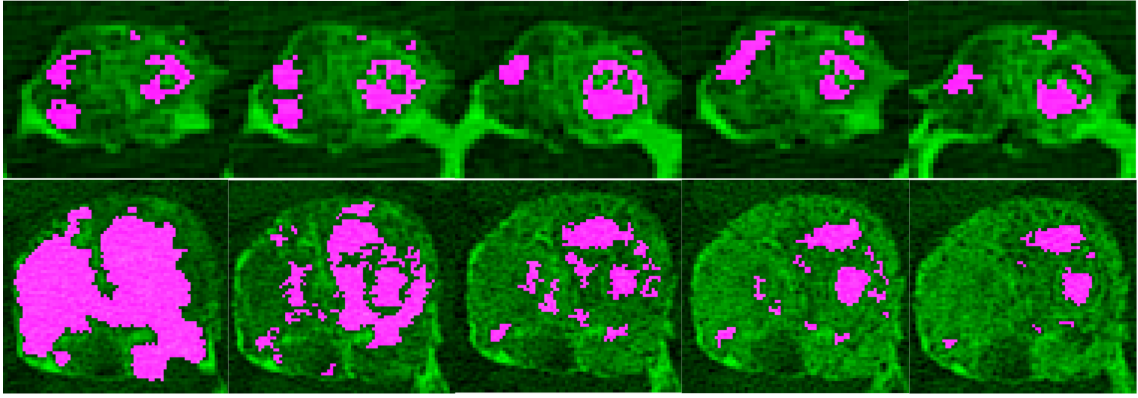
Figure 2:   Example of time-development treated lytic lesions from the third cervical vertebra (up) and fifth lumbar vertebra (down). It fused images of original data and created annotation, where the green color is non-lytic tissue and the pink color is lytic tissue. There was a gap of three months between the first three CT scans, a fourth scan was after the next five months, and a fifth after another four months.

2.2   THE ANALYSIS OF COMPACT LESIONS

This approach requires segmentation of the lesions which is a very challenging task because segmentation must be accurate and it is very time-consuming. If we manage to segment a compact tumor, various relevant features can be obtained from it. These features include statistical local features and features based on shape analysis. Statistical local features that are suitable for evaluating the time course of tumors include the mean value and the variance or standard deviation of the lesion intensity. Determining the relative values of these features, which are related to healthy trabecular tissue, also allows us to better assess how the density of tumor tissue changes.

Another group of features that can be evaluated for a given tumor is based on shape analysis. Since spinal tumors are 3D objects, it is convenient to choose descriptors for 3D image data. The most simple descriptor is the volume of the tumor. This might show changes in size over time, which may indicate a response to the treatment. However, normalization is required for this feature because the volume increases to the third power, and without some normalization, evaluation of the results would be difficult. Another suitable 3D descriptor is the roundness, which can, over time, show whether

the tumor is compact or whether it changes its shape and grows into the surrounding tissue. Other possible features include the surface, the center of gravity, the length of the main and the secondary axis, etc.

2.3 THE ANALYSIS OF THE WHOLE VERTEBRAL BODY

This approach requires segmentation of vertebra or localization of the center of body vertebra for next analysis. It could be less time-consuming because such accuracy is not required there and is suitable for very damaged vertebrae. Some vertebrae can be so extremely damaged by tumor tissue, that it makes shape analysis very problematic. In that case, another method of analysis might be more appropriate. One possible approach might be to calculate features from the whole vertebral body. Features that may be appropriate to extract from the entire vertebra include the mean and especially the variance of intensity. Assuming a severely affected vertebra with a lytic tumor (Fig. 2 below), if the tumor is being treated, which means that calcium is deposited to the destination of the tumor, the overall variance of intensity in the vertebra should increase. This parameter might indicate the time-development of tumors even for very severely affected vertebrae, in which compact tumors can't be segmented. [1]

Another possible approach might be the usage of texture analysis. This analysis could detect unique primitives that could be characteristic for the tissue which is going through a remodelation with calcium storage. This would detect the response of the lesion to the treatment in the time-development analysis.

3 REALIZATION AND DISCUSSION OF METHODOLOGY

3.1 DATA

In this work, an anonymized database of an oncologically ill patient examined by CT scan was used. The data were obtained through Philips Healthcare Brilliance iCT 256-slice scanner with experimental CT protocol from IRST Meldola, Italy. The database of this patient contains five scans which allow us to observe the time-development of lesions. A more detailed study of the data showed that this patient suffers from very extensive tumor damage of almost the whole spine, however, it consists of compact tumors as well. The annotation of compact tumors was created with a significant final cleaning refinement by the author of this paper and it was very time consuming.

3.2 IMPLEMENTATION TOOLS

The implementation was performed in the MathWorks MATLAB 2020a programming environment. The Image processing toolbox for visualization of data was used and the Statistics toolbox was used for statistical analysis. For segmentation of lesions and their refinement, the region growing method realized via Image Segmenter has been used which is included in the Image Processing toolbox.

3.3 EXPERIMENTS AND DISCUSSION

The data are from a patient who had a primary breast tumor with spinal metastases and was treated with bisphosphonates that inhibit bone de-mineralization. The patient suffers from severe damage to the tumor tissue, therefore it was very difficult to find a compact tumor in five consecutive scans. Lesions of different sizes and from different vertebrae were selected. The created database of five compact lytic lesions shows the time-development of the mean value of intensity (see Fig. 3 left) and volume from shape descriptors (see Fig. 3 right). In the mean intensity graph, we can see a slight decrease in mean intensity, which can characterize that the treatment does not work. In the graph of the volume we can see that two lesions are growing, two lesions are decreasing in size and one stays

the same. Next, a graph of the mean value and normalized variance from the whole body of the fifth lumbar vertebra is shown (see Fig. 2). An area of 40×45×40 voxels was selected and the features from this area were calculated. Their time-development was plotted (Fig. 4). It can be seen from the graph, that the mean intensity value and normalized variance increase over time. It could indicate a response to the treatment where calcium is deposited in the area of the lytic lesions.
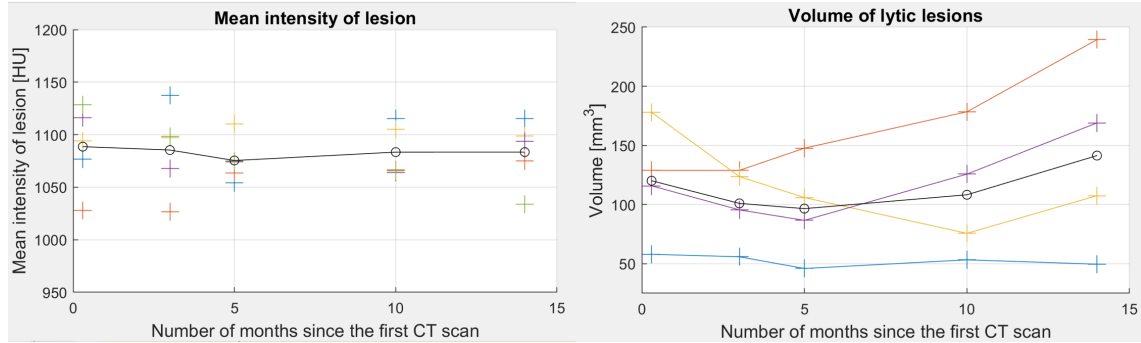
Figure 3: There are two graphs for the analysis of compact tumors. On the left is the mean intensity of lesions and on the right is the pseudo linear graph of shape descriptor. The volume was chosen for simplification. Each color of points and curves represents one lytic tumor and the black line shows the mean across all tumors for a given feature.
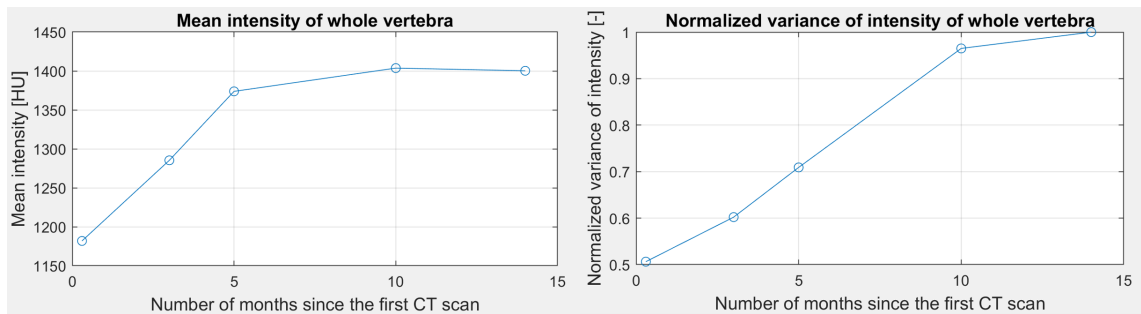
Figure 4: Pseudo linear graph of mean intensity from whole fifth vertebra body (left) and normalized variance of the intensity of this vertebra (right).

4 CONCLUSION

This contribution deals with the design of the methodology for the time-development analysis of vertebral tumors in CT data. The basic challenges for lesions analysis are described. In this paper has been designed a methodology of evaluation of time-development, which was divided into the analysis of compact tumors and the analysis of whole vertebral bodies. For each of both analysis methods, the potential relevant features have been suggested. A manually segmented annotation of lytic tumors was created and examples of statistical features were calculated and plotted by time-graphs.

REFERENCES

[1] MACEDO, Filipa, Katia LADEIRA, Filipa PINHO, Nadine SARAIVA, Nuno BONITO, Luisa PINTO a Francisco GONCALVES. Bone metastases: an overview. *Oncology Reviews* [online]. 2017, **11**(321), 43-49 [cit. 2020-11-22]. ISSN 1970-5565. Available from: doi:10.4081/oncol.2017.321

MODEL OF CHAOTIC SINGLE STAGE TRANSISTOR AMPLIFIER IN CLASS C

Miroslav Rujzl

Master Degree Programme (5), FEEC BUT

E-mail: xrujzl00@stud.feec.vutbr.cz

Supervised by: Jiří Petržela

E-mail: petrzelj@feec.vutbr.cz

Abstract: This article deals with design and experimental verification of a single stage transistor amplifier in class C which is introduced to chaotic state. The basic procedure of deriving differential equations and representation of dynamical system is presented. Equivalent circuit of transistor stage was transformed to the real model based on operation amplifier integrators. Chaotic behavior was confirmed by strange attractors captured by oscilloscope.

Keywords: dynamical system, chaos, chaotic system, transistor amplifier in C class

1 INTRODUCTION

Chaos often refers to a state in which the behavior of a system appears to be completely random and disordered. However, if we look at the chaos from scientific point of view, we talk about a deterministic chaos. Actual state of the dynamic system looks like a random state, but there is no random element in mathematical description. These dynamical systems are described by ordinary differential equations and governed by chaos theory. Chaotic behavior has been demonstrated, both theoretically and experimentally, in a large number of electronic circuits. For example, existence of robust strange attractors was observed in harmonic oscillators such as Collpits [1], Hartley, RC-feedback, analog filters like KHN topology [2], Sallen-Key structure, FIR filter, IIR filter, power circuits such as DC-DC converters [3], switched regulators, PLL circuit [4], AD converter with Σ-Δ modulation etc.

2 MODEL OF TRANSISTOR AMPLIFIER

Recently, a chaotic behavior has also been found in the single stage transistor amplifier working in class C. Despite simplicity of principal schematic, operating point of bipolar transistor is hypothetical and transistor itself is modelled as two-port described by full admittance matrix. The amplifier is loaded by resonant circuit which is lossless. According to [5] and Fig. 1 we can derive the differential equations that are described by change of input and output voltage u_1, u_2 and inductor current i_L.

$$\frac{du_1}{dt} = -\frac{y_{11}}{c_1}u_1 - \frac{y_{12}}{c_1}u_2 \tag{1}$$

$$\frac{du_2}{dt} = -\frac{y_{21}}{c_2}u_1 - \frac{y_{22}}{c_2}u_2 - \frac{1}{c_2}i_L \tag{2}$$

$$\frac{di_L}{dt} = \frac{1}{L}u_2 \tag{3}$$

where y_{xx} describe admittance parameters of transistor, c_1 is a parasitic capacitance between base and emitter and c_2 is a sum of a parasitic capacitance collector-emitter and capacitance in a resonant circuit. Chaotic behavior is subjected to definition of admittance parameters of transistor. We assume that input and forward transadmittance y_{11} and y_{21} is a constant and output admittance y_{22} converge to zero. A backward transadmittance y_{12} is created by nonlinear function which can be expressed by a polynomial:

$$y_{12}(u_2) = a + bu_2 + cu_2^2 + du_2^3 + eu_2^4 \qquad (4)$$

where a, b, c, d, e are parameters of the polynomial components. The polynomial expression is not limited to the admittance parameter y_{12}. Chaotic behavior exists also for the variant with polynomial expression of admittance parameter y_{21} or even exists a situation where both admittance parameters are determined by polynomials. Situation with polynomial y_{12} is used for better circuit realization and simpler form of mathematical model from the algebraic viewpoint.
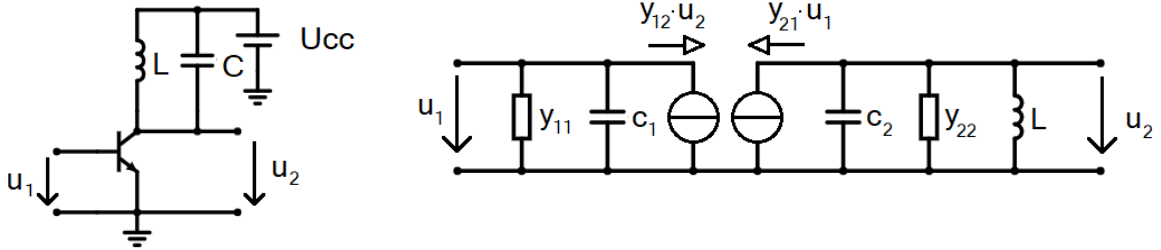
Figure 1: Transistor amplifier in class C and equivalent circuit model

3   INTEGRATOR MODEL

Based on the knowledge of the shapes of differential equations with the polynomial expression of admittance parameters, a real circuit can be constructed. This real circuit models with operational amplifiers by using function blocks. Electric current i_L, which acts as a state variable, is difficult to measure, so the third DR is realized by using a concept with an integrator, so the vector of state variables passes only to voltage values and the equations are representing the Kirchhoff´s voltage law. Another change compared to point 2 is in the application of the power forward transadmittance y_{21}. The integrators are formed by a standard connection with the operational amplifier TL082. The last integrator consists of an amplifier AD844 and nonlinear voltage elements are solved by analog multipliers AD633. The circuit can be described using the DR system as follows:

$$\frac{dx}{dt} = -\frac{1}{R_{T2}C_1}x - \frac{1}{R_{T1}C_1}y - \frac{K^2}{R_{T3}C_1}y^3 \qquad (5)$$

$$\frac{dy}{dt} = \pm\frac{K^2}{R_{T4}C_2}x^3 - \frac{1}{R_1C_2}z \qquad (6)$$

$$\frac{dz}{dt} = \frac{1}{R_2C_3}y \qquad (7)$$

where x, y, z are individual state variables representing voltages at the outputs of integrators, RT1 – 4 are values of resistors that simulate admittance parameters, C1, C2, C3 are capacitances of individual integrators and K is internal constant of analog multiplier (for circuit AD633 it holds that $K = 0,1$).

The realized connection can be seen in Fig. 2, resp Fig. 3. Individual state voltages are output using BNC connectors for easier measurement on the oscilloscope. Resistors simulating admittance are implemented as trimmers to allow the circuit to be tuned to a chaotic state. Simple shorting jumpers are used it switch the sign at the first term in equation 7. Fig. 4 shows the measured attractors on an oscilloscope, which with their complexity and geometric structure correspond to strange attractors calculated by numerical integration of differential equations (1), (2) a (3).
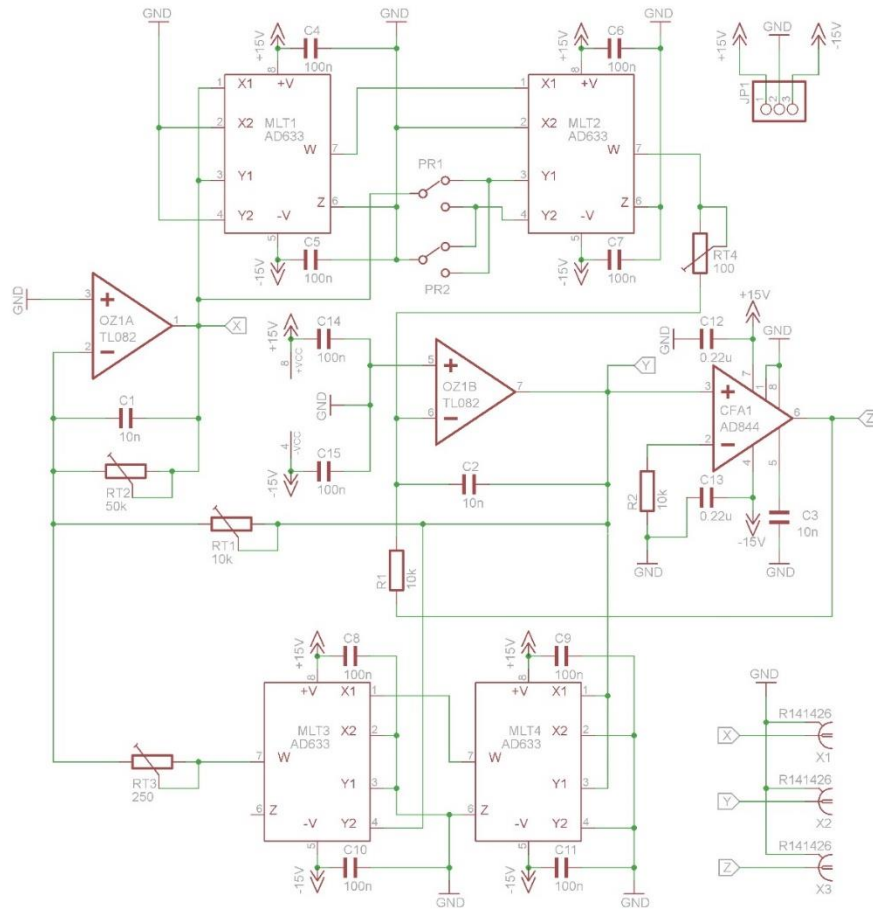
Figure 2: Schematic of the integrator realization
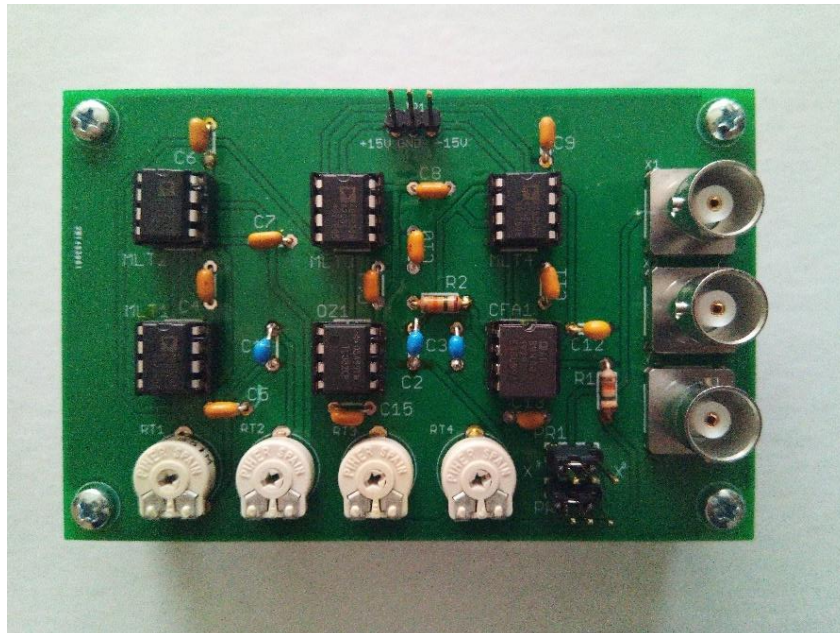
Figure 3: PCB of the integrator model of amplifier
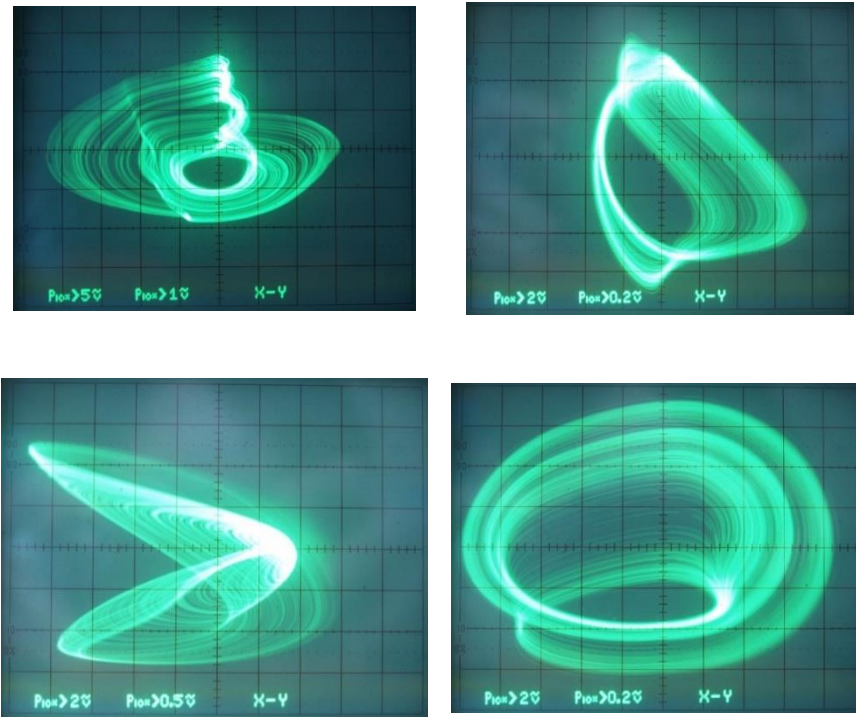
Figure 4: Attractors which were generated by integrator model

4 CONCLUSION

A prototype board modeling the differential equations of a transistor amplifier based on integrators confirmed the presence of chaotic behavior by generating structurally stable strange attractor. From the above findings, it follows that for the existence of a chaotic solution, only one bipolar transistor with a suitable setting of the operating point, i. e. admittance parameters, is needed. This fact could be important in the design of the amplifier itself and in common application to treat the conditions necessary for chaos to avoid bringing amplifier into a chaotic state. When using a multistage amplifier with generalized bipolar transistors, other chaotic or even hyperchaotic states could occur in this case.

REFERENCE

[1] KENNEDY, M.P. Chaos in the Colpitts oscillator. *IEEE Trans. Circuits Syst.* 1994, 41, 771–774.

[2] PETRŽELA, J. On the existence of chaos in the electronically adjustable structures of state variable filters. *Int. J. Circuit Theory Appl.* 2016, 11, 605–653.

[3] ZHOU, X.; J. LI, M. YOUJIE, Chaos phenomena in DC-DC converter and chaos control. *Procedia Eng.* 2012, 29, 382 470–473.

[4] ENDO, T.; L. O. CHUA, Chaos from phase-locked loops. *IEEE Trans. Circuits Syst.* 1988, 35, 987–1003.

[5] PETRŽELA, J. Evidence of strange attractors in class C amplifier with single bipolar transistor: polynomial and piecewise-linear case, *Entropy*, 2021, 23, 175.

WEB APPLICATION FOR THE DETECTION SYSTEM OF RISK SITUATIONS AT THE RAILWAY CROSSING

Jan Bachorec

Master Programme (2), FEEC BUT

E-mail: xbacho01@stud.feec.vutbr.cz

Supervised by: Zdeněk Martinásek

E-mail: martinasek@feec.vutbr.cz

Abstract: The paper describes a design and an implementation of the complex server application including the graphical user interface. The application is designed to detect dangerous situations at the rail crossing. The final own implementation consists of a server application and a database (full stack approach). The server application is based on .NET Core. Its purpose is to receive, process and securely store data from autonomous cameras into database. The database is based on the Cassandra database system. Graphical user interface is an interactive website implemented with Razor Pages technology.

Keywords: web application, visualisation, database systems, API services

1   INTRODUCTION

In recent years there has been an increase in traffic intensity. According to statistics from the Ministry of Transport [1], traffic on most types of roads increased by an average of almost 12 % between 2010 and 2016. Unfortunately, the higher traffic density is associated with an increase in the number of traffic accidents, as can be seen from police statistics [2]. Accidents at the railway crossing, as reported in the materials of the BESIP [3] organization, do not have a dominant share in the absolute number of accidents. On the other hand, when such an accident occurs, it tends to be very serious and often involves injuries and loss of life. The mentioned statistics clearly show the need to introduce additional safety mechanisms to help reduce the number of dangerous situations at a level crossings, thereby reducing damage to health and property.

In this paper, I am researching, developing and implementing a complex server solution for the autonomous system for detecting dangerous situations at railway crossings (ADEROS). The main purpose of the server solution is to collect and store data from individual camera modules monitoring railway crossings. The graphical user interface consists of the interactive web page and is intended to provide an overview of detected situations. Operator can analyze these situations.

2   OWN DESIGN AND IMPLEMENTATION OF THE SERVER SOLUTION

The system for autonomous detection of risk situations in transport consists of several basic components. The server solution, whose design and implementation I describe in this paper, consists of a data gathering and processing service. Collected data are stored in the database. Cameras autonomously detecting different traffic situations are used as a data source for the server solution. Interaction between the system operator and the system itself is provided by the own interactive web application. Its development is discussed in the following chapter. Figure 1 shows the complete architecture design of the system. The following text contains more detailed descriptions of the basic components of the server solution of the system for autonomous detection of risk situations at a railway crossing.
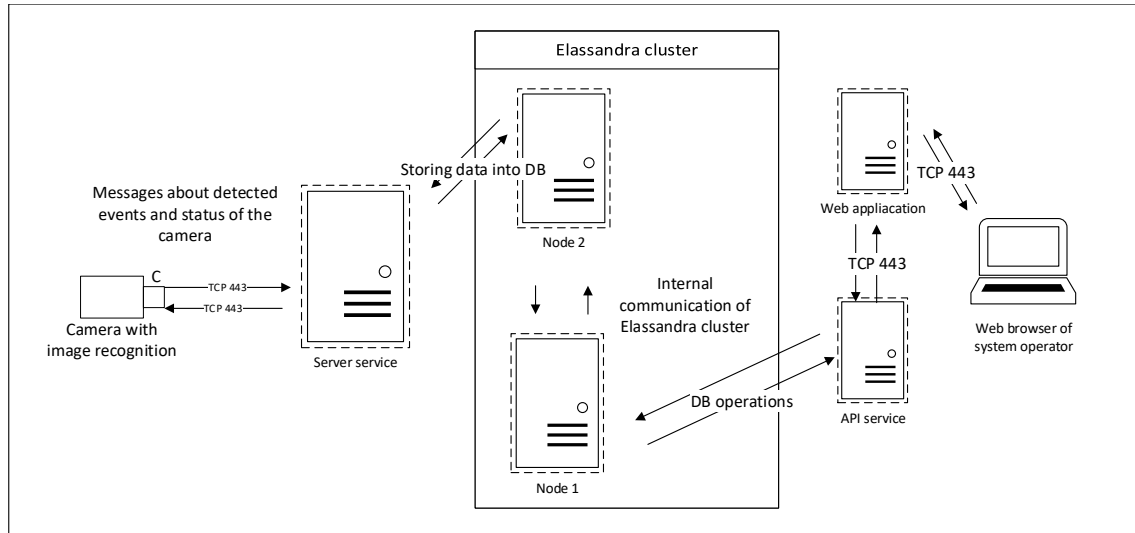
Figure 1: Design of the ADEROS server architecture.

The created **server service** provides an application interface for receiving messages from the cameras. And it also allows to process and store the messages into a database. The technology chosen for the implementation of the server service is the .NET Core platform and the C# programming language. With the help of these tools, I have implemented the REST API (Application Programming Interface) required to receive both event and status messages from the camera modules. Their detailed description is beyond the scope of this article, but it is necessary to describe communication interface of cameras. It consists of an HTTP module that allows sending camera status and event data in JSON (JavaScript Object Notation) format to the server solution. It also consists of an SSH (Secure Shell) server, which can be used to remote management of the cameras.

The created application consists of API controllers that are used to serve HTTP (Hypertext Transfer Protocol) requests from the cameras, containing information about their status and also about detected events. At the same time, these controllers, using the technique called DI (Dependency Injection), use other supporting classes and their methods to parse, process and store the data in the database. To work with the database server, the CassandraCSharpDriver version 3.16.1 is used. The program contains a namespace with classes that use this driver and provide communication with the database and correct mapping of data to the appropriate columns of the database tables. These classes include objects representing database cluster connections and sessions. In this case, the DI is used in such a way that only one instance of these objects is worked with during the program run in order to follow the recommendations for proper use of the CassandraCSharpDriver.

Microsoft IIS (Internet Information Services) was chosen as the web server technology to run the .NET Core server service. As mentioned in the documentation, it is a server bundled with the Windows operating system and Windows Server. More specifically, the current version of IIS 10.0 comes with the Windows 10 operating system and the Fall Creators Update package installed, as well as the Window Server 2016 v. 1709 server operating system.

Various database systems were considered for data storage. When deciding between them, I was guided by the requirement to store a large amount of data in real time. Furthermore, I preferred to use open source software. It was also necessary to consider the planned extension of the application. For example, it is planned to include more advanced statistical data processing. I decided to chose the **Elassandra** tool for data storage. It is a software combining the Cassandra NoSQL (Not only Structured Query Language) database system with the Elasticsearch search tool. Cassandra can ef-

ficiently handle a large number of data write operations if the schema design is appropriate, while providing considerable flexibility in data model design options. For example, through the support of user-defined data types. In the context of this work, Cassandra is used as a repository for detected events and status messages, as well as other information necessary for running developed systems. Elasticsearch provides a secondary index over tables.

3 OWN GRAPHICAL USER INTERFACE IMPLEMENTATION

The graphical interface is used to clearly display the detected situations to the ADEROS operator. I have created my own interactive web application using Razor Pages technology. This technology is part of the .NET Core framework, also used for the development of the server service described in previous section. In order to separate working with the database and other necessary services from the user interface itself, I created a REST API service for this purpose, also built on .NET Core and implementing all database operations.

Web application, whose interface is shown in the picture 2, contains four basic modules from the user's point of view. Within the camera module there is an overview of all available cameras in the system. The purpose here is to inform the operator if the camera is connected and communicating properly. It is also possible to click on a camera to view detailed information about its status and also, for example, to change its name in the system. The alert module informs about the latest detected risks. Again, you can find both the summary information about the latest alerts from all cameras and, after opening the appropriate alert, its detail. Within the events tab it is possible to view all recorded events. Even those that are not classified as alerts. The filtering of the displayed events uses name of the source camera and the time range of event detection. Administration allows, for example, to manage user accounts within the system.

The web application is secured against unauthorized access. A REST API service is used for user authentication. If the user is registered in the system, the API service returns a JSON Web Token (JWT) token, based on which the web application issues a Cookie to the user's browser, authorizing the user to access individual web pages.
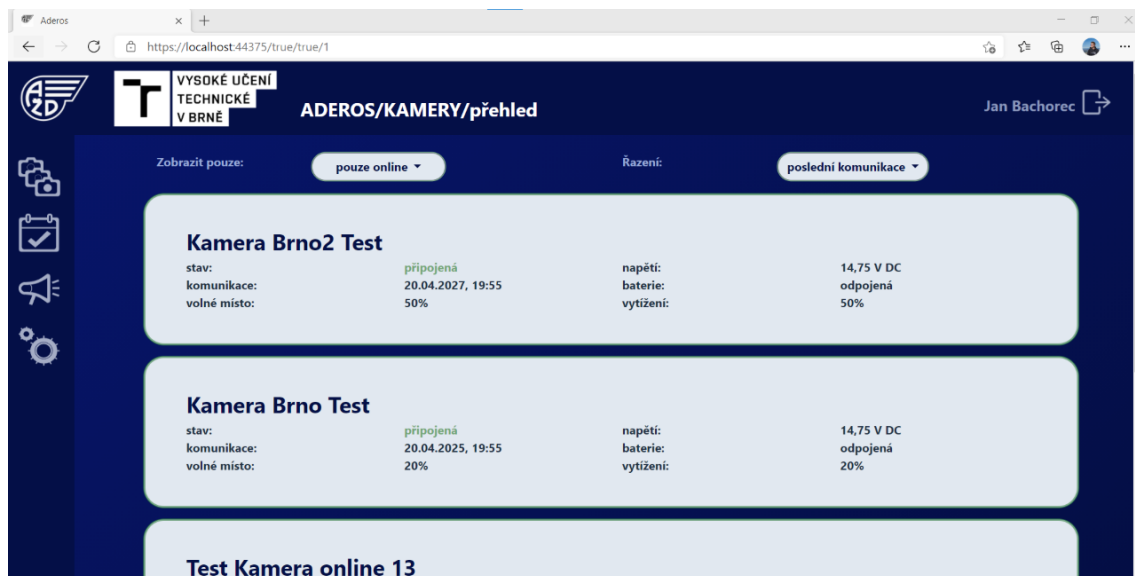

Figure 2: User interface of the own interactive web application.

API service is built on .NET Core and its task is to provide the necessary data to the web application via REST API so that information can be displayed to the user. The service contains both HTTP

controllers to handle requests from the web application, and classes that implement all supported operations. The interconnection between the controllers and the mentioned classes is implemented using a dependency injection technique. Working with tables in the Cassandra database is implemented similarly to the server service described above, using the CassandraCSharp driver. Communication between the web application and the API service is implemented using HTTP messages. Data is transferred in JSON objects within the body of the exchanged messages.

The service controllers are secured to prevent unauthorized access to the system. The service has implemented methods for authentication, in which it issues a JWT token to an authorized user based on knowledge of the login credentials. This is a JSON object signed with the service's secret key. This token contains information about the user and his role in the system. HTTP requests are then handled only if they contain a valid token.

4 CONCLUSION

This paper describes the research and the development of the server application for the system for detecting risky situations at a railway crossing. It is a complex system that allows to clearly display and analyze data obtained by cameras with autonomous image recognition. As a part of the actual design and development, I first had to implement a database schema design using Cassandra technology. I designed the scheme with an attention to high performance even in the case of connecting a large number of camera modules. The next step was to create a server service for collecting, processing and storing the data into the created database. I used the .NET Core platform to implement it. To enable interaction between the operator and the ADEROS system, I designed and implemented the interactive web application, built on Razor Pages technology. To provide this application with data, I also implemented the REST API service through which data is obtained from the database. The implementation of all components of the system has been tested and is fully functional. The advantage of the technologies used is the easy extensibility of the system if needed.

ACKNOWLEDGEMENTS

REFERENCES

[1] *Ředitelství silnic a dálnic zahájí v červenci Celostátní sčítání dopravy. In: Ministerstvo dopravy [online]. 2020, 24. 6. 2020 [cit. 2020-10-09].* Dostupné z:
https://www.mdcr.cz/Media/Media-a-tiskove-zpravy/Reditelstvi-silnic-a-dalnic-zahaji-v-cervenci-Celo.

[2] *Statistika nehodovosti 2019. In: Policie [online]. 2019 [cit. 2020-10-09].* Dostupné z:
https://www.policie.cz/clanek/statistika-nehodovosti-900835.aspx?q=Y2hudW09Mg%3d%3d.

[3] *Železniční přejezdy. Besip - Statistiky nehodovosti v Ceske republice [online]. 2019, , 28 [cit. 2020-10-09].* Dostupné z:
https://www.ibesip.cz/getattachment/Statistiky/Statistiky-nehodovosti-v-Ceske-republice/Dopravni-nehodovost-v-roce-2019/Zeleznicni-prejezdy/Zeleznicni-prejezdy.pdf.

[4] *Operating Cassandra. Cassandra Documentation [online]. [cit. 2020-11-27].* Dostupné z:
https://cassandra.apache.org/doc/latest/operating/index.html/.

IMPLEMENTATION OF APPLICATION THAT DEMON-STRATES MOBILE APPLICATION VULNERABILITIES

Karolína Šrůtková

Master Degree Programme (2), FEEC BUT

E-mail: xsrutk02@stud.feec.vutbr.cz

Supervised by: Zdeněk Martinásek

E-mail: martinasek@feec.vutbr.cz

Abstract: Nowadays, using mobile applications is a daily routine for all of us. This trend is misused by hackers who attack all kinds of mobile apps from which they can get not only sensitive data of users. The main contribution of this paper is a design and implementation of a mobile application that demonstrates mobile application vulnerabilities of Android mobile operating system and that could help developers to build more secure applications.

Keywords: Android, mobile application, vulnerability

1   INTRODUCTION

During the day users of the smartphones use several types of mobile applications that make their work, school and everyday life easier. This fact should lead to a greater emphasis on user security through the implementation of security features in both mobile operating systems and mobile applications. The growing trend of using mobile applications is a good thing for attackers who are trying to exploit their vulnerabilities and obtain any user data, mostly sensitive one. The main benefit of this article a design and implementation of a mobile application that demonstrates mobile application vulnerabilities on the Android operating system. This implemented application brings to developers a possibility to explore different vulnerabilities in real time and in real environment. Understanding the principle of vulnerability and its exploitation leads to the implementation of countermeasures and therefore the resulting application is safer.

2   CURRENT VULNERABILITIES OF MOBILE APPLICATIONS

Like any other operating system, Android has its vulnerabilities. Its biggest vulnerabilities include the vulnerability of exported activity, missing authorization, the vulnerability of broadcast receivers and services, the vulnerability of internal storage and unauthorized access to data.

The vulnerability of the exported activity means that when the exported activity is used, then all components of this activity can be accessed by any other application, and thus sensitive data can be easily obtained and processed by the application. *The missing authorization* is mostly mentioned in connection with API (Application Programming Interaction) calls and with Insecure Direct Object References (IDOR) vulnerability. IDOR vulnerability takes advantage of the fact that objects are accessed directly, most often using the user's ID (Identification). In an incorrectly implemented API request with method GET, it is enough to change the user ID in the URL (Uniform Resource Locator) path and the attacker will obtain information about any user.

The vulnerability of the other two basic components of the application, *broadcast receivers and services*, is also based on exportability, as in the case with activities. Moreover, incorrectly set permissions of these components will not ask for permission to access the application from another application. Files stored on *the internal storage* can be easily obtained by attackers if the content providers are not used to access the files, but the parameters are set directly to the files. These are

the MODE_WORLD_WRITEABLE parameters for the edit option and MODE_WORLD_READABLE for the read option. This type of implementation does not provide the possibility to restrict access to data to specific applications. ***Access to data*** in the application is controlled by content providers, which in case of poor implementation can be exploited by vulnerabilities SQL (Structured Query Language) Injection and Path Traversal. SQL Injection vulnerability can be exploited when reading data from the database in case the insertion of a specific SQL query into the input is not treated according to rules, which may lead to gain or modification of sensitive data. To exploit the Path Traversal vulnerability, the content provider methods for opening files *openFile ()* and *openAssetFile ()* must be poorly implemented. In these methods the Uniform Resource Identifier (URI) parameters can be incorrectly verified and thus the attackers can gain unauthorized access to files and directories that are stored outside the root directory by specifying the absolute path to the file.

3   DESIGN OF THE MOBILE APPLICATION

The main benefit of this article is the design and implementation of a mobile application, which demonstrates the vulnerabilities of the mobile application mentioned above and vulnerabilities that are present in communication with the web server. The mobile application is being developed as part of the PenterepMail project, which is still under development and will serve as a complement to the ongoing Penterep project focused on penetration testing. Once the entire PenterepMail project is completed, it will contain a vulnerable web and mobile application in the form of a mail client. Both applications will communicate with the virtual server using the PenterepMail API. Figure 1 shows the communication blocks of all parts of the PenterepMail project. Further in the text, only the actual design of the mobile application is described.
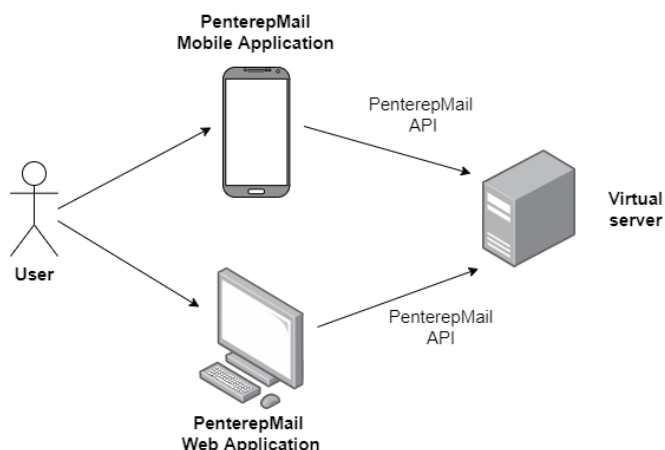
Figure 1: Communication blocks design of all parts of the PenterepMail project.

The mobile application is designed as a mail client. The user will be able to activate the application, log in and log out, view received and sent messages, send new messages and manage contacts, including their adding and removing. All these functionalities will be accompanied by the deliberate implementation of vulnerabilities based on the above analysis. The application is being developed in Android version 4.1 (API 16) for a clearer demonstration of vulnerabilities. Vulnerabilities will be also demonstrable on newer versions of the operating system, but not in a full scope.

After launching the mobile application, an activation screen will appear, through which the user must enter his phone number. The application sends an SMS to the entered number with a confirmation code, which was generated during the event for sending an SMS and which was saved in the application file. If the code stored in the application and the code sent in the SMS matches, the user is allowed to use the application.

The first screen after activation will be a welcome screen with options for login, for settings and for reading more information about how to use the application. Within the application settings, it will

be possible to set the IP address of the server in order to log in to the application. After logging in to the mobile application, the user will be able to view received and sent messages and add or remove contacts. It will also allow user to send new messages using the API and change the login details or IP address of the server. Screen designs before login and after login are available in Figure 2.
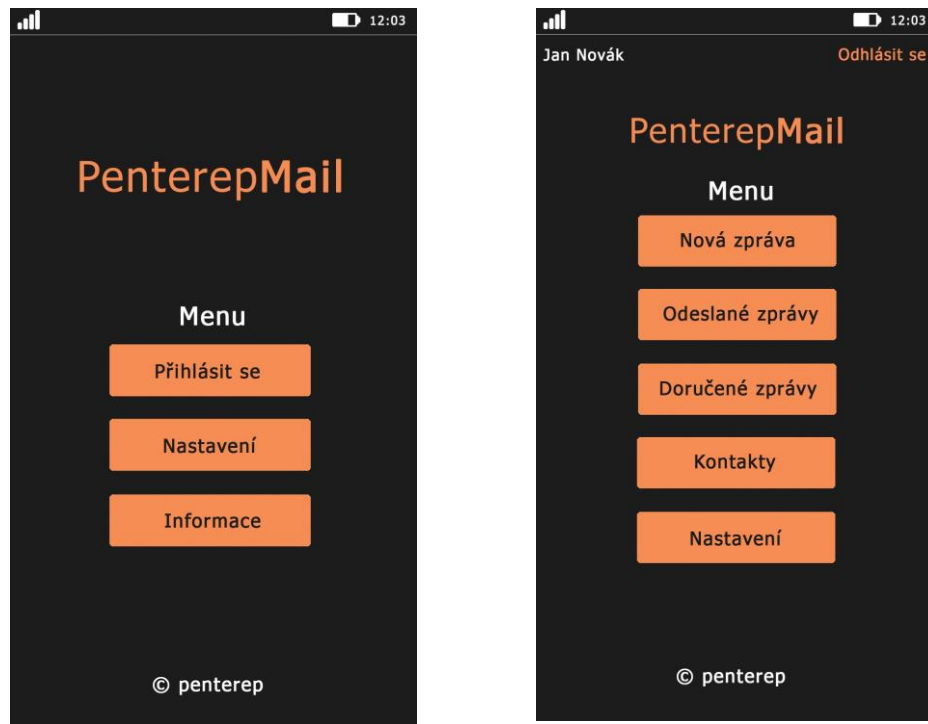
Figure 2: Screen designs before login (left) and after login (right).

The login credentials will be stored in the *settings.xml* file, which will be located in the internal storage. In addition, this file settings will have bad permissions, so it will be very easy to read its contents, including credentials. It will also be possible to access this file by exploiting the Path Traversal vulnerability. Registration will be allowed to the user only through the web application. Only a registered user will be able to log in to the mobile application. After a successful login, the user ID will be returned from the server with which the application will continue to work.

The application will be delivered as a trial version, from which it will not be possible to send messages. Messages can only be sent after entering the license key. The "New Message" activity will only run when the license is activated. However, the *settings.xml* file will contain a boolean value that indicates whether the application license is already activated. Due to the wrong permissions of the *settings.xml* file, the user can change this value from *false* to *true* and therefore achieve activation of the application license. It will also be possible to change the value while debugging the application. Another approach to bypassing the license will be a possibility to decompile the application. After decompilation, the user will be able to change the condition for starting the activity "New message" and after the subsequent compilation it will be possible to send messages also in the trial version of the application. The license key will not be provided to the user in any way. The user either bypasses entering the key using the mentioned steps above or finds it directly in the application code. In the code, the entered license key will be verified against the SHA1 hash, which is very easy to crack.

Contacts stored via the mobile application will be stored in a database in the device's internal storage. The database will be encrypted, but the password will be stored directly in the application code. An unexported content provider will be used to access the contacts and will be vulnerable to SQL Injection. The application will allow user to download attachments to the device's internal storage. Access to the files will be through a content provider which will be vulnerable to Path Tra-

versal. Sending messages will be solved by broadcast receivers, which will be exported and will be without appropriate authorization. Adding and removing contacts will be handled through exported intents. Sending an intent from another application will allow user to add or remove contacts without knowing the login. The application will also implement exported services running in the background, which will trigger requests at given intervals to see if a new message has arrived.

4   CUSTOM IMPLEMENTATION OF THE MOBILE APPLICATION

The first step in developing the application was to implement an activation screen. The activation screen is used in the application to demonstrate the vulnerability of the exported broadcast receiver. After entering the phone number and confirming it, the application calls an exported broadcast receiver called *SendSMSReceiver*. Its task is to send an SMS with the generated activation code to the given phone number. In order to send this broadcast receiver, an intent with the selected action name SEND_SMS must be passed to it. This intent also contains extra data representing the entered phone number and the generated code.

To test this vulnerability, mobile application penetration testing tool called Drozer was used. The tool is controlled from the command line. With its help, it was possible to eavesdrop on the sent data and find out to which telephone number and with what content the message is sent. With the acquired knowledge, it was also possible to use this broadcast receiver to send a message to any number with any message. The test result can be seen in Figures 3 and 4 as the output of the Drozer tool. Figure 3 shows a list of exported broadcast receivers in a given package. Figure 4 shows a command after which a message is sent to the selected number with the selected content.

```
drozer Console (v2.4.4)
dz> run app.broadcast.info --package com.example.penterepmailmobileapp
Package: com.example.penterepmailmobileapp
  com.example.penterepmailmobileapp.SendSMSReceiver
    Permission: null
```

Figure 3: List of found exported broadcast receivers in a given package.

```
dz> run app.broadcast.send --action SEND_SMS --
extra string phoneNumber "5554" --extra string
message "Zmeneny obsah zpravy"
dz>
```

Figure 4: Command for sending message to the selected number with selected content.

5   CONCLUSION

The resulting implementation of the application will be used primarily by developers to explore the vulnerabilities of the mobile application in a real environment and time, because it is a fully functional mail application. They will be able to test what can be caused by poor implementation of individual components and what is the impact due to leakage or modification of user data. The created mobile application should primarily serve as a tool that will contribute to the creation of more secure applications.

REFERENCES

[1]   Kotipalli, S. R; Imran, M. A. Hacking Android. Birmingham: Packt Publishing Ltd., 2016. ISBN 978-1-78588-314-9.

[2]   Nidecki, T. A. What Are Insecure Direct Object References [online]. 23. 3. 2020 [cit. 8. 3. 2021]. Dostupné z: https://www.acunetix.com/blog/web-security-zone/what-are-insecure-direct-object-references/

[3]   Security tips [online]. Google Developers, 2020 [cit. 7. 3. 2021]. Dostupné z: https://developer.android.com/training/articles/security-tips

BEEHIVE SECURITY SYSTEM AGAINST UNAUTHORIZED MANIPULATION

Martin Milota

Master Degree Programme (2nd year), FEEC BUT

E-mail: xmilot03@stud.feec.vutbr.cz

Supervised by: Petr Dejdar

E-mail: xdejda00@stud.feec.vutbr.cz

Abstract: The paper deals with the design of beehive security system against unauthorized manipulation. This system uses LoRaWAN wireless communication technology for data transmission. The device records the movement of the beehive using an accelerometer and then locates the beehive using a GPS module. This low-energy system works via a battery-powered ESP32 microcontroller and is located on a beehive.

Keywords: ESP32 microcontroller, GPS localization, Internet of Things, LoRaWAN network, LPWAN technologies, securing beehives

1 INTRODUCTION

Beekeeping is very widespread in the Czech Republic, and this is associated with a high rate of beehive theft. That is why sufficient hive security is a very important part of beekeeping. This is a security that uses modern wireless transmission technology in IoT. The introduction of the article explains the general issues of IoT and ESP32 microcontroller. For security design the article describes a suitable LoRaWAN transmission technology. The last chapter of the article is devoted to the final design of the security device and the test connection of part of the solution.

2 INTERNET OF THINGS

IoT (Internet of Things) can be defined as a network of physical objects that are interconnected, can communicate with each other, share and analyze data that provide a wealth of information for planning, decision making and management, which illustrates fig. 1. These physical objects are devices of various types and sizes such as vehicles, smartphones, toys, home appliances, industrial systems, medical instruments, buildings and more. IoT is not the only technology, but a combination of different hardware and software technologies. At present, we can already find many practical applications in many areas, such as security of property, agriculture, smart cities, houses and so on [1].

2.1 IoT TECHNOLOGIES

IoT applications have specific requirements, such as long range, low data rate, low power consumption and low cost. Frequently used technologies with short signals, such as ZigBee and Bluetooth, are not adapted to scenarios that require long-distance transmission and are therefore not suitable for such applications. Also, solutions that are based on mobile communication, such as 2G, 3G and 4G, provide a lot of coverage, but consume an excessive amount of device power. That is why these specific requirements of IoT applications have led to the emergence of new wireless communication technologies based on LPWAN (Low-Power Wide Area Network) [2].

Figure 1: Internet of Things.

3 MICROCONTROLLER ESP32

There are currently many different microcontrollers available in the market for IoT solutions. Such microcontrollers include, for example, the best-known Arduino, but also Xbee or WhizFi. However, most of these microcontrollers have a high weight or high cost and very few of them are open soruce. And this is the advantage of the device from Espressif Systems, which offers the ESP32 microcontroller with integrated 2.4 GHz Wi-Fi, Bluetooth and BLE (Bluetooth Low Energy) technology. The ESP32 system is designed and optimized for the best performance and size, with many advanced low-power chip features for a variety of applications. It is mainly focused on mobile applications, small electronics and IoT applications [3], [4].

4 LORAWAN

The LoRaWAN network has been designed for IoT applications to connect thousands of devices, modules and sensors. It operates in the unlicensed ISM (Industrial, Scientific and Medical) communication band and consists of two different layers. The physical layer of the standard is formed by LoRa modulation and above it is the MAC layer, the operation of which is ensured by the open-source standard LoRaWAN [5], [6].

LoRaWAN networks are formed by a star topology. The basic elements of the network are the end devices that send data to the LoRaWAN gateways. Then, these gateways send the data to the application servers. The operation of such a network can be seen in fig. 2. The end devices communicate with the LoRaWAN gateways wirelessly using the LoRa physical layer. The gateways communicate with application servers via the Internet using 3G / 4G mobile networks, Ethernet or other similar transport communication technology. All measured data thus reaches directly into the data storage, to which the user has access via a dedicated API (Application Programming Interface) or an administrative console running on the application server [7].

5 SECURITY SYSTEM DESIGN

The LoRaWAN network will be used to transfer data from the device to the server, where the data will be further processed. A simplified block diagram of the resulting security system can be seen in fig. 3. It consists of a LOLIN32 ESP-WROOM-32 microcontroller, to which the LoRa module SX1276 is connected for communication via the SPI bus. Next, the GPS module GY-NEO6MV2 connected via the UART bus, which is used to locate the beehive, and the module with the accelerometer and
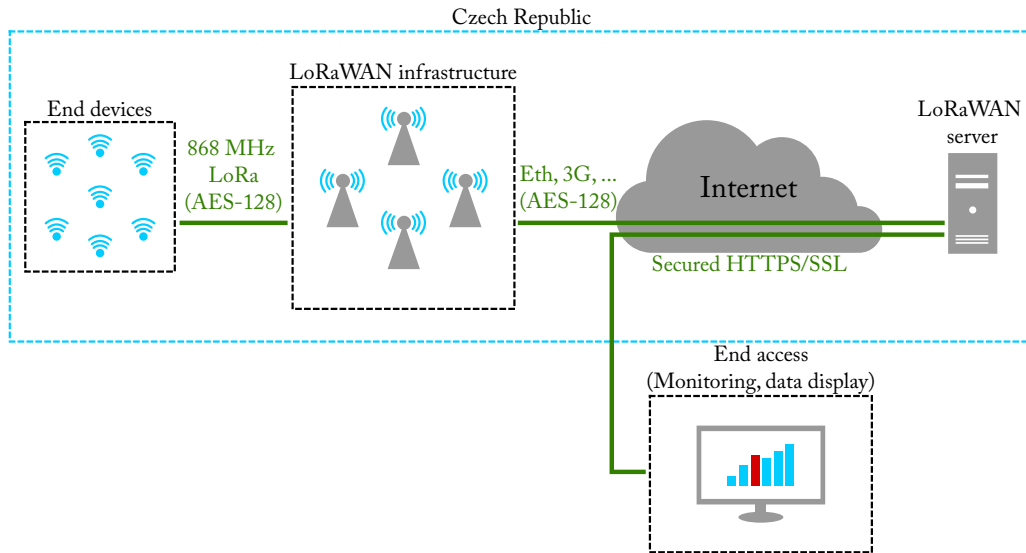
Figure 2: A network using the LoRaWAN protocol [7].

gyroscope GY-521 to record the movement of the beehive connected via the I^{2}C bus.
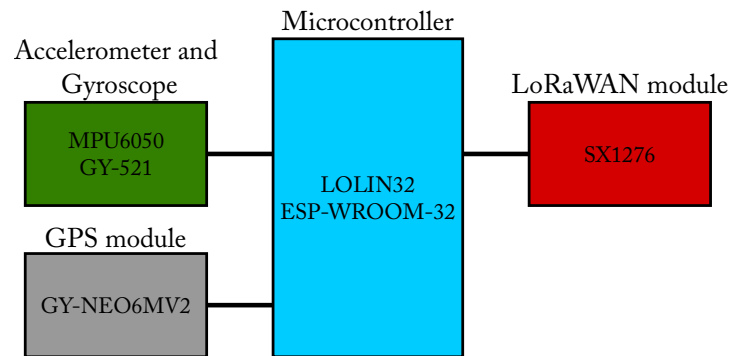
Figure 3: Simplified block diagram of a security system.

The system will be powered via the JST-PH-2 connector using a rechargeable and replaceable LiPol (Lithium Polymer) battery with a capacity of 5000 mAh, which should ensure uninterrupted operation for several months. The microcontroller itself offers 5 possible modes, where in addition to the active mode, the Deep-sleep Mode will be used most of the time with a consumption of 20 μA and will not be awakened until motion is detected by the accelerometer and gyroscope. The location of the device is also enabled by the LoRaWAN technology itself, but in the case of securing beehives, it is assumed that the device will be used outside the built-up area and the necessary network for precise positioning is not yet built in the Czech Republic. That is why the option was chosen to connect a GPS module to the device, which will achieve the required accuracy for locating the beehive. This module will be completely inactive during the operation of the device, a possible solution using MOSFET transistors to save battery power. It will only be activated if the microcontroller is woken to active mode. The LoRa module SX1276 also offers a sleep mode in which it consumes approximately 0.2 μA. After detecting the movement and localization, the command will wake the LoRa module and the data will be sent via the LoRaWAN network to the LoRaWAN server, where it will be further processed. The user will be sent an email about the beehive theft and it will be possible to track his current location using the GPS coordinates latitude and longitude.

6 CONCLUSION

This article focused on the design of a security system for beehives. In the introduction, the concept of IoT and the introduction of the selected ESP32 microcontroller were explained. The following chapter describes the LoRaWAN technology and the last chapter was devoted to the final design of a device for securing beehives. A test connection of a part of the design for functional testing was carried out. This connection consisted of the ESP32 microcontroller itself and a module with an accelerometer and a gyroscope to record the handling of the beehive.

Next, the device will be connected to the LoRaWAN network. A suitable box will be produced for placing the device in the hive, connecting a GPS module to track the position of the hive. The device will then be practically tested to see if it meets the basic requirements for securing beehives.

REFERENCES

[1] PATEL, K. K.; PATEL, M. S. *Internet of Things-IOT: Definition, Characteristics, Architecture, Enabling Technologies, Application & Future Challenges.* International Journal of Engineering Science and Computing, 2016, vol. 6, no. 5, s. 6122–6131. ISSN 2321-3361.

[2] MEKKI, K.; BAJIC, E.; CHAXEL F.; MEYER F. *A comparative study of LPWAN technologies for large-scale IoTdeployment.* The Korean Institute of Communications and Information Sciences, 2019, vol. 5, no. 1, s. 1–7.

[3] MAIER, A.; SHARP, A.; VAGAPOV, Y. *Comparative analysis and practical implementation of the ESP32 microcontroller module for the internet of things.* Internet Technologies and Applications (ITA), Wrexham, 2017, s. 143–148.

[4] PURBA, J.; WAHYUDIN, D. *Bluetooth Low Energy (BLE) Based Power Window System.* IOP Conference Series: Materials Science and Engineering, Bandung, Indonesia, 2017, vol. 384.

[5] DE CARVALHO SILVA, J.; RODRIGUES, J. J. P. C.; ALBERTI, A. M.;SOLIC, P.; AQUINO, A. L. L. *LoRaWAN — A low power WAN protocol for Internet of Things: A review and opportunities.* 2017 2nd International Multidisciplinary Conference on Computer and Energy Science (SpliTech),Split, 2017, s. 1–6.

[6] DAVCEV, D.; MITRESKI, K.; TRAJKOVIC, S.; NIKOLOVSKI, V.; KOTELI, N. *IoT agriculture system based on LoRaWAN.* 2018 14th IEEE International Workshop on Factory Communication Systems (WFCS), Imperia, 2018, s. 1–4.

[7] MAŠEK, P.; ŠTÚSEK, M.; FUJDIAK, R.; MLÝNEK P.; HOŠEK J. *Komunikační systémy pro IoT.* Brno: Vysoké učení technické v Brně, Fakulta elektrotechniky a komunikačních technologií, 2019

IPMSM DRIVE SYSTEM PARAMETERS TUNING - AN ANALYTIC SOLUTION

Lukas Zezula

Master Degree Programme (1), FEEC BUT

E-mail: xzezul06@stud.feec.vutbr.cz

Supervised by: Petr Blaha

E-mail: blahap@feec.vutbr.cz

Abstract: This paper describes an analytical way for tuning the parameters of vector controlled permanent magnet synchronous machines. Current controllers are tuned to guarantee a $60°$ phase margin for the identified parameters of the system. Gains of the controllers are modified to ensure amplitude margin 2 of the current loop even for variable parameters of the machine. The velocity controller gain is tuned to maximize phase margin which is limited by a speed controller time constant placement.

Keywords: PMSM, vector control, parameters tuning, drive system analysis

1 INTRODUCTION

Permanent magnet synchronous machines (PMSMs) are becoming more and more popular due to their reliable construction, high efficiency and power density. PMSMs are classified as surface-mounted PMSMs (SPMSMs) and interior PMSMs (IPMSMs) depending on the permanent magnets position. The IPMSM has a more complex electrical model than the SPMSM due to the saliency effect. However, because of the saliency, the IPMSM has a higher torque and power density.

Typically, a vector control is used to control alternate current (AC) machines. Vector control is a variable-frequency drive control method in which the stator currents of a three-phase AC machine are transformed into the two orthogonal components, also referred to the rotor reference frame $d-q$. Direct axis current i_d defines the magnetic flux and quadrature axis current i_q defines the electromagnetic torque T_e of the machine. Proportional-integral (PI) controllers are usually used to keep the measured current components i_d and i_q at their requested values i_d^* and i_q^*. If the SPMSM operates in the constant torque region then the flux linking of the machine should be given only by the permanent magnets flux linkage λ_{pm}. This is achieved by the zero value of i_d which is provided by the zero requested value $i_d^* = 0$. Due to the magnets position in the IPMSM i_d also contributes to the electromagnetic torque. Then the most efficient current combination which provides the required torque can be determined based on the maximum torque per amper curve. However, the contribution of i_d to T_e is depended on the difference between the direct and quadrature axis inductance L_d and L_q which is very low for most industrial IPMSMs. Then T_e of the IPMSM can be controlled using only i_q. This instantaneous torque control loop which is also called the current loop (CL) is usually extended by the mechanical angular velocity ω_m PI controller. The control action of this PI controller is the requested value of the quadrature axis current i_q^* thus it is the requested electromagnetic torque.

Even though vector control is well known and commonly used, the controllers are still mostly tuned manually, often with the trial and error approach. The aim of this paper is to determine controllers parameters that can be calculated out of machine parameters and ensure the robustness of the solution.

2 CURRENT LOOP - PARAMETERS PROPOSAL

Sul described the electrical model of PMSM in [1]. Assuming that the velocity loop dynamics is much

slower than the current loop dynamics the electrical angular velocity ω_e can be taken as a constant parameter. Permanent magnets flux linkage λ_{pm} can be involved into the input signal vector as a constant input signal. Under these conditions PMSM can be described using the linear model as (1).

$$\begin{bmatrix} \frac{di_d}{dt} \\ \frac{di_q}{dt} \end{bmatrix} = \begin{bmatrix} -\frac{R_s}{L_d} & \frac{L_q\omega_e}{L_d} \\ -\frac{L_d\omega_e}{L_q} & -\frac{R_s}{L_q} \end{bmatrix} \cdot \begin{bmatrix} i_d \\ i_q \end{bmatrix} + \begin{bmatrix} \frac{1}{L_d} & 0 & 0 \\ 0 & \frac{1}{L_q} & -\frac{\omega_e}{L_q} \end{bmatrix} \cdot \begin{bmatrix} u_d \\ u_q \\ \lambda_{pm} \end{bmatrix} \tag{1}$$

In (1), R_s, u_d and u_q stand for the stator windings equivalent resistance and $d-q$ voltage inputs.

As Xingye described in [2], the coupling between currents i_d and i_q may be compensated using the diagonal compensation decoupling current controller as shown in Figure 1.
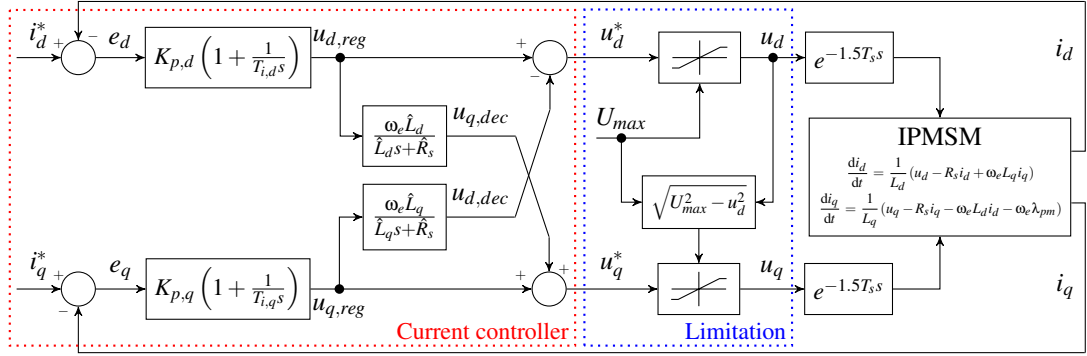

Figure 1: Current loop control structure (neglected anti-windup)

In Figure 1, s is the Laplace operator, $\hat{R}_s$, $\hat{L}_d$, $\hat{L}_q$ stand for the identified stator resistance and $d-q$ inductances, $K_{p,d}$ and $K_{p,q}$ are gains of the current controllers and $T_{i,d}$ and $T_{i,q}$ are integral time constants. The control is implemented in a discrete form so the transport delay (1.5 of the sampling period T_s) caused by sampling and by calculation of the control action is present.

If machine parameters don't vary and they correspond to the identified parameters then currents are given in the Laplace transform as (2).

$$\begin{bmatrix} I_d(s) \\ I_q(s) \end{bmatrix} = \begin{bmatrix} \frac{K_{p,d}(T_{i,d}s+1)e^{-1.5T_s s}}{R_s T_{i,d} s\left(\frac{L_d}{R_s}s+1\right)} & 0 & -\frac{L_q\omega_e^2}{L_d L_q s^2 + R_s(L_d+L_q)s + R_s^2 + L_d L_q \omega_e^2} \\ 0 & \frac{K_{p,q}(T_{i,q}s+1)e^{-1.5T_s s}}{R_s T_{i,q} s\left(\frac{L_q}{R_s}s+1\right)} & -\frac{\omega_e(L_d s + R_s)}{L_d L_q s^2 + R_s(L_d+L_q)s + R_s^2 + L_d L_q \omega_e^2} \end{bmatrix} \cdot \begin{bmatrix} E_d(s) \\ E_q(s) \\ \frac{\lambda_{pm}}{s} \end{bmatrix} \tag{2}$$

Zeros of the PI current controllers are placed to compensate poles of the system and gains are designed to ensure CL phase margin of at least $60°$. Controllers parameters are then described as (3).

$$T_{i,d} = \frac{\hat{L}_d}{\hat{R}_s} \qquad T_{i,q} = \frac{\hat{L}_q}{\hat{R}_s} \qquad K_{p,d} = C_{cl}\frac{\hat{L}_d}{T_s} \qquad K_{p,q} = C_{cl}\frac{\hat{L}_q}{T_s} \tag{3}$$

Where $C_{cl} \in (0, \frac{\pi}{9})$ is a scaling coefficient of current controllers gains. Value $C_{cl} = \frac{\pi}{9}$ corresponds to the phase margin $60°$ of the identified system.

3 CURRENT LOOP - ROBUSTNESS

Parameters design which was presented in the previous section relies on the parameters received from the identification. However, real parameters of PMSM vary depending on the operating point. Stator winding resistance R_s increases with increasing temperature and inductances decrease with increasing current. Parameters R_s, L_d, L_q of the machine can be then expressed as uncertain parameters using identified parameters $\hat{R}_s$, $\hat{L}_d$, $\hat{L}_q$ and multiplication factors $G_{R_s} \in \langle G_{R_s}^-, G_{R_s}^+ \rangle$, $G_{L_d} \in \langle G_{L_d}^-, G_{L_d}^+ \rangle$, $G_{L_q} \in \langle G_{L_q}^-, G_{L_q}^+ \rangle$ as $R_s = G_{R_s}\hat{R}_s$, $L_d = G_{L_d}\hat{L}_d$, $L_q = G_{L_q}\hat{L}_q$.

If the transport delay is neglected then the characteristic polynomial of IPMSM with varying parameters is derived as $\Delta = a_4 s^4 + a_3 s^3 + a_2 s^2 + a_1 s + a_0$ where:

$$a_4 = L_d L_q T_s^2$$
$$a_3 = (L_d + L_q) R_s T_s^2 + 2C_{cl} L_d L_q T_s$$
$$a_2 = L_d L_q \left(C_{cl}^2 + T_s^2 \omega_e^2\right) + \left(\hat{R}_s + R_s\right)(L_d + L_q) T_s C_{cl} + R_s^2 T_s^2$$
$$a_1 = (L_d + L_q) \hat{R}_s C_{cl}^2 + \left(L_d \hat{L}_q + L_q \hat{L}_d\right) T_s \omega_e^2 C_{cl} + 2R_s \hat{R}_s T_s C_{cl}$$
$$a_0 = \left(\hat{R}_s^2 + \hat{L}_d \hat{L}_q \omega_e^2\right) C_{cl}^2 \tag{4}$$

The stability of the system can be evaluated using Routh-Hurwitz algebraic criterion which results in the condition (5).

$$a_1 \left(a_2 a_3 - a_1 a_4\right) - a_0 a_3^2 > 0 \tag{5}$$

Due to the dependence of the Δ coefficients on each other, the replacement with interval polynomials as a successive use of the Kharitonov's theorem leads to the very restrictive values of C_{cl}. However, the stability can be evaluated numerically using (5). Changes of R_s, L_d, L_q and ω_e can be expressed using 4 nested *for* loops and the value of C_{cl} which ensures stability can be found on the interval $C_{cl} \in (0, \frac{\pi}{9})$ using the Bisection method. The electrical angular velocity ω_e is then limited by the interval $\omega_e \in \langle 0, \omega_{e,max} \rangle$ where the highest possible electrical angular velocity $\omega_{e,max}$ is usually given by the manufacturer in the datasheet or it is limited by the Shannon sampling theorem as $\omega_{e,max} = \frac{\pi}{T_s}$. If the most restrictive C_{cl} is divided by 2 then the amplitude margin 2 for the worst case is guaranteed.

4 VELOCITY LOOP - PARAMETERS PROPOSAL AND ROBUSTNESS

The electromagnetic torque of IPMSM can be found e.g. in [1] as $T_e = \frac{3}{2} p i_q (\lambda_{pm} + (L_d - L_q) i_d)$. However, part of the equation which is related to i_d may be considered as a disturbance signal and torque can be controlled only by i_q as $\hat{T}_e = \frac{3}{2} p i_q \lambda_{pm}$. Where p stands for the number of pole pairs. Mechanical angular velocity ω_m can be then described by the motion equation as $\omega_m = \frac{1}{Js}\left(\hat{T}_e - T_{dist}\right)$. Where J stands for the total moment of inertia and T_{dist} stands for the disturbance torques which can be determined as $T_{dist} = \frac{3}{2} p (L_q - L_d) i_q i_d + sign(\omega_m)(B|\omega_m| + T_{dry}) + T_{load}$. Where T_{dry} is dry friction, B is a viscous friction coefficient and T_{load} is the torque load on the shaft of the machine.

If CL is robustly tuned and lower boundary of the R_s nearly corresponds with the identified value $\hat{R}_s$ then the closed-loop transfer function of the q-axis current may be approximated with the first-order system and open-loop transfer function of the velocity loop is formed as (6).

$$\Omega_m(s) = \frac{K_{p,\omega} 3p\lambda_{pm} (T_{i,\omega} s + 1) e^{-1.5 T_s s}}{T_{i,\omega} 2Js^2 \left(\frac{T_s}{C_{cl}} s + 1\right)} \left(\Omega_m^*(s) - \Omega_m(s)\right) \tag{6}$$

Parameters of the PI velocity controller are designed as (7).

$$T_{i,\omega} = 10\frac{T_s}{C_{cl}} \qquad\qquad K_{p,\omega} = \frac{20 J C_{vl}}{3p\lambda_{pm} T_s} \tag{7}$$

Where C_{vl} stands for the velocity loop scaling coefficient and integral time constant $T_{i,\omega}$ is placed to ensure slope $-20\ dB/dec$ for one decade.

Assuming parameters tuning (6) and using a substitution $\varepsilon = T_s \omega$ the open-loop transfer function can be written in a pseudo-angle representation as (8). Where the pseudo-angle of the system ε is related to the real frequency ω and sampling period T_s.

$$|F_{0,\omega}(j\varepsilon)| = \frac{C_{vl} C_{cl} \sqrt{100\varepsilon^2 + C_{cl}^2}}{\varepsilon^2 \sqrt{\varepsilon^2 + C_{cl}^2}} \qquad \Phi(\varepsilon) = -\pi + \arctan\left(\frac{10}{C_{cl}}\varepsilon\right) - \arctan\left(\frac{1}{C_{cl}}\varepsilon\right) - 1.5\varepsilon \tag{8}$$

Phase margin *PM* of this system is then defined as $PM = \pi + \Phi(\varepsilon)$. Solving equation $\frac{\partial PM}{\partial \varepsilon} = 0$ on intervals $C_{cl} \in (0, \frac{\pi}{9})$ and $\varepsilon > 0$ provides a pseudo-angle ε_{max} which ensures maximum phase margin

PM_{max}. Due to the $T_{i,\omega}$ placement and system characteristic PM_{max} is approximately limited by the interval $PM_{max} \in \langle 46°, 55° \rangle$. The lowest boundary of this interval $46°$ corresponds with the C_{cl} highest permissible value $\frac{\pi}{9}$. Then C_{vl} is calculated to ensure PM_{max} as (9).

$$\varepsilon_{max} = \sqrt{\frac{2C_{cl}\left(\sqrt{9801C_{cl}^2+14520C_{cl}+3600}-101C_{cl}-60\right)}{400}} \qquad C_{vl} = \frac{\varepsilon_{max}^2\sqrt{C_{cl}^2+\varepsilon_{max}^2}}{C_{cl}\sqrt{C_{cl}^2+100\varepsilon_{max}^2}} \qquad (9)$$

This design also ensures robustness to the velocity loop parameters changes. Moment of inertia J may change several times and the velocity loop will still be stable.

5 PARAMETERS TUNING VALIDATION - MEASUREMENT ON A REAL MACHINE

Proposed tuning was validated on the servo-motor with the following parameters:

series resistance $\hat{R}_s$ $[m\Omega]$	$6;\ G_{R_s} \in \langle 0.9, 1.5 \rangle$	moment of inertia J $[kg \cdot m^2]$	0.6
d axis inductance $\hat{L}_d$ $[\mu H]$	$145;\ G_{L_d} \in \langle 0.8, 1.2 \rangle$	max. velocity $\omega_{e,max}$ $[rad/s]$	10000
q axis inductance $\hat{L}_q$ $[\mu H]$	$200;\ G_{L_q} \in \langle 0.7, 1.2 \rangle$	sampling period T_s $[\mu s]$	100
magnets flux linkage λ_{pm} $[mWb]$	11.7	DC bus voltage U_{dc} $[V]$	200
number of pole pairs p $[-]$	10	max. allowed current I_{max} $[A]$	160

Table 1: Motor and inverter parameters

Controllers parameters were calculated as: $C_{cl} = 6.3 \cdot 10^{-3}$, $K_{p,d} = 9 \cdot 10^{-3}$, $K_{p,q} = 13 \cdot 10^{-3}$, $T_{i,d} = 24 \cdot 10^{-3}$, $T_{i,q} = 33 \cdot 10^{-3}$, $C_{vl} = 0.2 \cdot 10^{-3}$, $K_{p,\omega} = 67$ and $T_{i,\omega} = 0.16$. Control structure was discretized, enhanced with the antiwindup back-calculation and implemented into processor which drives the inverter. The setpoint of angular velocity was $\omega_m^* = 100\,rad/s$ and after $5\,s$ a torque load $10\,Nm$ was generated using dynamometer. Figure 2 presents measured characteristics.
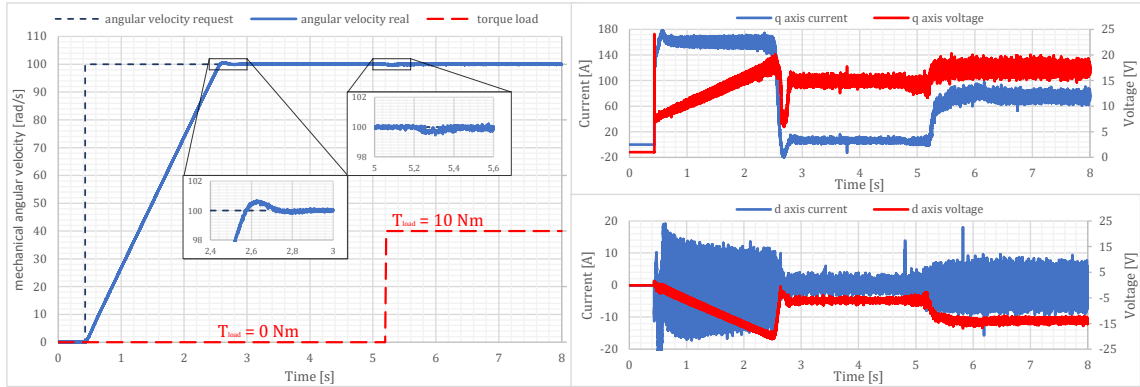
Figure 2: Measured mechanical angular velocity, $d - q$ currents and $d - q$ voltages

CONCLUSION

Analytical tuning strategy for IPMSM drive system was presented and designed parameters were used in the real IPMSM drive system. Proposed tuning ensures robustness to parameters changes and the drive system is stable even if winding faults are emulated. The proposed solution can be further improved by a flux weakening design and by tuning of the anti-windup tracking time constants.

REFERENCES

[1] S. K. Sul, Control of electric machine drive system. Hoboken, N.J.: Wiley-IEEE Press, 2011. ISBN 978-0-470-59079-9.

[2] G. Xingye, L. Chuang, Z. Yuefei and W. Kai, "Analysis and dynamic decoupling control schemes for PMSM current Loop," 2016 IEEE International Conference on Aircraft Utility Systems (AUS), Beijing, China, 2016, pp. 570-574, doi: 10.1109/AUS.2016.7748115.

TEXT DOCUMENT PLAGIARISM DETECTOR

Lukáš Kořínek

Master Degree Programme (5), FEEC BUT

E-mail: xkorin12@stud.feec.vutbr.cz

Supervised by: Petr Petyovský

E-mail: petyovsky@feec.vutbr.cz

Abstract: This paper provides an overview of diploma thesis concerned with research on available methods of plagiarism detection and then with design and implementation of such detector. Primary aim is to detect plagiarism within academic works or theses issued at BUT. The detector uses sophisticated preprocessing algorithms to store documents in its own NoSQL corpus. Implemented comparison algorithms are designed for parallel execution on graphical processing units and they compare a single subject document against all other documents within the corpus in the shortest time possible, enabling near real-time detection capabilities.

Keywords: text processing, parallelization, CUDA, NoSQL, C++

1 INTRODUCTION

Writing a paper, thesis, or simply any text available to general public demands that every idea coming from a different author has to be properly cited according to a widely accepted norm (e.g. ISO 690, [2]). Lack of citations deem the text to be a product of process known as plagiarism. Occurence of plagiarism is considered as a serious ethical issue, especially within grounds of the academia.

Therefore, it is vital to prevent releasing plagiarised texts to avoid any future scandals. A specialized software, plagiarism checker/detector, may help with the prevention, as its main function is to examine various textual bases and to measure if and how much are examined texts similar to each other. In general, plagiarism detectors can be based on one of three main principles. Intracorpal systems compare selected text against a well-known finite set of documents. Extracorpal systems search in open databases, typically utilizing web search engines. Intrinsic systems analyze stylistic features of a single text and try to find differences within its parts. A plagiarism detector may compare original fulltexts or data extracted from considered documents.

The plagiarism detection task itself is not an exact discipline and any of the approaches always results in a compromise between practical feasibility, speed of processing, quality of inference and flexibility of usage on various text sources. The essential part is human interaction with the system, because any of the findings (matches between documents) have to be evaluated by a human to avoid unfair accusations due to occuring false positives.

The diploma thesis designs and then develops a new plagiarism detection system for the BUT, where it should be used to compare submitted theses. Its design uses findings from some of the previous papers (see [4], [5]) to construct an optimal detection algorithm deliverable as a complete out-of-the-box solution, not as just a prototype. Main requirements are high speed comparison, acceptable quality of output and integrability into existing web infrastructure. The ultimate goal is to enable on demand plagiarism checking. Challenge in designing such a system is in the volume of data which has to be processed, because the expected database of documents (corpus) may have up to hundreds of thousands of documents. This scale requires parallelization on both the processor and graphical cards. Available system resources have to be handled as effectively as possible.

2 SOLUTION OVERVIEW

Implemented system is an introcorpal plagiarism detector that always compares a single selected document against all other documents within its own corpus.

The system reads texts in PDF format and BUT theses identified by their ID (available on an internal API). Document indexation (addition into the corpus) requires each text to pass a series of eleven pre-processing steps. Data formed by the preprocessor is then suitable for fast comparison. Preprocessing involves text parsing, removal of cited sections and words without informative value, tokenization (splits the text into a list of words) and lemmatization (replaces words with their uninflected form). Not only document content gets saved, but also other relevant metadata, like timestamps, information about its author and page counts.

The processed document contents are stored as so called n-grams, or groups of N adjacent words forming a sequence based on the original text. These n-grams are represented as numeric indices into n-gram tables persisted within application memory. Viability of the described approach had been previously tested in thesis [4]. Usage of numeric values for the n-grams improves detection performance, because comparison of numbers is less compute-intensive than comparison of strings.

```
Tabs in your browser. => 'tabs|browser', 7082612
```

Illustrative text and a corresponding n-gram table record (2-grams).

Detection task (document comparison and matches extraction) is highly paralellized and the selected text gets compared against batches of up to thousands of documents. One after another, batches are loaded from the corpus and searched for matches. Obtained matches from every batch are then joined together. Furthermore, compared documents in batches are divided into multiple smaller chunks to further speed up match searching.

Matches extraction involves comparison of documents' n-grams and a subsequent clustering based on a min-max distance metric. Matched clusters (typically on a scale of paragraphs) are then filtered by thresholding both their size and equality of their contents. Line and page markers are used to locate the match within original document fulltexts. Resulting clusters are grouped by compared documents and presented to the user. A web-based graphical user interface is also included.

Figure 1: Graphical user interface of the detection system.

3 SOFTWARE ARCHITECTURE

The plagiarism detector solution includes following components:

Application service The service makes a key element of the solution. Written in C++, it provides an interface for all implemented functions of the detector, like document indexation and match searching. It is also responsible for management of available resources. Output of the service can be obtained in either line-delimited JSON objects or a human readable structured text. Parameters of the service are configurable in a standalone file.

API client The client is a slim web application written in NodeJS. It provides a REST API and a web interface for tasks run by the service. Communication between the client and the service is based on TCP sockets. The expectation is that a user or supersystem will communicate with the service solely throughout this client.

Document corpus NoSQL MongoDB database containing indexed documents and their metadata.

Additional data memory Also handled by MongoDB, persists data critical for indexation tasks.

All the functions of the application are available in form of callable Tasks. Default thread of the application performs only task scheduling, while every called task gets executed on its own worker processor thread. According to their needs, specific tasks may then utilize creation of new subtasks or graphical processing unit threads based on Nvidia CUDA platform. The graphical threads are used not only for the match searching, but also for some of the preprocessor steps.

In order to standardize hosting environment and installation of the solution, application's components have been containerized facilitating *docker* and *podman* services. The application has been designed to run on a UNIX-based operating system with an Nvidia graphics card and a multicore processor.

4 TESTING

A set of 11 documents has been created for qualitative testing of the detector (forming a so called annotated corpus). These documents contain both copy&paste plagiarisms (identical copies) and paraphrases. Comments attached to the documents annotate source and placement of plagiarised sections.

Apart from this set, a corpus of real BUT theses from years 2018 and 2020 has been indexed. In total, for testing there were 4 498 documents consisting of almost 9 million unique n-grams (2-grams; preliminary tests on longer n-gram variants seemed to perform worse in detection of paraphrases). Length of the documents varied between 1 and 300 pages. The comparison was performed in 10 batches, each of them utilizing 32x256 graphical threads (optimal combination found for testing hardware Nvidia GTX 1660S).

document	page count	time [s]
samples/sample_2.pdf	37	18
samples/sample_1.pdf	56	29
internal://2018:112346	158	65

Table 1: Time-oriented test results.

Found matches fully cover copy&paste plagiarism in the annotated database, most of the paraphrases and a noticeable number of false positives. The table above doesn't show found match counts, because their inclusion would be meaningless without more details. Improvements in indexation and optimalization of threshold parameters would help to eliminate some of the false positives found.

5 CONCLUSION

A new software solution for plagiarism detection has been designed and implemented. The solution is suitable for integration into BUT systems. Deployed system may replace or accompany an existing solution used at the moment of writing this paper. A possible arrangement might be, that a detection on newly turned-in theses would execute two plagiarism detection softwares in series, where one of them is faster and the other one has better result quality.

However, the solution is independent on university infrastructure and can be applied to any textual match searching task in situations where a finite corpus is known and the languages concerned use the latin script. The solution is delivered in form of a container cluster, enabling it to be deployed in a cloud environment.

As for the next steps, further speed up of the match searching algorithm would make the system usable in an on-demand basis. To achieve that, brute force approach in bringing multiple and more powerful graphical processing units, or memory access optimizations (enabling maximalization of threads run on every graphics card) may both be used. Improvements in preprocessing steps' selectivity and match searching process parameters would help eliminate most of the false positives.

In the future, an extension of the application searching for keywords in online sources is possible. Inclusion of web pages in the corpus would increase the likelihood of a successful match for plagiarisms which do not come from other theses, but from texts available on the web.

REFERENCES

[1] *ANON. Zákon č. 121/2000 Sb.: Zákon o právu autorském, o právech souvisejících s právem autorským a o změně některých zákonů (autorský zákon)*. In: 121/2000. 2000, 36/2000. Available from: https://www.zakonyprolidi.cz/cs/2000-121

[2] *BIERNÁTOVÁ, Olga a Jan SKŮPA. Bibliografické odkazy a citace dokumentů: dle ČSN ISO 690 (01 0197) platné od 1. dubna 2011 [online]*. 2011, , 27 [cit. 2020-10-01]. Available from: https://www.citace.com/CSN-ISO-690.pdf

[3] *POTTHAST, Martin. Overview of the 6th International Competition on Plagiarism Detection [online]*. 2014, 32 [cit. 2020-09-28]. Available from: http://ceur-ws.org/Vol-1180/CLEF2014wn-Pan-PotthastEt2014.pdf

[4] *KOBATH, M. Detekce plagiátů*. Brno: Vysoké učení technické v Brně, Fakulta elektrotechniky a komunikačních technologií, 2015. 59 s. Vedoucí bakalářské práce Ing. Petr Honzík, Ph.D.

[5] *PŘIBIL, Jiří. Efektivní metody detekce plagiátů v rozsáhlých dokumentových skladech*. Jindřichův Hradec, 2010. Dizertační práce. Vysoká škola ekonomická v Praze - Fakulta managementu v Jindřichově Hradci. Vedoucí práce Radim Jiroušek.

[6] *CUDA Zone - NVIDIA developer [online]*. 2021 [cit. 2021-03-24]. Available from: https://developer.nvidia.com/cuda-zone

NEGATIVE ELECTRODE FOR SODIUM-ION BATTERIES

Antonín Šimek

Master Degree Programme (1), FEEC BUT

E-mail: xsimek31@vutbr.cz

Supervised by: Jiří Libich

E-mail: libich@feec.vutbr.cz

Abstract: This thesis is mainly focused on research of electrode materials for negative electrodes of sodium-ion batteries, which do not use lithium for their function. It is mainly focused on the synthesis and characteristics of sodium metatitanate, which is used as active material in negative electrode mass.

Keywords: negative electrode, sodium titanate, theoretical capacity, energy storage, sodium-ion battery, renewable energy

1   INTRODUCTION

Today, battery research is a very popular topic, as they are used in almost every area of human activity. Lithium-ion batteries are used for most of this applications today, due to their large capacity and proven functionality. However, their problem is the limited amount of lithium in the earth's crust and high cost of lithium as a material. Therefore, alternatives to batteries that do not use lithium are being sought. Promising candidates are sodium-ion batteries, which could replace lithium-ion batteries in some areas. They use sodium for their function, which is abundant in the earth's crust and its price is not nearly as high as in the case of lithium.

The purpose of this work is synthesis of suitable sodium titanate, preparation of negative electrodes and their characterization in order to find a potential successor of some lithium batteries.

2   SODIUM-ION BATTERIES

Sodium-ion batteries were developed in parallel with lithium-ion batteries in the 1970s and 1980s, but their research was almost interrupted due to their lower capacity and more complex technological processes. In the current research, we come out from proven electrode materials, which work reliably in the case of lithium-ion batteries. However, when implementing sodium in these electrode materials, we are getting relatively different results because sodium ions are much larger than lithium ions and the resulting capacity is much smaller. However, the operating principle of sodium-ion batteries together with their basic parts is very similar to that of lithium-ion batteries. [1]

2.1 PROPERTIES OF SODIUM

The main disadvantages of sodium are larger dimensions and greater weight compared to lithium. The relative atomic weight of sodium is approximately 22.99 $g.mol^{-1}$, for lithium it is only 6.94 $g.mol^{-1}$. The potential of sodium is 0.33 $E°.V^{-1}$ lower than that of lithium. Due to the greater weight, the theoretical capacity of sodium is significantly lower compared to lithium. The theoretical capacity of lithium is 3829 $mA.h.g^{-1}$, for sodium it is only 1165 $mA.h.g^{-1}$. The ionic radius of sodium is 102 pm, for lithium it is 76 pm. For this reasons, it is very difficult to find a suitable crystalline material that would accept sodium ions and at the same time show sufficient capacity and stability during cycling. [2, 3]

3 SYNTHESIS OF SODIUM TITANATES

The synthesis of sodium titanates was carried out by the sol-gel method, i.e. by converting the colloidal suspension into a high-purity gel, from which an oxide material in the form of a powder is formed during high-temperature sintering. The basic precursors for the synthesis of sodium titanate were: absolute ethanol (CH_3CH_2OH), Ti(IV)-ISOPropoxide ($C_{12}H_{28}O_4Ti$) and sodium carbonate (Na_2CO_3).

The preparation procedure was as follows:
1. Using a pipette, dispense 30 ml of ethanol into a conical flask (according to Erlenmeyer).
2. Place the flask on a hot-plate, set the temperature to 90 °C and the stirring speed in the range of 120-250 rpm. Then wait for 15 minutes for ethanol to warm up.
3. Using a pipette, add 3.3 ml of Ti-ISOPropoxide to the flask and wait another 15 minutes.
4. Weigh 1,98 g of sodium carbonate and add to the flask. Wait 60 minutes for the individual substances to react.
5. Pour the resulting substance into an sintering dish and place it in to a vacuum oven.
6. Set the following parameters on the oven: temperature 800 °C, time 5 hours (300 minutes), temperature rise 3 °C.min^{-1}, air atmosphere.

3.1 SYNTHESIS OF INDIVIDUAL SAMPLES OF SODIUM TITANATES

During the synthesis, a total of 12 samples were prepared. The preparation of each sample was based on the above procedure. Different ratios of basic substances were chosen for prepared samples in order to achieve the highest possible percentage of $Na_2Ti_3O_7$ phase in the final material, and to determine the amount of colloidal suspension, which can be reliably sintered in the laboratory furnace with sufficient purity. The prepared titanate should be a snow-white powder. Based on this information, it was therefore possible to determine whether the sample did not contain a large amount of carbon and achieved the required purity. Table 1 summarizes all the samples prepared and the amounts of starting materials used in their synthesis.

Identification	Absolute ethanol [ml]	Ti-ISO Prop. [ml]	Na₂CO₃ [g]	Sintering dishes [pcs]	Colour of the sample
1a, b, c, d	90	9,9	5,94	3	×
2a	30	3,3	1,98	1	Pure white
3a	30	9,9	5,94	1	Light gray
4a	30	9,9	1,98	1	Pure white
5a	60	19,8	3,96	2	Black
6a, 7a, 8a	30	9,9	1,98	1	Dark grey
9a	30	9,9	1,98	1	Pure white

Table 1: Summary of prepared samples and the amount of starting materials.

4 CHARACTERIZATION OF PREPARED SAMPLES

First, X-ray Powder Diffraction (XRD) was performed on Samples 1c, 2a, 3a and 4a. The reason for selecting these samples was their pure white colour and the assumption of a high content of sodium titanates without other impurities. Sample 4a was the best from this analysis. This sample was consists of 51,1 mol. % Na2Ti3O7 (sodium metatitanate) and of 46.6 mol. % Na8Ti5O14 compound. Sample 4a was therefore used for further analysis.

For further characterization, it was necessary to obtain a sufficient amount of sodium titanate. The same method of preparation sample 4a was repeated four times, thus obtaining a sample designated 9a in a total amount of 1.38 g. Scanning Electron Microscopy (SEM) as well as SEM elemental analysis were used. The result of this analysis was the following composition: 60.61 at. % of oxygen (O), 12.43 at. % of sodium (Na), 19.59 at. % of titanium (Ti) and 7.37 at. % of carbon (C).

5 PREPARATION OF ELECTRODES

A total amount of 0.5 g of electrode mass was prepared, where 10 wt. % was binder PVDF (polyvinylidene fluoride), 20 wt. % was seductive additive super C65 and the remaining 70 wt. % was sodium titanate (sample 9a). After preparation of the electrode mass, an 80 μm thick layer was applied to the polished side of a 15 μm thick copper foil. Then it was placed in an dryer at 60 °C for about 3 days.

Subsequently, the electrodes themselves were punched with a 16 mm diameter punch, which were then pressed at a pressure of approximately 25 kN for 5 seconds. The pressing process was followed by re-drying of the electrodes, but this time in a vacuum dryer at 50 °C and a pressure of 0.9 bar for about 2 days.

After drying, the electrodes were assembled into electrochemical measuring cells. A sodium wheel with a diameter of 16 mm and a thickness of approximately 1 mm was first placed in the cells. After that the porous separator with a diameter of 18 mm was added. This was followed by the addition of 170 μl of liquid electrolyte type EC/DEC (cell no. 3 and no. 4) and type EC/PC (cell no. 2). Finally, the prepared electrode was added. The assembly took place in a glove box in the presence of an argon protective atmosphere.

6 RESULTS OF MEASUREMENTS

The measurement was performed on a multi-channel Bio-Logic potentiostat. First, an open circuit voltage (OCV) measurement was performed for 24 hours. This voltage stabilized at values in the range of 2.51 to 2.89 V for individual cells. Subsequently, Cyclic Voltammetry (CV) was performed. The plot of CV is shown in Figure 1a. Peaks A and B lying below the x-axis are located in the reduction region, where the electrode mass is charged, i.e. sodium cations are intercalated into the structure. Peaks C and D lying above the x-axis are located in the oxidation region, where the electrode mass is discharged, i.e. sodium ions are deintercalated from the structure. The measurement was compared with the measurement listed in the article [3], which we consider to be the reference. When comparing these dependences, peaks were found that lie at approximately the same potential as in the case of our measurement. Peaks A and D represents the reduction and oxidation of the $Na_2Ti_3O_7$ phase. The remaining two peaks B and C then represents the reduction and oxidation of the $Na_8Ti_5O_{14}$ phase.

The second measurement procedure was Galvanostatic Cycling with Potential Limitation (GCPL), which was based on CV measurements. From the theoretical capacities obtained during the CV measurement (approximately 90 mA.h.g^{-1}), the hourly current was subsequently calculated as 1C (approximately 0.21 mA.h). The electrode discharge was chosen to be 0.1C (current approximately 0.21 μA) to avoid unnecessary electrode overloading. This measurement showed a decrease in discharge capacity during cycling, which corresponds to be approximately 1.2 % for each cycle. The results of this measurement are shown in Figure 1b. Furthermore, horizontal sections in GCPL were measured, whose potential difference is approximately 0.2 V and which are proof of a two-phase mechanism according to the equation: $Na_2Ti_3O_7\ 2\ Na^+\ 2\ e\ \rightleftharpoons\ Na_4Ti_3O_7$. The GCPL results for the first two cycles are shown in Figure 1c.

The last measuring procedure was the measurement of Rate Capability (RC). In this measurement, a decrease in the discharge capacity was observed with increasing load due to the slow kinetics of the ion transition. However, the discharge capacity increases again with decreasing load, which is evidence of reversible electrochemical processes. The results of this measurement are shown in Figure 1d.
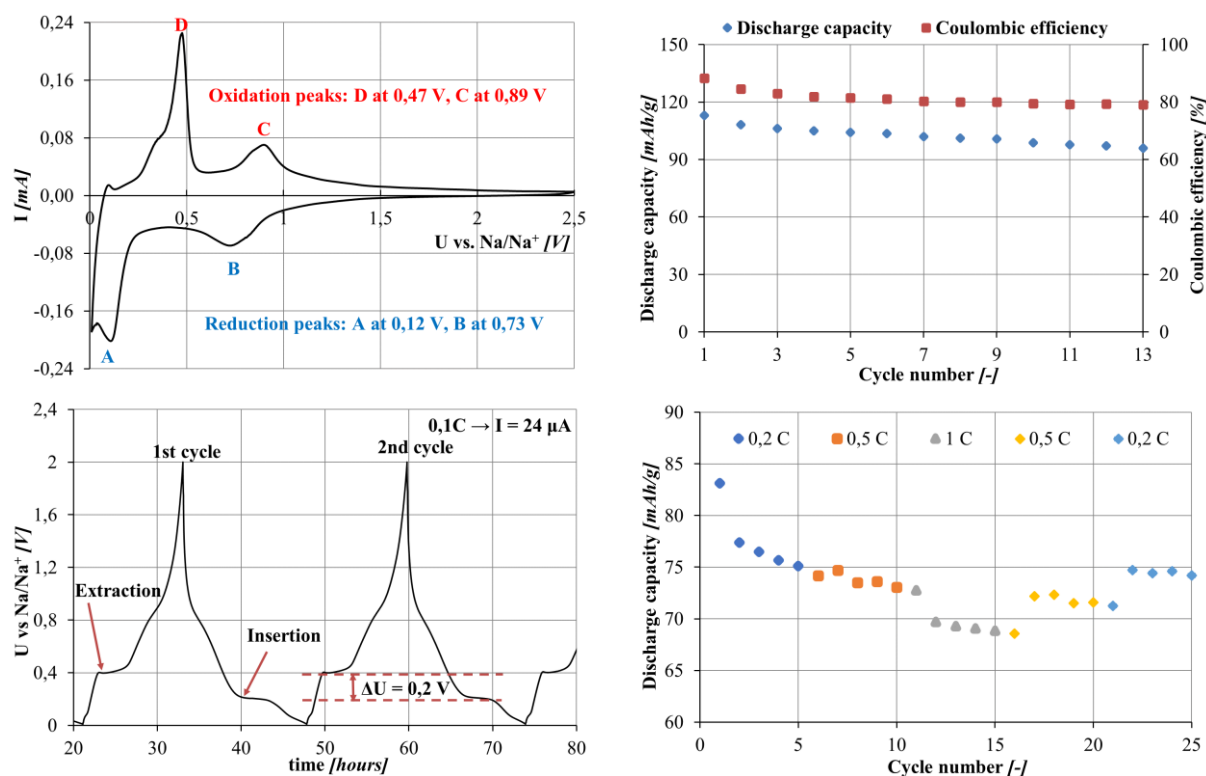
Figure 1: The results of measurement: (a) Cyclic voltammetry, (b) and (c) Galvanostatic Cycling with Potential Limitation, (d) Rate Capability.

7   CONCLUSION

The purpose of the work was to synthesize a suitable sodium titanate, to create negative electrodes and to perform their basic electrochemical characterization. The prepared electrodes showed good reversible capacity, but with a significant decrease in capacity during cycling. Nevertheless, this material is a promising candidate for the future development of sodium-ion batteries. The overlap of this work is to identify and eliminate the cause of the decrease in capacity during cycling, prepare phase-pure sodium titanate and ensure the simplest and most efficient synthesis, which could eliminate the shortcomings of existing samples..

REFERENCES

[1]   Chem Soc Rev. ROYAL SOCIETY OF CHEMISTRY [online]. 2017, 46(12), 3485-3856 [cit. 2018-12-14]. ISSN 0306-0012. Available from: https://pubs.rsc.org/is/content/articlepdf/2017/cs/c6cs00776g

[2]   ŠÁTEK, Dominik. *Negative electrodes for sodium-ion batteries*. Brno, 2019. Bachelor thesis. Brno University of Technology, Faculty of Electrical Engineering and Communication. Department of Electrotechnology. Thesis supervisor Ing. Josef Máca, Ph.D.

[3]   KULOVA, T. L., Y. O. KUDRYASHOVA, A. A. KUZ'MINA, A. M. SKUNDIN, I. A. STENINA, A. A. CHEKANNIKOV, A. B. YAROSLAVTSEV a J. LIBICH. Study of degradation of Na2Ti3O7-based electrode during cycling. *Journal of Solid State Electrochemistry* [online]. 2019, 23(2), 455-463 [cit. 2020-06-04]. DOI: 10.1007/s10008-018-4154-1. ISSN 1432-8488. Available from: http://link.springer.com/10.1007/s10008-018-4154-1

GRAPHENE FIELD EFFECT TRANSISTOR PROPERTIES MODULATION VIA MECHANICAL STRAIN INDUCED BY MICRO-CANTILEVER

Jan Brodský

Master Degree Programme (2), FEEC BUT

E-mail: jan.brodsky@ceitec.vutbr.cz

Supervised by: Imrich Gablech

E-mail: imrich.gablech@ceitec.vutbr.cz

Abstract: This work presents a new method, which enables the electrical characterization of graphene monolayer with induced mechanical strain. The device is a combination of two-dimensional field effect transistor (2DFET) and a MEMS cantilever, both of which can be used to alter graphene properties. The first method applies external electric field to the graphene monolayer. The second method is based on mechanical bending of the cantilever by external force, which induces mechanical strain in the characterized layer. By sweeping the gate voltage (V_{GS}) in range from -50 V to $+50$ V and measuring the current between drain and source (I_{DS}) with fixed drain-source voltage (V_{DS}) at 1 V, Dirac point of graphene is found at ≈ 9.3 V of V_{GS}. After bending of the cantilever, the sweep is performed again. The induced strain shifts the position of the Dirac point by ≈ 1.3 V to $V_{GS} = 8$ V. Because the fabrication process is compatible with silicon technology, this method brings new possibilities in graphene strain engineering.

Keywords: graphene, Dirac point, MEMS, FET, cantilever, mechanical strain

1   INTRODUCTION

Graphene is a carbon material with hexagonal structure, ideally a monolayer. It has been under extensive research since its first preparation in 2004. Because of its unique properties, such as exceptional chemical stability, mechanical flexibility and high charge carrier mobility, it is a promising material for nanoelectronics. Graphene is a semiconductor with zero bandgap, valence and conduction band meet at Dirac points. It is possible to tune the bandgap, e. g. by application of external electric field, doping or mechanical strain engineering [1]. Mechanical strain inducing is advantageous, because no absorption of secondary substances is involved (doping) and the process is fully recoverable, leaving no permanent change in the graphene [2].

2   DEVICE DESIGN AND FABRICATION

2.1   DESIGN

Using software tool CNST Nanolithography Toolbox together with Python, parametrized layout was created. The dimensions of the chip were (6×6) mm^2, it had four sets of cantilevers, with respective lengths of (100, 200, 300 and 400) µm and width of 60 µm. The total number of cantilevers was 64, while each cantilever has its own source and drain electrodes with varying channel length from 1 µm to 50 µm. The width of the graphene was set to 10 µm. Silicon-on-Insulator (SOI) wafer was used as the substrate. The gate electrode was contacted from the top by an opened window in SiO$_2$ layer. The entire layout is shown in **Figure 1**.

Gate | Topside cantilever shaping | Graphene
Cr/Au conductive paths | Backside cantilever etching
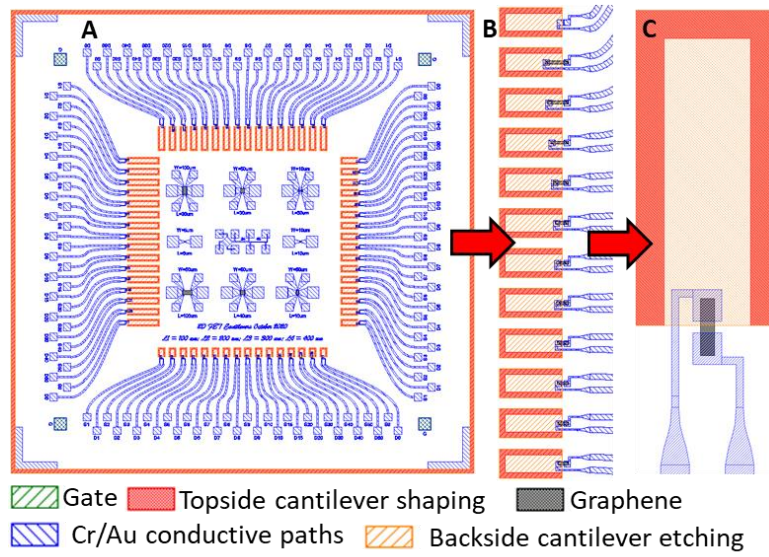
Figure 1: (A) Design of the whole chip with dimensions of (6×6) mm^2; (B) Set of cantilevers with electrodes; (C) Detail of a single cantilever.

2.2 FABRICATION

The whole process consisted of 5 lithography steps (**Figure 2**). Thin layer of SiO$_2$ with thickness of ≈ 90 nm, serving as the gate insulator, was created on the SOI wafer by thermal oxidation process employing O$_2$.

Top SiO$_2$ | Si | Bottom SiO$_2$ | Graphene | Au
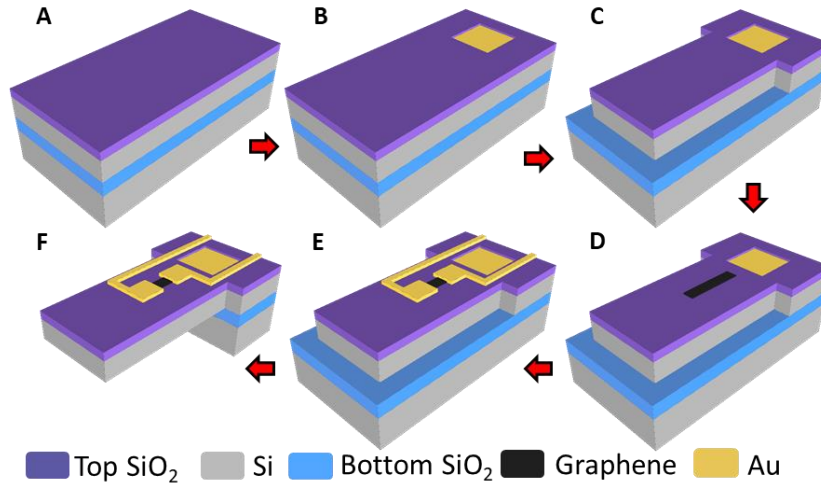
Figure 2: Schematic representation of the fabrication process. (A) SOI substrate; (B) Gate contact; (C) Topside cantilever shaping; (D) Graphene shaping; (E) Electrode lift-off; (F) Shaping cantilevers from the bottom.

During the first lithography, window in the SiO$_2$ was created by reactive ion etching (RIE), followed by the deposition of Au layer to create the gate contact. In the second step, cantilevers were shaped from the top side by Bosch deep reactive ion etching (DRIE) of the ≈ 3 µm thick device Si. Afterwards, wafer was properly cleaned in NMP solvent, rinsed in IPA, and finished by O$_2$ plasma cleaning. Such prepared wafer was used for wet-transfer of CVD grown graphene. In third lithography step, graphene was shaped by covering the pattern in PMMA/AZ 5214 e-beam and photoresist combination and etching in O$_2$ plasma by RIE. PMMA served only as a protective layer for the graphene. In following step, lift-off of Cr/Au electrodes by e-beam evaporation was realized. In the final step, lithography was performed from the backside of the wafer. Using DRIE once more,

the handle Si (≈ 500 µm, bulk silicon) was etched, shaping the cantilevers from the bottom side of the SOI wafer. The final release was also achieved by SiO_2 dry etching. As seen in **Figure 1**A in red, the chips were separated by the DRIE, meaning no dice cutting was necessary. The fabricated chip is shown in Figure 3.
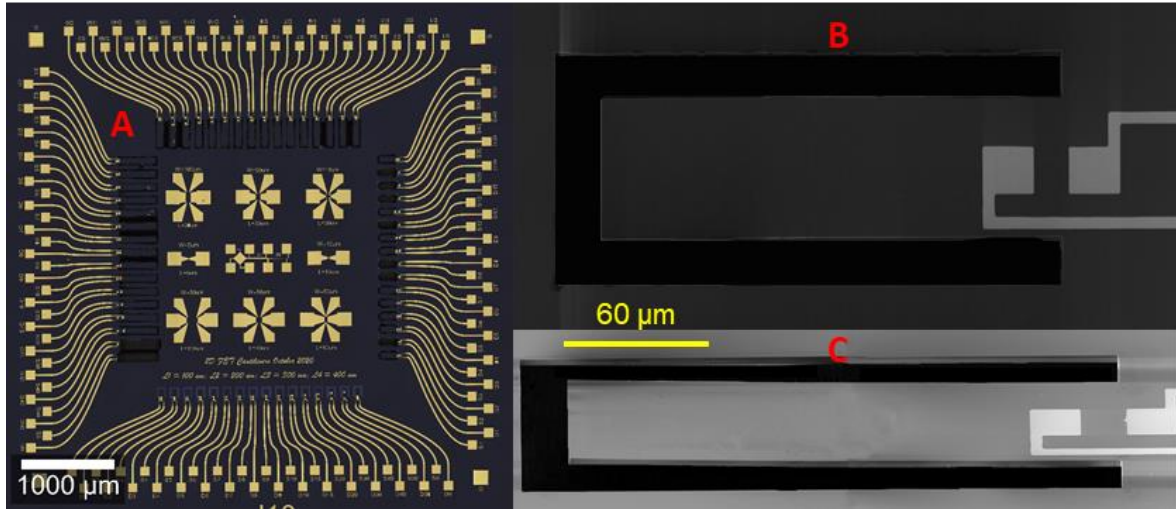
Figure 3: (A) Optical image of the whole chip; (B) SEM image of cantilever from top side; (C) SEM image of cantilever tilted by 55 °C.

3   EXPERIMENTAL

The individual chips were first cleaned in NMP to remove photoresist residues. Using Raman spectroscopy, it was confirmed that the transferred graphene (Figure 4) is a monolayer [3]. Prior to electrical characterization, the samples were cleaned in acetone and then annealed in vacuum furnace at ≈ 150 °C for 36 hours with low temperature ramp of ≈ 3 °C·min^{-1}. Dirac point of pristine graphene should be at value of $V_{GS} \approx 0$ V. However, when exposed to air, the graphene becomes p-doped and the Dirac point moves towards higher values of V_{GS}. The cleaning of graphene samples in acetone and annealing partly restores the original position of Dirac point. Using probe station MPS150 and parameter analyzer Keithley 4200A-SCS, V_{GS} was swept in range from -50 V to $+50$ V. With fixed voltage V_{DS}, current I_{DS} was measured. The sample was placed in N_2 atmosphere during the measurement, to prevent the doping of the graphene resulting in unwanted shift of the Dirac point. The V_{GS} sweeps were performed three times for each sample, before the bending of the cantilever, with the cantilever bent down and finally with the cantilever returned to its original position.
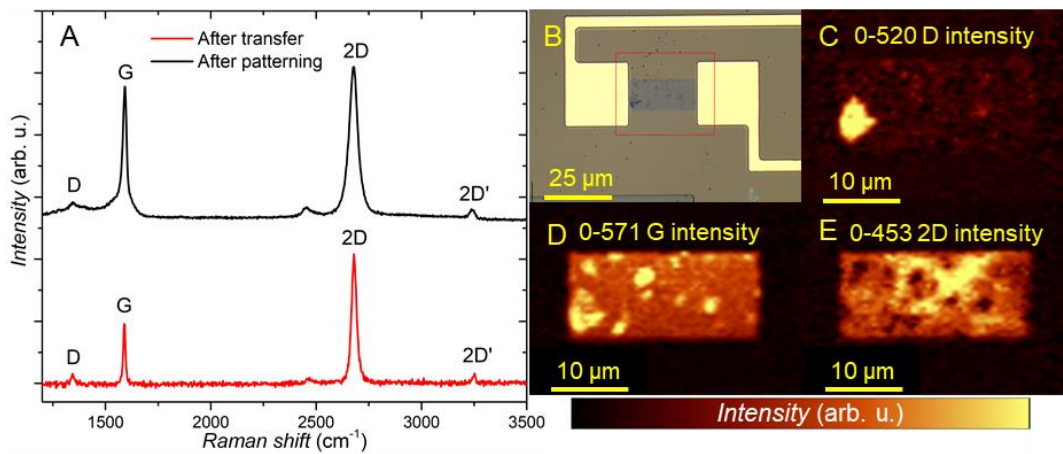
Figure 4: Raman spectra of transferred and patterned graphene: (A) single spectra; (B) optical image of electrode with graphene; Raman map of (C) D peak; (D) G peak; (E) 2D peak.

4 RESULTS

By running the first sweep, Dirac point voltage of slightly p-doped graphene was found at $V_{GS} \approx 9.3$ V (Figure 5). When the cantilever was bent down, the Dirac point voltage shifted by 1.3 V from $V_{GS} \approx 9.3$ V to $V_{GS} \approx 8$ V. For the third sweep, the cantilever was returned to its original position. After the cantilever returned to its initial position, the Dirac point voltage was almost the same as original value with only slight difference of ≈ 0.2 V. The bending process was repeated with other five structures, which show similar behavior with negligible Dirac point voltage differences.
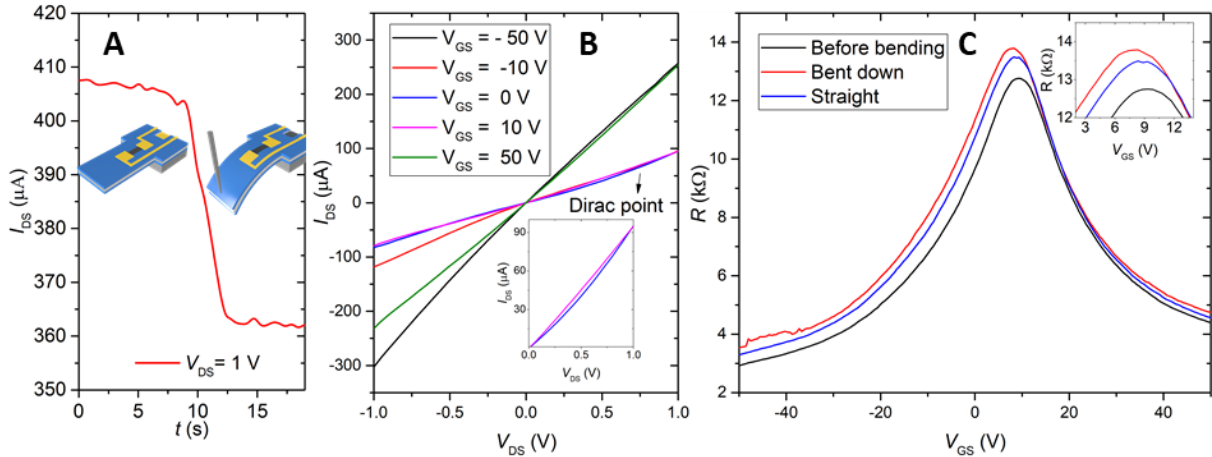
Figure 5: Dirac point voltage shift with induced mechanical strain. (A) Bending of the cantilever in time shows significant decrease of I_{DS}; (B) I-V characterization with V_{GS} as parameter, hinting the Dirac point voltage (C) Shifting of the Dirac point due to induced mechanical strain by bending.

5 CONCLUSION

In this work, a novel technique for controlling graphene properties through two methods employing electrical field and mechanical strain was introduced. The shift of Dirac point voltage of graphene was achieved by inducing mechanical strain into the monolayer. This opens up new possibilities in graphene strain engineering directly on Si substrates used in integrated circuits. Promising outlook is the possibility to control the stress of deposited films (e. g. ion-beam sputtering using Kaufman source) directly during chip fabrication process which will lead to required strain in graphene after structure release.

REFERENCES

[1] D. R. Cooper *et al.*, "Experimental Review of Graphene," *Condens. Matter Phys.*, vol. 2012, 2012, doi: 10.5402/2012/501686.

[2] G. Hu *et al.*, "Controlling the Dirac point voltage of graphene by mechanically bending the ferroelectric gate of a graphene field effect transistor," *Mater. Horizons*, vol. 6, no. 2, pp. 302–310, Feb. 2019, doi: 10.1039/c8mh01499j.

[3] J. Bin Wu, M. L. Lin, X. Cong, H. N. Liu, and P. H. Tan, "Raman spectroscopy of graphene-based materials and its applications in related devices," *Chemical Society Reviews*, vol. 47, no. 5. Royal Society of Chemistry, pp. 1822–1873, Mar. 07, 2018, doi: 10.1039/c6cs00915h.

ENERGY DISPERSIVE X-RAY SPECTROSCOPY OF DOPED PVDF FIBERS

Tereza Smejkalová

Master Degree Programme (2), FEEC BUT

E-mail: xsmejk25@vutbr.cz

Supervised by: Dinara Sobola

E-mail: sobola@vutbr.cz

Abstract: This work aims to further improve properties of Polyvinylidene fluoride (PVDF), one of the most promising electroactive polymers, by the inclusion of powders of piezoactive materials. PVDF was formed by electrospinning into fibres with a thickness of $0.5 - 1.5\mu m$ and then examined in a scanning electron microscope including energy-dispersive X-ray spectroscopy. The obtained properties of doped PVDF could be used in the design of sensors. Before that, it is essential to perform a series of analysis to support or deny the use of adjusted fibres for sensor design.

Keywords: PVDF, energy dispersive X-ray spectroscopy, scanning electron microscopy, electrospinning, composition, morphology, piezoelectricity

1 INTRODUCTION

PVDF is a semicrystalline polymer having five crystalline polymorphs, including α-, β-, γ- δ- and ε-phases. The β-PVDF-phase exhibits excellent ferroelectric and piezoelectric properties, making PVDF useful in a wide range of applications (including actuators, biosensors, energy harvesting materials, audio devices and many more). Piezoelectric materials generate an electric charge in response to applied stress or slight mechanical deformation, thus eliminating the need for external power sources for electrical stimulation. Therefore, it is possible to obtain a high dielectric constant by doping PVDF with piezoelectric ceramics. As a result, polymer-ceramic composites can be an excellent choice to achieve miniaturisation of energy storage devices.

The fibres were doped by five types of piezoelectric ceramics. This paper, because of its required length, focuses only on two resulting composites – on fibres doped by barium calcium zirconate titanate $((Ba,Ca)(Zr,Ti)O_3$, BCZT) and barium titanate (BT). Both of the ceramics are lead free, environmentally friendly ferroelectric materials applicable in a variety of applications, due to its excellent dielectric, ferroelectric and piezoelectric properties. [1]

2 ELECTROSPINNING

For the preparation of PVDF fibres doped by ceramics was used the method called electrospinning. Electrospinning is a simple method to produce nanoscale fibres both in a laboratory and industrially. The fabrication process starts with the application of a high electric field to the polymer or solvent solution. A Taylor cone is formed at the surface of the polymer solution extrusion and above critical voltage electrostatic repulsion overcomes the surface tension of polymer droplet. The charged polymer jet is then ejected from the needle tip towards a grounded collector, forming randomly oriented fibres. The droplet is refilled by pumping new polymer solution to the needle tip through the needle.

The fibre diameter and porosity of the fibrous scaffolds depend on the processing parameters such as applied voltage, solution flow rate, type of solvent, polymer concentration in the solution, and the distance between the needle and collector. Electrospun fibres can be aligned by controlling the

fabrication parameters. Because of the chaotic trajectory of the polymer jet, the fibres collected on a grounded collector generally exhibit random orientation.

collector voltage	50 kV
distance between the needle and collector	20 cm
collector speed	300 rpm
duration of the process	100 min

Figure 1: The electrospinning processing parameters for manufacturing PVDF fibres doped by piezoelectric ceramics BCZT and BT

3 SCANNING ELECTRON MICROSCOPY

The morphology of the nanofibers was characterized by scanning electron microscopy. A SEM microscope uses a fine beam of focused electrons to scan a sample's surface; the process is based on applying kinetic energy to produce signals on the interaction of the electrons. [2] Electrons reflected by the specimen are used to form a magnified, black and white three-dimensional image.

When working with a scanning electron microscope, proper sample preparation plays an important role. While metals do not require complicated preparation, non-metals need to be coated with a conductive material to interact with electrons. [3] To ensure conductivity, resp. for charge dissipation, samples were sputtered with a thin layer of gold or carbon (10 nm) before the analysis. The images were taken on a Tescan Lyra 3 scanning electron microscope at an accelerating voltage of 5 or 10 kV.

(a)   (b)
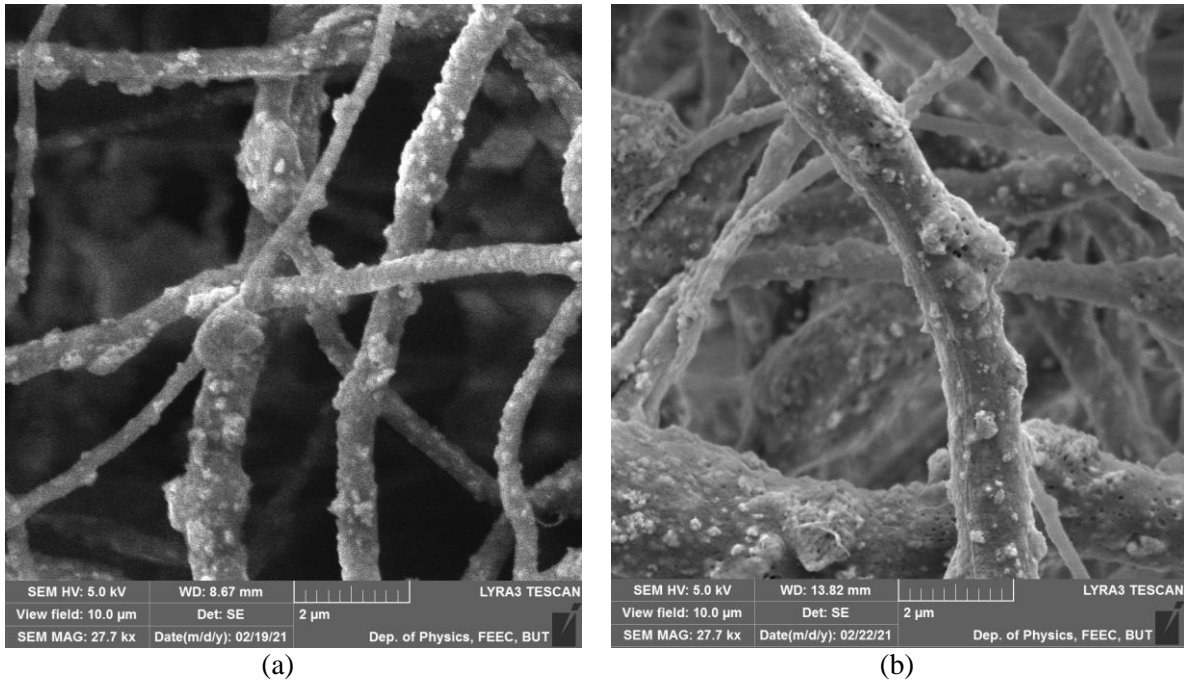
Figure 2: SEM micrograph of PVDF doped by (a) BCZT (b) BT

The fibres have an inhomogeneous diameter; the thin fibres are predominantly smooth, while the wider fibres have a rough texture with a layered structure. It suggests that the ceramic particles have been incorporated into the fibres; however, to support this argument, EDX analysis must be performed.

4 ENERGY DISPERSIVE X-RAY SPECTROSCOPY

The presence of ceramic particles in the fibres were confirmed by mapping using energy dispersive spectroscopy (EDX mapping, Tescan Lyra 3, AZtec software). The measurement was performed at

an accelerating voltage of 10 kV and energies of 0–10 keV were measured. For sample A (doping by BCZT), it was barium, calcium, carbon, fluorine, oxygen, titanium, and zirconium, and for sample B (doping by BT), it was barium, carbon, fluorine, oxygen and titanium. The aim was to create as homogenous composite as possible, where properties are uniform throughout the fibre, i.e. they do not depend on the position inside the fibre body.

Figure 2 presents a map analysis of the major elements; these elements are colour-coded by EDX software. It is visible that the added ceramic relatively homogeneously copies the structure of PVDF nanofibers, which means that the additive is homogeneously dispersed in the polymer fibres formed by electrospinning; thus, the desired result was obtained. Similar results were obtained for the remaining samples of nanofibers (TiO2, ZnO, KNN), the analysis of which is not part of this article.

(a)

(b)
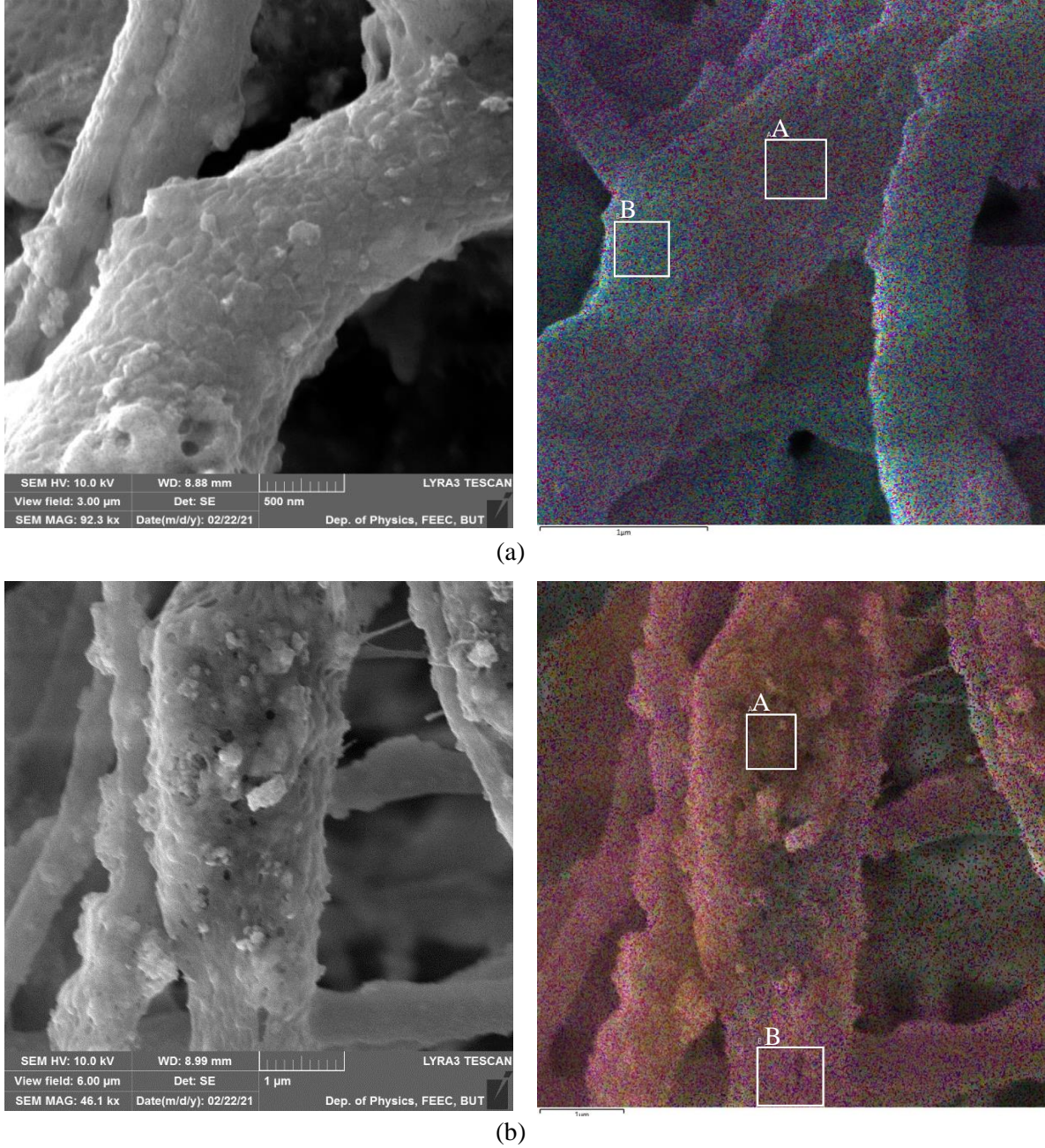
Figure 2: SEM and EDX micrographs of PVDF fibres doped by (a) BCZT (b) BT

EDX analysis also presents a spectrum that plots the number of X-rays detected versus their energies. In other words, it displays the peaks correlated to the elemental composition of the investigated sample and allows the identification of elements present in the sample.

Using the AZ software, it is also possible to concentrate only on a specific fibre section. This feature is ideal for comparing elemental compositions of various fibres areas - see Section A and Section B highlighted in EDX micrographs. These sections are further investigated in figures 3 and 4.
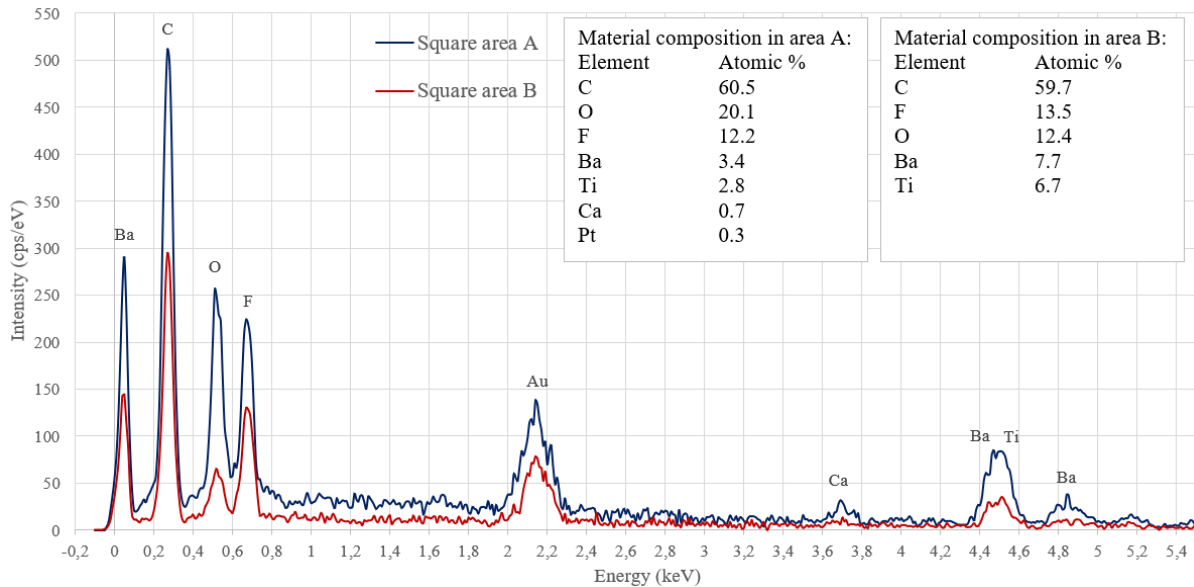
Figure 3: EDX spectrum of PVDF fibres doped by barium calcium zirconate titanate (BCZT)
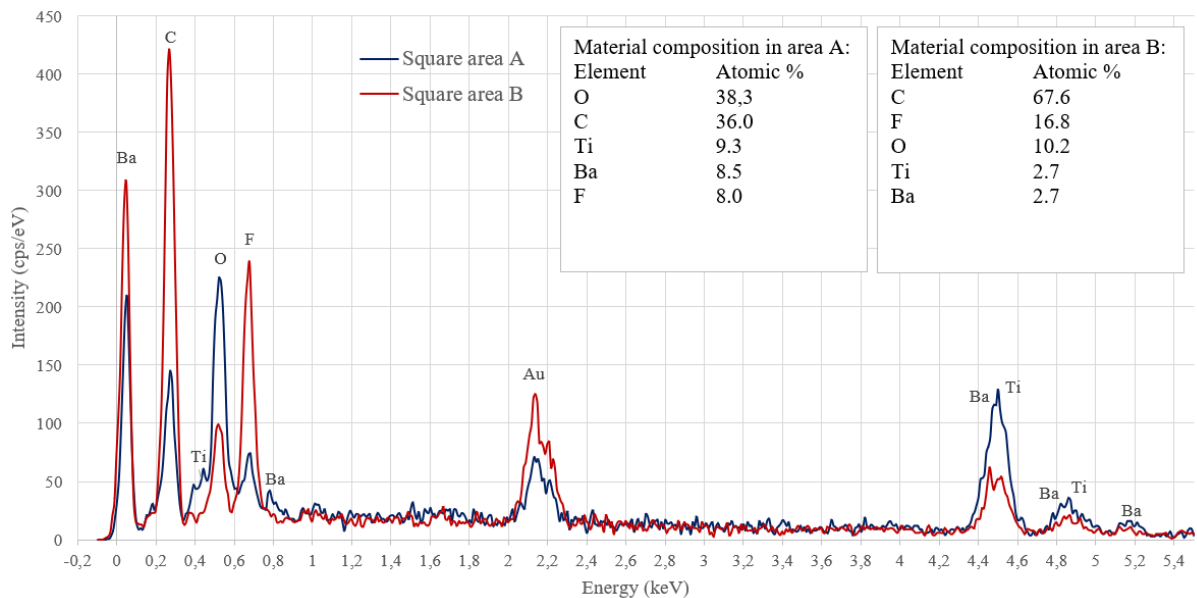
Figure 4: EDX spectrum of PVDF fibres doped by barium titanate (BT)

See also the atomic percentage of each section in the upper right corner of each graph. These reveal the atomic composition of each area, not influenced by the weight of each element. Only a slight difference in atomic percentages of elements in each section's composition supports the previous arguments. It can be concluded that fibres have been doped successfully with even distribution.

5 CONCLUSION

Principally, the modification of materials provides additional features and aptitudes while maintaining the fundamental characteristics of the materials. By improving polyvinylidene fluoride properties by the inclusion of piezoelectric ceramics, it is possible to induce β-phase in PVDF. Therefore, polymer-ceramic composites can be an excellent choice to achieve miniaturisation of energy storage devices by combining the merits of polymers and ceramics.

After fabricating the fibres from the polymer solution, the fibres' morphology was characterised by a scanning electron microscope. SEM offered a transparent picture of thin fibres enriched with ceramics particles; fibres exhibited a rough texture with a layered structure and, therefore, proposed that the ceramic particles were uniformly distributed throughout the fibres.

The presence of ceramic particles in the fibres was later confirmed using energy dispersive spectroscopy. EDX provided mapping of the elements of doped ceramics using a colourful micrograph. It is possible to distinguish building elements of the doped fibres in the pictures and their even dissipation in the fibres. By exporting data from AZtech, an EDX software, it was possible to create a series of spectra, revealing the element composition in percentage. Comparison of elemental compositions of various fibres areas is sufficient evidence of satisfactory dopant distribution. Therefore, it is possible to continue with development of doped PVDF fibres.

ACKNOWLEDGEMENTS

REFERENCES

[1]   Parjansri, P., Intatha, U., Pengpat, K. et al. Improvement in the electrical properties of BCZT Ceramics induced by self-seeds. Appl. Phys. A 125, 421 (2019). https://doi.org/10.1007/s00339-019-2711-9

[2]   Om Prakash Choudhary and Priyanka. Scanning Electron Microscope: Advantages and Disadvantages in Imaging Components. Int.J.Curr.Microbiol.App.Sci. 6(5): 1877-1882. doi: https://doi.org/10.20546/ijcmas.2017.605.207

[3]   Smejkalová, Tereza. Energeticky disperzní rentgenová spektroskopie dopovaných vláken PVDF [online]. Brno, 2020 [cit. 2021-03-03]. Available at: https://www.vutbr.cz/studenti/zav-prace/detail/131971.

SENSITIVITY ANALYSIS OF AN INDUCTION MACHINE WITH A SOLID ROTOR

Vladimír Bílek

Master Degree Programme (2), FEEC BUT

E-mail: xbilek22@stud.feec.vutbr.cz

Supervised by: Jan Bárta

E-mail: bartaj@feec.vutbr.cz

Abstract: The main focus of this paper is to show the results of sensitivity analysis of an induction machine with a solid rotor. For the case study, a three-phase induction machine with a squirrel cage is chosen. However, the rotor of this machine is replaced with an axially slitted solid rotor. The electromagnetic model is analyzed by transient simulation using the FEM-based program Ansys maxwell. For the sensitivity analysis, several parameters are changed in a wide range to see the difference in the machine overall performance. Some of the results of these simulations are shown and discussed at the end of this paper.

Keywords: Sensitivity analysis, Induction machine, Solid rotor, FEM, Simulation, Transient analysis

1 INTRODUCTION

In recent years, the demand for high-speed electrical machines has increased. This is due to the fact that some industrial applications require high rotational speed, as pumps, turbo-compressors, turbo-circulators, and others. For this reason, electrical machines with gearboxes are being replaced with electrical machines with a solid rotor. The main advantages of electrical machines with a solid rotor are withstanding high centrifugal forces, peripheral speed, and temperatures. They are mechanically rigid, highly reliable, compact, and easy to manufacture. Moreover, high-speed electrical machines have better electromagnetic performance and efficiency at higher speeds. This work deals with the methodology of analysis and sensitivity analysis of an induction machine with an axially slitted solid rotor, which is the most widespread solid rotor construction.

2 ANALYZED MACHINE

For the case study, an industrially produced three-phase, 4 pole induction machine with single-layer winding and squirrel cage, was chosen. The material of the stator and rotor core is M470-50A. Catalog parameters of the analyzed machine are listed in Table 1.

Parameter	Unit	Value
Rated output power	kW	1.5
Rated torque	Nm	9.905
Rated input voltage	V	3x400 (Y)
Rated input current	A	3.43
Frequency	Hz	50
Rated speed	rpm	1446.2
Efficiency	%	84.62

Table 1: Parameters of analyzed induction machine.

However, the rotor of this machine was replaced with an axially slitted solid rotor. The geometry of the rotor is shown in Figure 1. This rotor is made of a single piece of ferromagnetic material, where the material is stainless steel 1008. With the changed rotor, the overall performance of the machine will be worse, due to the lower rotational speed and the fact that a solid rotor tends to have a large slip and worse performance compared to the rotor with a squirrel cage.
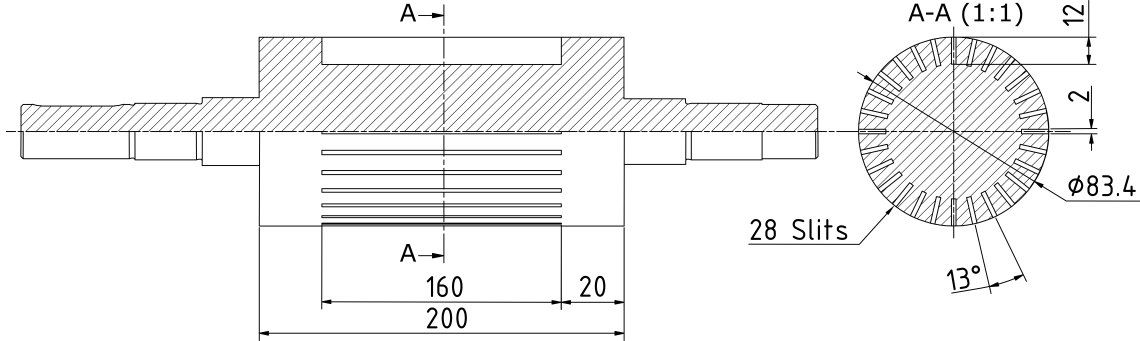
Figure 1: Geometry of axially slitted solid rotor.

3 THE METHODOLOGY OF ANALYSIS OF ELECTRICAL MACHINES WITH A SOLID ROTOR

Due to the complex physical nature of electrical machines with a solid rotor, analytical methods for the calculation of machine's performance is not the best choice. Most of these analytical methods include some simplification, which impairs the accuracy of the calculated results. This is due to a very complex distribution of relative permeability in the rotor which depends on the material and its B-H curve. Also, the distribution of rotor losses is very complex and depends on the penetration depth of the higher air-gap frequencies. Both of these variables are most pronounced in the axially slitted solid rotor. A much better choice would be 3D FEM-based transient analysis, due to the nature of the rotor construction. However, this type of analysis takes a lot of computation power, memory, and is highly time-consuming. So, the last and best option is 2D FEM-based transient analysis, which takes less computation power, memory and is much less time-consuming. However, the 2D electromagnetic model is not able to take into account the 3D aspects of the rotor construction, such as resistivity and leakage inductance of the rotor end regions. This would result in wrong calculated results, where for example the value of the torque would be too optimistic. In order to take into account the end regions of the rotor, corrective end-effect factors were proposed, as it is stated in [1]. The corrective end-effect factors are adjusting the conductivity of the rotor in the electromagnetic model in a way that includes the resistivity of the end regions of the rotor, with following simple equation:

$$\sigma_{Corr} = \sigma_{Fe} \cdot k, \tag{1}$$

where σ_{Fe} is the conductivity of rotor material and k is the corrective end-effect factor. There are two types of corrective end-effect factors that depend on the geometry of the rotor and the slip frequency of the rotor. The chosen formula of corrective end-effect factor based on geometry has according to [2] shape:

$$k_{Russell,M} = 1 - \frac{\tau_p}{\pi l_s} \frac{tanh\left(\frac{\pi l_s}{\tau_p}\right)}{\left(1 + tanh\left(\frac{\pi l_s}{\tau_p}\right) tanh\left(\frac{\pi l_{end}}{\tau_p}\right)\right)}, \tag{2}$$

where l_s is active length of the machine, τ_p is pole pitch of the rotor and l_{end} is length of the rotor end regions. The chosen formula of corrective end-effect factor based on slip frequency has according to [3] shape:

$$k_{PAN} = 1 - c \cdot \omega_r^{\frac{3}{4}}, \tag{3}$$

where c is an experimental adaptation coefficient (estimated as $c = 0.022$) and ω_r is angular slip frequency of the rotor. The product of these two corrective end-effect factors gives a total correction factor which is then used in equation (1).

4   PROCESS OF SENSITIVITY ANALYSIS

The sensitivity analysis was done by changing individually the value, in a wide range with reasonable limits, of the following 8 parameters: active length of the machine, length of rotor end region, number of stator winding conductors, stator tooth width, stator slot height, length of the air-gap, depth of the axial slits, and width of the axial slits. Each of the new geometry is made with the FreeCAD program, which is an open-source CAD program. The simulations of the sensitivity analysis are done with the 2D FEM program Ansys maxwell. The type of simulation which was used is transient with constant speed, where the constant speed was selected with the respect to the highest efficiency of the machine with the solid rotor. The speed was estimated to be $1322\,rpm$, which means that the slip of the machine increased significantly. The whole process of the simulation was automated using the programing language Python. The algorithm of the sensitivity analysis is shown in Figure 2.
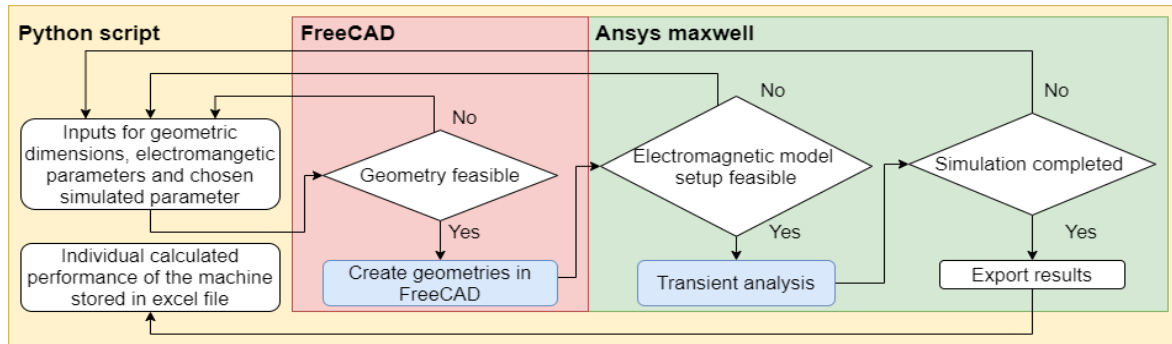
Figure 2:   Algorithm of sensitivity analysis of an electrical machine with a solid rotor.

5   SENSITIVITY ANALYSIS RESULTS

Within the sensitivity analysis, the focus was put mainly on four output parameters of all simulations. These parameters include electromagnetic torque, electromagnetic efficiency, output power, and phase current. All of the results of these four parameters are shown in Figure 3. Here, on the x-axis of the graphs, is the normalized value of a variable, which represents the relative growth of changed parameters. This way, it is possible to put all results together. It is clear from the results, that the overall performance of the machine with the solid rotor is worse, compared to the original one with squirrel cage. This output was expected. It can also be seen that in order to raise effectively torque, output power, and efficiency, it is necessary to increase the active length of the machine. However, this will also increase the size of the machine and phase current, which results in a higher current density of the stator slot. Another improvement could be made by narrowing the axial slits in the rotor. This modification would raise torque, output power and efficiency. It would also decrease phase current and current density of the stator slot. But this modification depends on the manufactures capabilities. By changing slit depth to 150 % it would be possible to achieve higher torque, output power, and efficiency. But this would also increase phase current. After this point, the performance of the machine would be worse. By lowering the number of conductors, the torque and output power would raise. But the efficiency would be lower and phase current would raise. The length of the air-gap is in this machine perfectly design. On the other hand length of the end region does not have much effect on the performance of the machine. And by changing either length or width of the stator slot, the performance of the machine is minimally improved or gets worse.
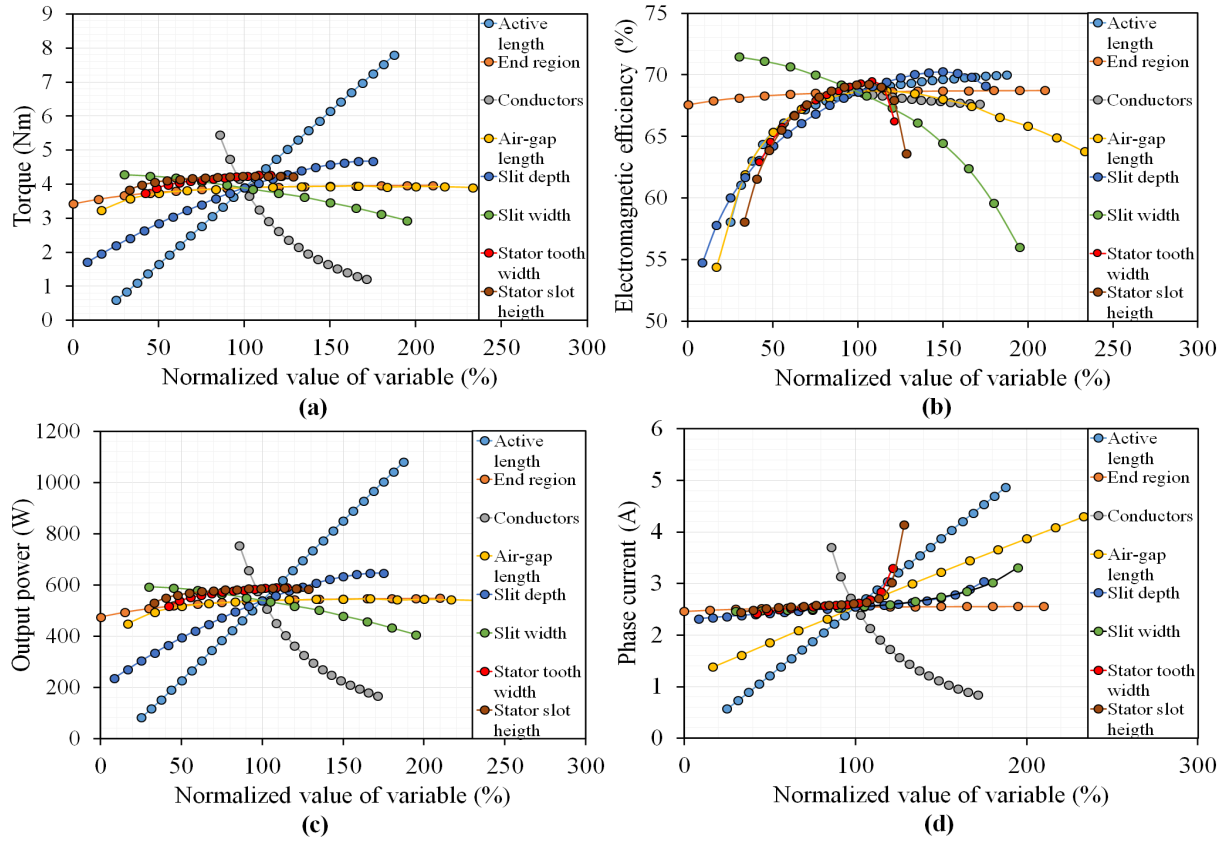
Figure 3: Function of: torque (a), electromagnetic efficiency (b), output power (c) and phase current (d) on normalized value of variable.

6 CONCLUSION

This paper showed the methodology and process of sensitivity analysis of the induction machine with the axially slitted solid rotor, using the FEM-based program Ansys maxwell. From the results, it can be seen that the machine has worse performance compared to the original machine with squirrel cage. But otherwise, it is a very well-designed machine with some room for improvement. This improvement could be done by adjusting some of the machine parameters described in this paper.

ACKNOWLEDGEMENT

REFERENCES

[1] T. Aho, "Electromagnetic analysis of a solid steel rotor motor for demanding operation environments," PhD thesis, Lappeenranta University of Technology, 2007.

[2] J. Pyrhonen, J. Nerg, P. Kurronen and U. Lauber, "High-Speed High-Output Solid-Rotor Induction-Motor Technology for Gas Compression," in IEEE Transactions on Industrial Electronics, IEEE, vol. 57, no. 1, pp. 272-280, Jan. 2010, doi: 10.1109/TIE.2009.2021595.

[3] J. Pyrhonen, J. K. Nerg, T. Aho and P. T. Kurronen, "Solid rotor end effects - Analytic and experimental results for high-power high-speed machines," IEEE EUROCON 2009, IEEE, St. Petersburg, Russia, 2009, pp. 688-695, doi: 10.1109/EURCON.2009.5167708.

THERMAL ANALYSIS PROCESSING OF STATOR PART OF HIGH-POWER ELECTRICAL MACHINE IN ANSYS WORKBENCH

Martin Světlík

Master's Power Electrical and Electronic Engineering (1), FEEC BUT

E-mail: xsvetl06@stud.feec.vutbr.cz

Supervised by: Marek Toman

E-mail: Marek.Toman@vutbr.cz

Abstract: This paper deals with thermal calculation of a high-power electrical machine by final elements method using. The first part of the paper is focused on preprocessing. Therefore geometry preparation by CAD software Inventor, meshing and setting of boundary conditions by Ansys. The second part of the paper deals with thermal analysis and results are discussed in conclus.

Keywords: 3D thermal analysis, finite element method, Steady-state

1 INTRODUCTION

Thermal analysis is an undivided part of the construction of almost all technical mechanisms. This is why is necessary to be very considerate of the thermal analysis in the case of electrical rotating machines. This analysis is one of the three main parts needed to design and construct rotating machines. With the development of computer technology, it is possible to use the finite element method, which is suitable for accurate solutions of temperature distribution in electrical machines. This software computing allows the development of optimized geometries to increase the efficiency of these machines and verification right functionality without the production of test specimens.

2 MODEL PREPARING

2.1 GEOMETRY PREPARING

Geometry creation is one of the key components of model preparation in so-called preprocessing. It is because the time and computational complexity of the model analysis depend on the geometry preparation. Due to time and computational complexity might be used some simplifying of the geometry with minimal effect on analysis accuracy but it makes the analysis shorter and less computational complex. These simplifications also simplify the setting complexity of mesh which is necessary for computing and also minimizing preprocessing time. There is shown an example of geometry made by Inventor software in the figure 1. Geometry made in this software can be export to Ansys software where is Steady-state thermal package used. The advantage of using Inventor is the possibility of parameterization of the model and its subsequent parametric transfer to the already mentioned Ansys. [1]
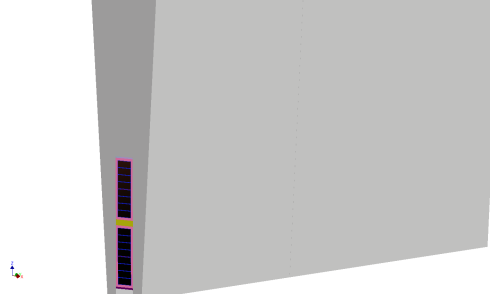
Figure 1: Stator model designed in Inventor

2.2 CREATION OF FINITE ELEMENT NETWORK

Creation of newtork also known as meshing very closely related with prepared geometry. The simplifications used to design this geometry will now be of interest. To right aproximization of the problem it is very important to set generated mesh fine enough but in the case of too fine mesh the complexity of analysis increase without obtaining more accurate results. In the case of "Steady State thermal", the optimal shape of the network segments is set to the shape of a hexagon. This shape do not have to be regular but this regularity is better for quickly analysis. Exmple of designet mesh is shown in the figure 2.
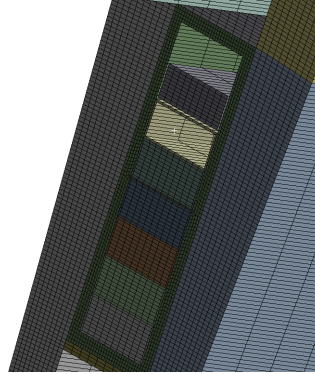
Figure 2: Mesh designet in "Steady state thermal" software

2.3 CONTACT AREA SETTINGS

As such, the Ansys software is not able to precisely set the contact surfaces of the individual bodies as they actually touched, so it is necessary to verify and possibly correct the incorrect setting of these surfaces. Without the correct setting, the model could not be solvable, as heat would not be removed from any of its parts and this would lead to an increase in its temperature to an infinitely large value. When setting the contact surfaces, it is also possible to define the thermal resistance of heat transfer between individual surfaces, which can be calculated according to [2] as

$$R = \frac{\Delta T}{qA},\tag{1}$$

where R is heat transfer resistance, ΔT is temperature drop, q is heat flux value and A is apparent constact surface. The method of setting the contact surfaces is shown in the figure 3
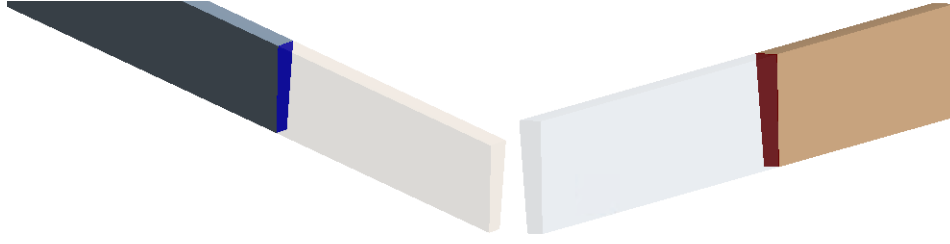
Figure 3: Setting of contact surfaces example

2.4 BOUNDARY CONDITIONS SETTING

The setting of boundary conditions is used to find a solution corresponding to the specific conditions that are expected to affect the machine. According to [3], these boundary conditions include the specified temperature, heat flux density, convection and radiation. Depending on the expected location of the model in real use, some of these conditions can be neglected, but incorrect neglect or setting can lead to erroneous results of solving this model. Therefore, it is necessary to pay attention to their correct setting so that the model corresponds to reality as much as possible and the results have a meaningful value. [4], [5] As part of setting boundary conditions, there is a table 1, in which these conditions are summarized.

Total volume losses in yoke	[kW/cm^3]	14,48
Total volume losses in winding	[kW/cm^3]	3,87
Convection heat transfer from the front of the yoke	[W/cm^2]	900
Convection heat transfer into the air gap	[W/cm^2]	4000
Heat transfer via heat sink	[W/cm^2]	11400

Table 1: Table of set boundary conditions during calculation preparation

After setting all the previous points, it is possible to continue to the process of solving the model. This process runs automatically in Ansys software, which uses a general heat transfer relationship in the form [3] for static analysis.

$$\frac{\partial^2 T}{\partial x^2} + \frac{\partial^2 T}{\partial y^2} + \frac{\partial^2 T}{\partial z^2} + \frac{\dot{e}_{\text{gen}}}{k} = 0. \tag{2}$$

By solving this model, the resulting temperatures were obtained, which are indicated in figure 4. It can be seen in this figure that the warmest parts of the stator are the windings. This is due to its poor cooling, which is related to double insulation. The term double insulation is used here to mean the insulation of the conductors themselves and subsequently the insulation of each layer of conductors, which is clearly seen in figure 2. From the figure 4 it is further evident that the lowest temperature was on the outside of the stator, especially in the part where water cooling was considered. This water cooling allowed better heat dissipation from the entire stator.
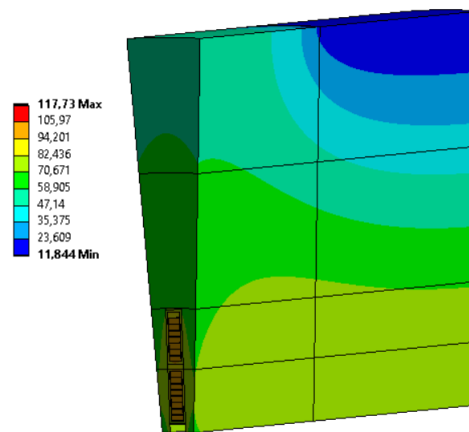
Figure 4: Resulting temperatures

3 CONCLUSION

The article was devoted to the preparation and evaluation of the stator model using Ansys software and it defined the basic relationships through which this calculation can be performed. An important part of this model was the evaluation and justification of the temperatures shown in figure 4, which showed the resulting temperatures. From these temperatures it is clear that the warmest part of the model was the wires placed in the groove, which was the expected result.

REFERENCES

[1] *AUTODESK INVENTOR 2021: Welcome to Inventor Help* [online]. 2021 [cit. 2021-03-14]. Dostupné z: http://help.autodesk.com/view/INVNTOR/2021/ENU/

[2] MANTELLI*, M. B. H. a M. M. YOVANOVICH. *Thermal Contact Resistance*. Federal University of Santa Catarina, Florianopolis, Brazil, University of Waterloo, Waterloo, Canada., , 40.

[3] CENGEL, Yunus A. a Afshin J. GHAJAR. *HEAT AND MASS TRANSFER: FUNDAMENTALS & APPLICATIONS*. 5th edition. New York: McGraw-Hill Education, 2015. ISBN 978-0-07-339818-1.

[4] *ANSYS Steady-State Thermal: Initial Temperature* [online]. 2020 [cit. 2021-03-14]. Dostupné z: https://forum.ansys.com/discussion/13915/ansys-steady-state-thermal-initial-temperature

[5] BARRETO a JORGE. *Applying Boundary conditions for Thermal analysis on a plate*. 2016.

IMPACT OF UNCONTROLLED CHARGING OF ELECTRIC VEHICLES ON PHOTOVOLTAIC HOSTING CAPACITY OF FINNISH LV DISTRIBUTION NETWORKS

Robin Filip

Master Degree Programme (2), FEEC BUT

E-mail: xfilip42@stud.feec.vutbr.cz

Supervised by: Martin Paar

E-mail: paar@feec.vutbr.cz

Abstract: This paper deals with impact of uncontrolled charging of electric vehicles on photovoltaic hosting capacity of low-voltage distribution networks. Monte Carlo based simulation is used to generate randomized profiles of electric vehicle uncontrolled charging and to compare photovoltaic hosting capacity of Finnish suburban residential areas with different penetration of electric vehicles.

Keywords: Electric vehicles, photovoltaic hosting capacity, Monte Carlo, distribution network

1   INTRODUCTION

World-wide pressure on sustainable development of our society urges us to pursue more environmentally friendly technologies. Transportation, responsible for more than a quarter of all greenhouse gasses production in USA [1], is under immense pressure to change. Car manufacturers are forced by political pressure to promptly reduce production of cars with internal combustion engine (ICE) and seek more environmental-friendly solutions.

Now, the most prospering technology, able to substitute ICE vehicles, are electric vehicles (EVs). Their large-scale utilization is conditioned by a significant increase of transmission capabilities of our electrical networks which, if not handled properly, can endanger strict requirements for electricity supply.

The goal of this paper is to study EVs interconnection with another prosperous technology – photovoltaic (PV) systems – and to investigate impact of uncontrolled EV charging on grids ability to incorporate PV.

2   METHODOLOGY

2.1   LV DISTRIBUTION NETWORKS

Centerpiece of this study are low-voltage (LV) distribution networks (DNs) in suburban areas. Approach, based on method used in [2] , utilizes transformer of nominal power 315 kVA that feeds 3 feeders with 40 customers in 10 nodes. Conductors between the nodes are sized to meet customers demand based on their household consumption and type of heating.

2.2   UNCONTROLLED EV CHARGING

To study effect of EV charging, parametrization of EV owners and their behavior have to be done. Monte Carlo (MC) analysis is based on stochasticity and thus key-parameters – time of arrival and daily mileage – are described by probability density functions shown in Fig. 1 a) [3]. To capture the worst scenario, 100 % EV penetration (each household has one EV) is assumed. Structure of EV chargers depicts Fig. 1 b).
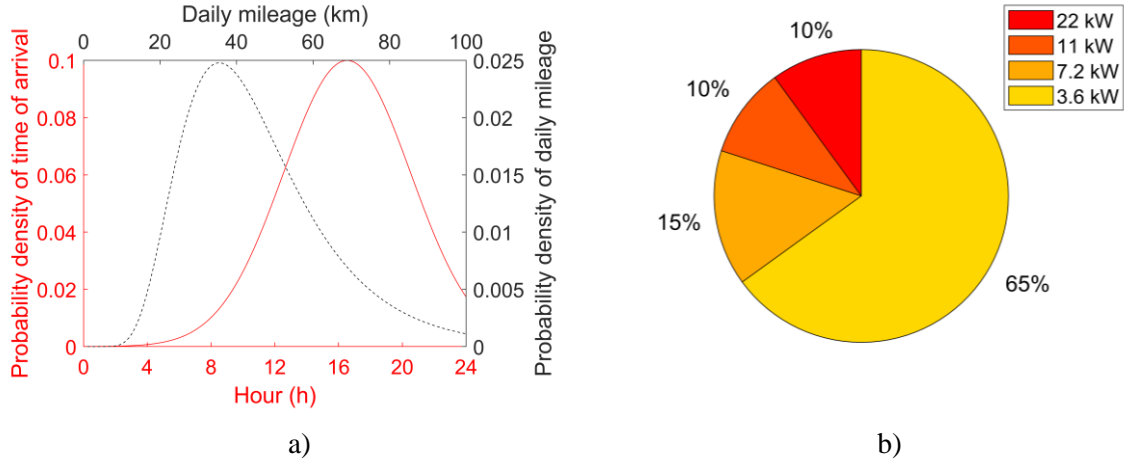
Fig. 1: Probability density functions (a), structure of EV chargers (b)

Based on the time of arrival (connection EV to the charger), daily mileage, energy consumption per km (changing during the year), charger and its efficiency, 600 random annual charging profiles for 1 EV were generated. To ensure an acceptable calculation time, charging profiles are in form of average hourly data. Example of annual charging profile of all customers in a LV DN (40 customers/EVs) can be seen in Fig. 2.

2.3 MONTE-CARLO BASED ALGORITHM

Uncontrolled EV charging is characteristic by high stochasticity and is therefore ideal for MC approach that randomly parametrize the simulation, obtains results and starts next iteration. In this study, 1000 iterations per case are used. In each iteration, load and EV charging profiles differ.

PV hosting capacity (HC) is searched in each iteration by gradually increasing PV rated power and verifying conformity with the limits of risk parameters like voltage increase/drop, conductor's ampacity or transformer load. When any of these limits is reached, simulation stops, and the PV installed power is titled as HC for that iteration. In order to ensure comparability between hosting capacities in different areas, hosting capacity can be defined as PV rated power (downstream the transformer) over the rated power of the transformer.
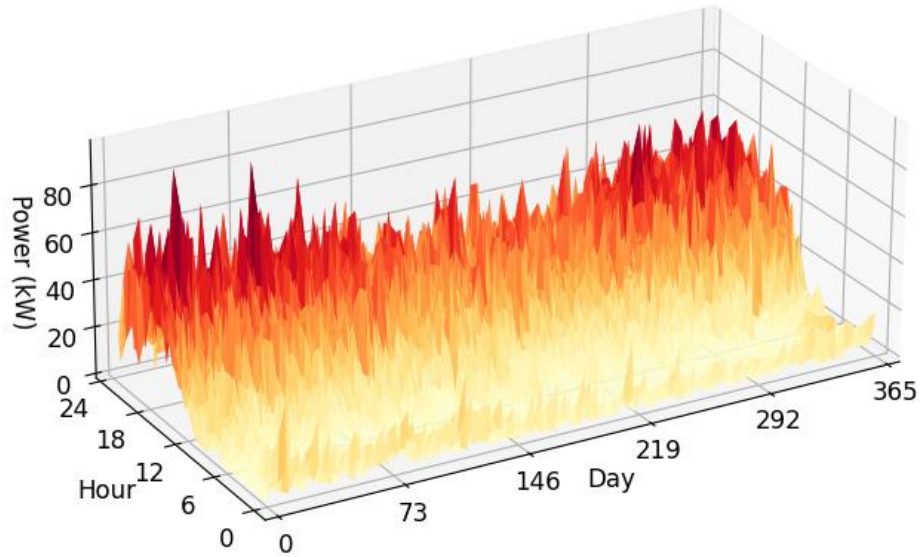
Fig. 2: Example of uncontrolled EV charging in LV DN (summation over all nodes)
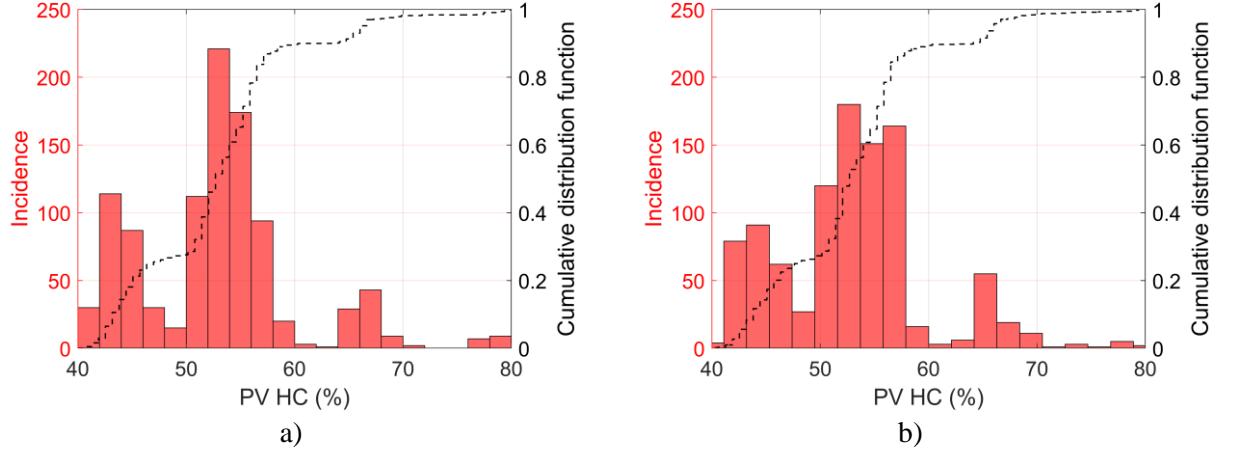
Fig. 3: PV HC distribution with 0% (a) and 100% (b) EV penetration

2.4 RESULTS

Using a MC based simulation, PV HC has been analyzed for both 0 % and 100 % EV penetration. Comparison of the results in Fig. 3 shows only insignificant differences. In both cases, HC ranged from 40 % to 80 % with average 52 %. Given the transformer of nominal power 315 kVA, average analyzed DN is able to absorb PV of rated power 163.8 kWp. With the 40 customers in the area, each of them can install PV up to 4.1 kWp without endangering the electricity supply.

These results are caused by mismatch between EV charging and PV power profiles. This study showed that EV charging and its stochastic behavior consistent with [3] (arrival time peaking around 5 PM) does not have significant impact on HC of PV peaking around the noon.

Another interesting result from the simulations is relation between total load (households consumptions + EV charging) with PV production in the worst case (WC) – the hour in which the limits were exceeded. This is shown in Fig. 4. In some cases, electricity supply is not endangered even when PV production is 10 times higher than the load (left side of Fig. 4). On the other hand, there is certain probability that the limits can be endangered even when the PV production is only 2 times higher than consumption (right side of Fig. 4).
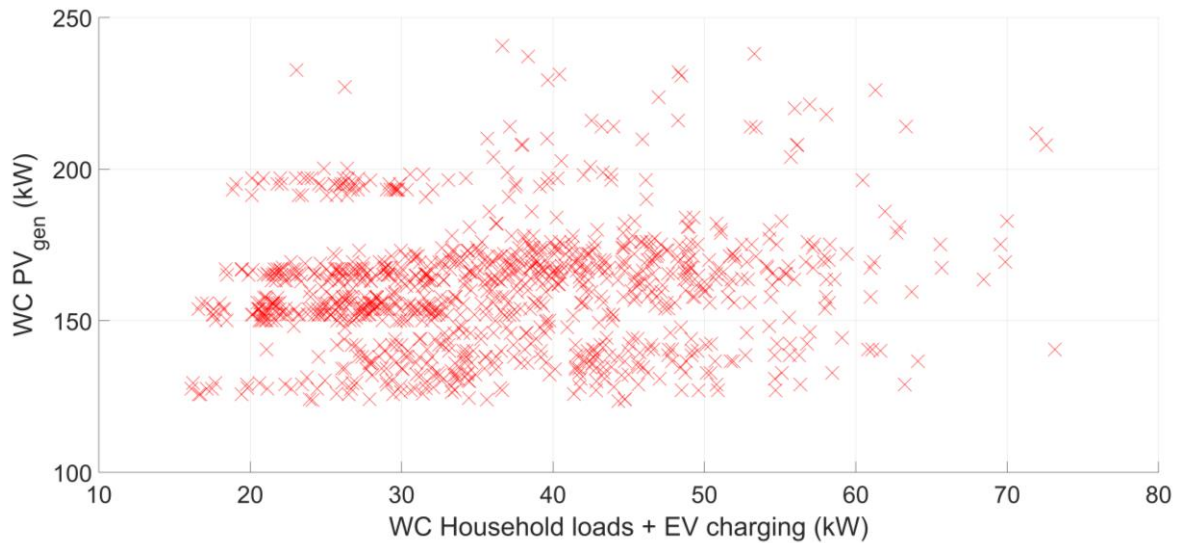
Fig. 4: Comparison of PV power with loads in distribution network in the worst case (WC)

3 DISCUSSION

It was seen that uncontrolled EV charging does not have significant impact on PV hosting capacity. However, it must be noted that the results are valid only for the locations assumed in this paper. Behavior of the customers corresponds with the current behavior of Finnish drivers and is valid only for residential areas. In nonresidential areas, behavior and arrival time could be closer to the PV peak and thus increase local consumption and increase PV HC.

This study also omits the difference between customers behavior on workdays and weekends. It can be expected that the behavior and arrival time differ, and its impact should be studied further.

Finding that there is no significant influence of uncontrolled charging on PV HC also opens the door to questions about more sophisticated forms of EV charging and their influence, especially in cooperation with battery energy storage systems (BESS), should be investigated.

4 CONCLUSIONS

A Monte Carlo based simulation was used to analyze impact of uncontrolled EV charging on photovoltaic hosting capacity of Finnish suburban residential areas. PV HC in two different EV penetration cases – 0 % and 100 % - was analyzed in low-voltage distribution networks with 3 feeders and 40 customers equally distributed among 10 nodes. In these grids, PV installed power was increased step by step and concordance with electricity parameters like voltage rise/drop, cable ampacity etc. was analyzed.

The results showed no significant impact of uncontrolled EV charging on PV HC as PV peak is several hours before connecting EV to the chargers in residential areas. Average HC corresponds to 52.8 %, which is equal to PV of nominal power 163.8 kWp. Given the 40 customers in the area, each of them can install PV up to 4.1 kWp without endangering the electricity supply.

REFERENCES

[1] Sources of Greenhouse Gas Emissions: Transportation Sector Emissions [online]. United States Environmental Protection Agency. Available at: https://www.epa.gov/ghgemissions/sources-greenhouse-gas-emissions#transportation

[2] LINDNER, Martin and Matti LEHTONEN. 2017. An Analysis of Photo-Voltaic Hosting Capacity in Finnish Low Voltage Distribution Networks. Energies [online]. Basel: MDPI, 10(11), 1702. Available at: http://search.proquest.com/docview/1977967915/

[3] O. Lehtinen, S. Pitkäniemi, A. Weckman, M. Aikio, M. Mabano and M. Lehtonen. Electric Vehicle Charging Loads in Residential Areas of Apartment Houses 2020 21st International Scientific Conference on Electric Power Engineering (EPE), Prague, Czech Republic, 2020, pp. 1-6, doi: 10.1109/EPE51172.2020.9269191

USING MACHINE LEARNING TECHNIQUES IN THE VISUAL DETECTION OF STARLINGS IN VINEYARDS

Josef Novotný

Master Programme (2), FEEC BUT

E-mail: xnovot0v@vutbr.cz

Petr Marcoň

E-mail: marcon@feec.vutbr.cz

Abstract: This paper deals with the visual detection of starlings. The aim is to design an early warning detection system that protects crops from flocks of starlings. This system uses computer vision and machine learning algorithms. In the first phase, the activity in the vineyard was collected. Further, the neural network model using a cloud-based AutoML platform was trained and evaluated. The final classifier distinguishes objects into several categories. These categories include individual birds, flocks, and various unintended objects such as flies and bees. Overall, the flock detection algorithm achieved 89 % accuracy and 94 % recall.

Keywords: Starling detection, Flock detection, Computer vision, Machine Learning

1   INTRODUCTION

Vineyards belong to the main area endangered by voracious swarms of starlings. They present the most significant threat of every winemaker or fruit grower. Starlings refer to one of the most invasive species. They leave behind grape crop damaged, which reflects economic losses for winemakers. The use of various scaring systems can prevent the consequences of starling invasion. However, the vast majority of bird deterrent devices do not consider the presence of starlings. It significantly reduces the efficiency of such scaring devices. This paper aims to design a detection algorithm based on computer vision methods to supplement the scaring techniques. According to the current state of the art [1], several studies proposed bird detection based on neural networks (R-CNN, R-FCN, YOLO). The other authors [2] focus on designing a system based on the use of Gaussian and Gabor filters and HOG in combination with convolutional neural networks. This paper proposes to verify another method, namely a neural network architecture search (NAS) using the cloud-based tool Google AutoML to detect individual birds and starling flocks.

2   DATA COLLECTION

First, the activity on the vineyard was collected using a camera and microcomputer Nvidia Jetson Nano by employing an algorithm based on the differential method supplemented by filtering the background image, as shown in fig. 1. This method works as a motion detector with background subtraction. It eliminates the erroneous detection caused by the motion of vine leaves and surroundings.
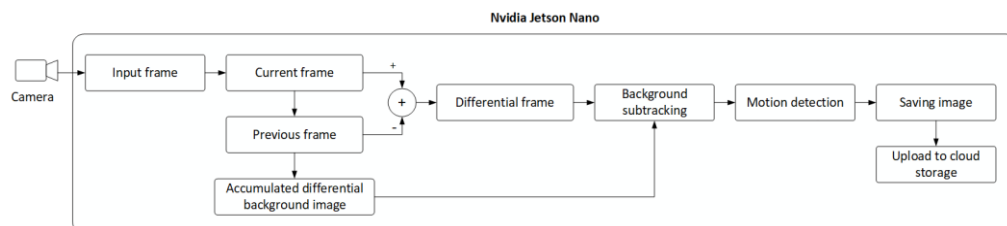
Figure 1:   Flow chart of the motion detector for data collection.

3 SET OF COLLECTED IMAGES

The collected images were further divided into three main categories: bugs, birds, and bird flocks. The dataset also includes a helicopter captured entirely randomly. Due to different visual characteristics, birds and bugs were divided further into subcategories, as shown in fig. 2. The set of images contains 993 images, which were annotated for the purpose of supervised machine learning. The annotation of objects was conducted by marking areas of their occurrence with the classification name, detailed described in the tab. 1.
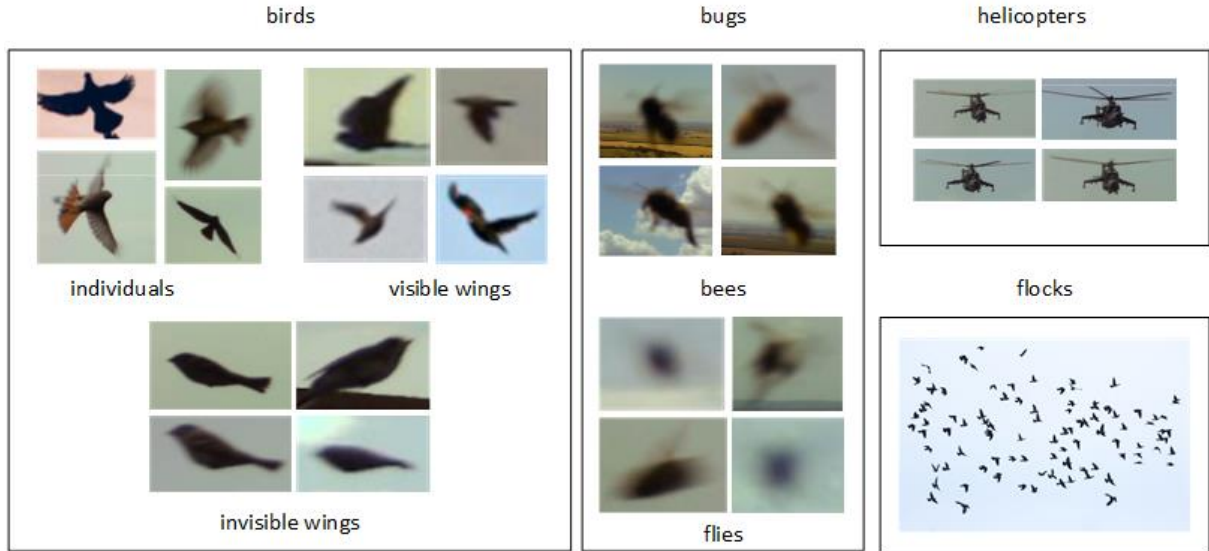
Figure 2:   Categorizing the collected images.

The data collected did not contain enough relevant images of bird flocks, so they were supplemented from open datasets. Most of the images contained more objects, so the total number of objects exceeds the number of images collected. The images were divided according to the ratio 8: 1: 1 into a training, validation, and test set within machine learning.

Category		Training set	Validation set	Test set
Bugs	flies	442	41	43
	bees	36	7	4
Birds	individuals	2308	445	228
	visible wings	55	6	6
	invisible wings	160	5	3
Flocks		59	9	6
Helicopters		11	2	4

Table 1:   Overview of categories within the training, validation and testing set.

4 MACHINE LEARNING ALGORITHM

The application of cloud-based Google AutoML Vision provided a tool for training a machine learning model. This approach presents Neural Architecture Search (NAS), which searches for an optimal neural architecture to a given problem using a recurrent neural network, as shown in fig. 3. The algorithm designs a basic set of hyperparameters, the number of layers and nodes of the convo-

lutional neural networks. In subsequent iterations, the feedback specifies the individual parameters. This process is performed until the algorithm gradually finds the most optimal neural architecture for a particular dataset. Then, the resulting model is selected from a set of convolutional neural networks based on the optimal properties such as sufficient accuracy and recall of detection. [3] [4]
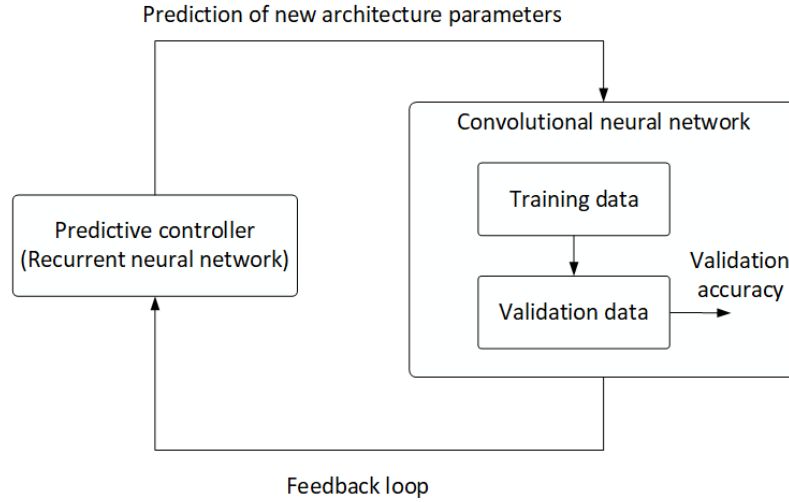
Figure 3:   Block scheme of the neural architecture search algorithm Google AutoML. [4]

5   CLASSIFIER RESULTS

The classification model was deployed to the Tensorflow Lite framework for further implementation and evaluation. For evaluation purposes, a separate dataset was used. The classifier result is based on a comparison of classified and annotated ground truth. Objects in images are marked as true-positive if the detected area with the annotated overlap by at least 50 %. The results of the classifier are further presented in tab. 2.

	Birds	Bugs	Flocks
Precision	75.6 %	98.0 %	88.9 %
Recall	76.0 %	67.6 %	94.1 %
F1 score	75.8 %	80.0 %	91.4 %

Table 2:   Precision, recall and F1 score of the classifier from image test set.

Precision represents the proportion of true-positive detections from the whole set of detected objects. In the case of birds, it is 75.6 %, which is the lowest value of accuracy from the set of all objects. It was mainly due to the occurrence of a larger number of false-positive images caused by a variability of the object size. For bugs, the accuracy reaches 98 %. The bugs were well recognizable since they occurred mainly close to the camera. The flocks have an accuracy of 88.9 %. Bird flocks were generally defined as a grouping of five or more birds. Another parameter was the recall of birds and bugs, which shows 76 %, resp. 67.9 %. It was caused by a low capture of relatively small and indistinct objects at greater distances. The sensitivity of the bird flock detection reaches 94.1 %, which is due to the relatively well-defined structure of the flock object, as shown in fig. 4. The F1 score evaluates the harmonic average of accuracy and recall. Bird flocks show 91.4 %.

From the results, it can be concluded that the biggest obstacle to the correct classification is too small size of the objects given by the distance from the camera. Conversely, large groups of birds can be detected with a high success rate, presented by accuracy and recall.

Figure 4: Examples of detected objects and corresponding confidence score.

6 CONCLUSION

The aim of this paper was to secure visual detection of starling flocks. First, a data collection of activity in the vineyard was performed using differential motion detection algorithms. However, this method was found unsuitable for detecting starling flocks due to a large amount of erroneous detection caused mainly by flying insect. The collected data were further used as a training set for the purpose of a supervised machine learning algorithm, which presents a cloud-based tool Google AutoML Vision. This method searches optimal neural network architecture and tunes the hyperparameters automatically. In the next step, the classifier was exported to the Tensorflow Lite framework. Subsequently, the results were evaluated on a separate test set of images. The accuracy of detection of bird flocks reaches 88.9 %, with a recall of 94.1 %. The biggest obstacle for the classification algorithm was mainly the distance of detected objects. However, it has been confirmed that the detection method is sufficiently effective and can be used as part of starling scaring device systems.

REFERENCES

[1] Hong, S.-J.; Han, Y.; Kim, S.-Y.; Lee, A.-Y.; Kim, G. *Application of Deep-Learning Methods to Bird Detection Using Unmanned Aerial Vehicle Imagery*. Sensors 2019, 19, 1651, doi:10.3390/s19071651.

[2] Ghosh, S.K.; Islam, Md.R. *Bird Species Detection and Classification Based on HOG Feature Using Convolutional Neural Network. In Proceedings of the Recent Trends in Image Processing and Pattern Recognition*; Santosh, K.C., Hegadi, R.S., Eds.; Springer: Singapore, 2019; pp. 363–373.

[3] Zoph, Barret; Le, Quoc V.: *Neural Architecture Search with Reinforcement Learning*, 2016.

[4] *Cloud AutoML Documentation* [online]. Available at: https://cloud.google.com/automl/docs

UTILIZATION OF CONVOLUTIONAL NEURAL NETWORKS FOR SEGMENTATION OF MOUSE EMBRYOS CARTILAGINOUS TISSUE IN MICRO-CT DATA

Veronika Poláková

Master Degree Programme (2), FEEC BUT

E-mail: xpolak35@vutbr.cz

Supervised by: Jiří Chmelík, Jan Matula

E-mail: chmelikj@feec.vutbr.cz, Jan.Matula2@ceitec.vutbr.cz

Abstract: Automatic segmentation of the biological structures in micro-CT data is still a challenge since the object of interest (craniofacial cartilage in our case) is commonly not characterized by unique voxel intensity or sharp borders. In recent years, convolutional neural networks (CNNs) have become exceedingly popular in many areas of computer vision. Specifically, for biomedical image segmentation problems, U-Net architecture is widely used. However, in case of micro-CT data, there is a question whether 3D CNN would not be more beneficial. This paper introduces CNN architecture based on V-Net as well as the methodology for data preprocessing and postprocessing. The baseline architecture was further optimized using advanced techniques such as Atrous Spatial Pyramid Pooling (ASPP) module, Scaled Exponential Linear Unit (SELU) activation function, multi-output supervision and Dense blocks. For network learning, modern approaches were used including learning rate warmup or AdamW optimizer. Even though the 3D CNN do not outperform U-Net regarding the craniofacial cartilage segmentation, the optimization raises the median of Dice coefficient from 69.74 % to 80.01 %. Therefore, utilizing these advanced techniques is highly encouraged as they can be easily added to any U-Net-like architecture and may remarkably improve the results.

Keywords: CNN, micro-CT, image segmentation, cartilaginous tissue, V-Net

1 INTRODUCTION

Developmental biologists use mouse embryos to investigate the complex process of morphogenesis but also abnormal growth of the mammalian face (Figure 1). For this task, various imaging techniques are used including X-ray micro computed tomography (micro-CT). However, since a manual segmentation of the craniofacial cartilage from 3D image data is time-consuming and operator dependent, there is a demand for an automatic approach. Several image processing and machine learning attitudes were already examined for solving this issue [6]. Nevertheless, they were entirely surpassed by a deep learning approach (U-Net) [6] encouraging for a greater research in the area of convolutional neural networks (CNNs).

Figure 1: Craniofacial cartilage in the context of mouse embryo head.

A question that arises considering volumetric data is whether using a 3D approach could improve the segmentation results. Regrettably, the craniofacial cartilage is oblong in all directions meaning that even with a reasonable rate of downsampling the neural network cannot see the cartilage (or at least its sufficient part) at once because of memory limits. Nevertheless, the hypothesis is that even 3D *piecewise* segmentation can enhance the segmentation results since the neural network can benefit from surrounding slices. Although 2D CNNs can accept 3D inputs (set of slices), they cannot provide spatial information about the extracted features in contrast to 3D CNNs. Therefore, 3D CNN solution

based on V-Net [7] was proposed as a baseline network which was further refined in a subsequent optimization process. The paper emphasizes an utilization of advanced architectural modifications as they can greatly improve the segmentation performance compared to the vanilla architecture.

2   AVAILABLE DATA, PREPROCESSING AND POSTPROCESSING

For experiments, 14 micro-CT datasets of mouse embryos in 17.5-th day of development were available. Three of the mice were mutants, the rest of them were wild types. A voxel was always isometric, however, voxel size depends on the given dataset ($4.5\,\mu m$ to $6.2\,\mu m$). The median size of the micro-CT data was $1348 \times 1916 \times 1419$. For each dataset, a binary mask defining the craniofacial cartilage was available (see Figure 1 and Figure 2a for a better understanding of the segmentation problem).

Before the network's offline learning, several preprocessing steps were performed. Firstly, to provide enough global information to the network, data were four times downsampled in the xy plane. After the downsampling, slices for each dataset were center cropped or padded to 320×448 size in order to unify the size of slices for different datasets [6]. Next, a window-level transformation was applied on the whole 3D image data to enhance the contrast between tissues (Figure 2). Finally, the 3D image data were normalized using z-score.

(a)   (b)   (c)
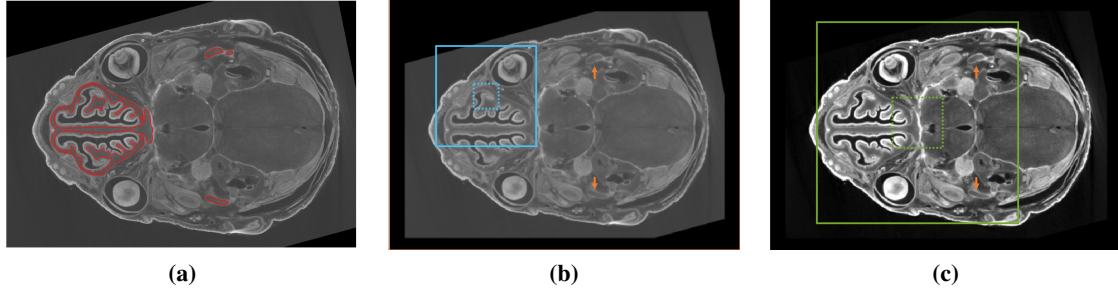
Figure 2: Example of the original slice with the craniofacial cartilage bounded in red (a), the same slice after the downsampling and padding to the 320×448 size (b) and after the contrast enhancement (c). Also, it is shown that the input size to the original V-Net (solid blue) was replaced by 256×256 size (solid green) in the xy plane which leads to major improvements in the segmentation of the problematic parts (orange arrows). Finally, it is demonstrated that without downsampling, the network would see only limited global information (dotted blue and green) insufficient for good results.

As the network performs piecewise segmentation, the pieces (patches) of predictions had to be put together in the postprocessing phase. Supposing algorithm taking care of dividing to patches, exactly reverse algorithm was used for assembling them. After that, upsampling by the factor of four was performed in the xy plane to obtain an estimate of the craniofacial cartilage segmentation in the original resolution.

3   BASELINE NETWORK AND LEARNING

The baseline network (Figure 3a) was inspired by V-Net [7] that was further modified to be more suitable for the given segmentation problem:

1) V-Net's input size is designed for small compact objects of interest which is not a case of the craniofacial cartilage. Thus, the input size was changed from $128 \times 128 \times 64$ to $256 \times 256 \times 24$ (Figure 2) implying sufficient global information in the xy plane as well as the presence of surrounding slices providing local information. With this arrangement, the 5th level was omitted from the architecture since the depth size in the 4th level was only 3.
2) Supposing only two prediction classes (craniofacial cartilage versus background), the softmax activation function was replaced by the sigmoid one for simplicity.

3) Assuming patches with very different contents, it is crucial to have sufficient batch size for effective learning. In V-net the decoder has twice as large capacity than the encoder which is rather unusual in the literature. Regarding this fact, the decoder capacity was experimentally reduced to the encoder one, which did not lead to decrease in training results. Therefore, the architecture was adjusted to have the same capacity in the encoder and the decoder part implying a bigger possible batch size and also a smaller chance of overfitting.

To prevent overfitting even more, data augmentation (random flipping, shifting, zooming and noise addition) was performed as well as weight decay. Nevertheless, even though the augmentation is beneficial for the final testing results, these more difficult inputs frequently caused trapping in local optima at the very beginning of training. This early-stage divergence phenomenon is identified in the literature and may be easily suppressed by a learning rate warmup strategy [4]. Specifically, the learning rate was linearly increased from zero to the required learning rate for a few initial steps implying a more careful movement in the optimization landscape as well as avoiding the divergence at the beginning of training. As far as the weight decay is concerned, it was proved that implementing weight decay through L2 regularization is inaccurate regarding adaptive learning rate optimizers as the parameters with bigger gradients are less regularized [5]. For that reason, the widely used Adam optimizer was replaced by AdamW [5] that penalizes the norm of all parameters equally.
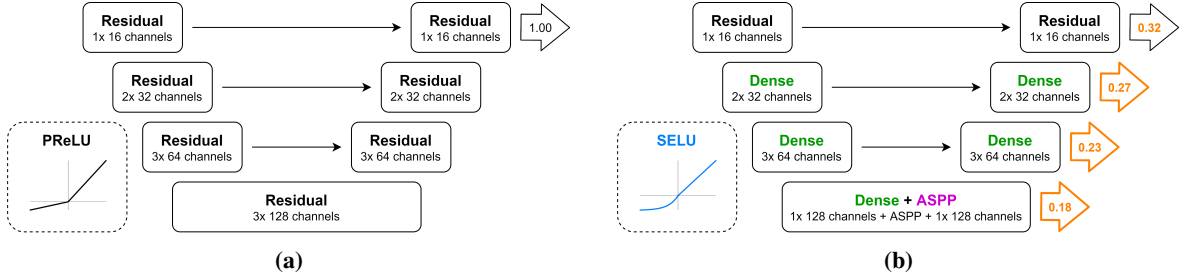
Figure 3: Comparison of the baseline CNN architecture (a) and the enhanced one (b). Hyperparameters of learning: batch size 8, Tversky loss (FN weight 0.7), AdamW (weight decay 1e−6), learning rate warmup from 0 to 1e−4 (initial 250 batches), after that decaying the learning by $e^{-0.15}$ each epoch.

4 NETWORK OPTIMIZATION

The baseline architecture was further optimized (Figure 3b and Graph 1) by incorporating the following architectural techniques:

Atrous spatial pyramid pooling In order to compensate the removal of the 5th level from V-Net and to add even more multi-scale information an atrous spatial pyramid pooling (ASPP) module [1] was placed in the lowest level of the CNN architecture. Proposed ASPP includes $1 \times 1 \times 1$ convolution as well as a series of $3 \times 3 \times 3$ convolutions with various dilation rates (1, 2, 4, 8) in the xy plane (no dilation was used for z axes since the depth size was only 3 in that level).

SELU activation function Scaled exponential linear unit (SELU) [3] is an activation function designed to have a so-called self-normalizing property mitigating the internal covariate shift problem. A great advantage of SELU over the other normalization techniques is no need for hyperparameter tuning as well as no dependency on the batch size.

Multi-output supervision To ensure more stable learning and a high quality of the features in the network's whole depth, the learning was dependent on the Tversky loss in all levels of the CNN architecture. To compensate various sizes of predictions in each level, the final loss was computed as the *weighted* average of the individual losses.

Dense blocks In each CNN architecture level, dense connections [2] were added to reuse the extracted features as much as possible.
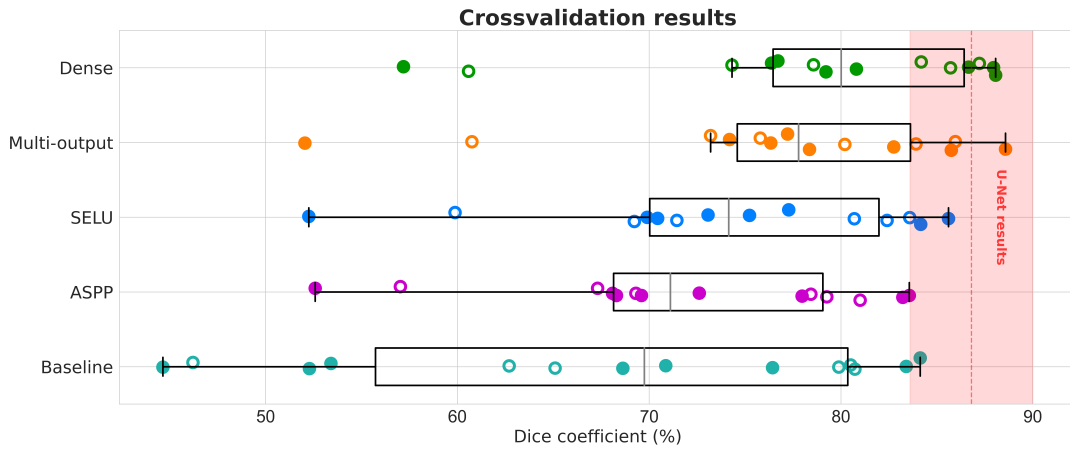
Graph 1: 7-fold crossvalidation results mapping the optimization process. The components on the *y* axis were added to the architecture in the previous step (e.g. SELU boxplot was created using baseline architecture with ASPP and SELU). Also, the comparison with the previous research (U-Net, chosen set of testing samples [6] represented by empty points) is depicted.

5 CONCLUSION

Based on a search of literature, the network learning as well as the architectural improvements were proposed raising the median of Dice coefficient from 69.74 % to 80.01 %. According to the results, in case of the oblong structures (such as the craniofacial cartilage) it is more advantageous to use a slice by slice segmentation instead of the 3D piecewise segmentation. Nonetheless, it was justified that the advanced architectural modifications remarkably improved the CNN performance, thus, we highly encourage to exploit them more in further research. Also, the proposed optimization process can be easily applied in U-Net and probably improve the benchmark for the given segmentation problem.

ACKNOWLEDGEMENT

REFERENCES

[1] CHEN, Liang-Chieh et al. DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs. *IEEE Transactions on Pattern Analysis and Machine Intelligence* [online]. 2018, **40**(4), 834–848 [cit. 2021-03-09]. ISSN: 1939-3539. *https://ieeexplore.ieee.org/document/7913730*.

[2] JEGOU, Simon et al. The One Hundred Layers Tiramisu: Fully Convolutional DenseNets for Semantic Segmentation. In: *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)* [online]. 2017, 1175–1183 [cit. 2021-03-09]. ISSN: 2160-7508. *https://ieeexplore.ieee.org/document/8014890*.

[3] KLAMBAUER, Günter et al. Self-Normalizing Neural Networks. In: *Advances in Neural Information Processing Systems* [online]. 2017, 30, 972-981 [cit. 2021-03-09]. ISSN: 1049-5258. *https://arxiv.org/abs/1706.02515*.

[4] LIU, Liyuan et al. On the Variance of the Adaptive Learning Rate and Beyond. Presented at: *International Conference on Learning Representations (ICLR) 2020* [online]. 2020 [cit. 2021-03-08]. *https://arxiv.org/abs/1908.03265*.

[5] LOSHCHILOV, Ilya & Frank HUTTER. Decoupled Weight Decay Regularization. Presented at: *International Conference on Learning Representations (ICLR) 2019* [online]. 2019 [cit. 2021-03-08]. *https://arxiv.org/abs/1711.05101*.

[6] MATULA, Jan. *Segmentace chrupavčité tkáně ve 3D mikro CT snímcích myších embryí* [online]. Brno, 2019, Diplomová práce (Ing.). VUT v Brně, FEKT, ÚBMI. Vedoucí práce: Ing. Jiří Chmelík. [cit. 2021-03-07]. *https://www.vutbr.cz/studenti/zav-prace/detail/118361?zp_id=118361*.

[7] MILLETARI, F., N. NAVAB & S. AHMADI. V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation. In: *2016 Fourth International Conference on 3D Vision (3DV)* [online]. 2016, p. 565–571. [cit. 2021-03-07]. ISSN: 2378-3826. *https://ieeexplore.ieee.org/document/7785132*.

PATHWAYS IDENTIFICATION IN THE GRAM-POSITIVE BACTERIUM ANEURINIBACILLUS SPECIES H1

Jana Musilová

Doctoral Degree Programme (2), FEEC BUT

E-mail: musilovajana@feec.vutbr.cz

Supervised by: Karel Sedlář

E-mail: sedlar@feec.vutbr.cz

Abstract: Here, we present the first insight into the Gram-positive, promising polyhydroxyalkanoates producer *Aneurinibacillus* species H1. Both static and dynamic properties are described in this paper with the aim on identification of pathways occurring in the organism. The genome consists of a circular, 3,663,644 bp long chromosome and contains 4,654 protein-coding sequences and 129 RNAs in total. The GC content is 44.8%. Functional properties identification showed that CDS divide into 26 categories with the most prevalent Amino Acids and Derivatives group. Pathways inference revealed 201 pathways. The most represented group is metabolic pathways, which is further divided into 12 groups.

Keywords: PHA, genome assembly, pathway, protein-coding sequences

1   INTRODUCTION

With the increasing pollution of the planet by waste, such as plastics as one of the biggest sources of waste, new ways must be found to avoid it. The pollution may be significantly reduced by replacing these harmful materials with environmentally friendly ones. Polyhydroxyalkanoates (PHA) are polyesters, also labelled as biodegradable plastics, produced by microorganisms such as *Aneurinibacillus* species H1 [1]. This thermophilic, Gram-positive bacterium can synthetize PHA as intracellular granules to store unused energy.

Although *Aneurinibacillus* sp. H1 is a promising bacterium in a way of PHA overproduction, almost nothing is known about its genomic and phenotypic properties. The genome assembly data is not available; functional properties such as information about metabolic pathways is missing as well. Therefore, we present the first insight into the genome as well as its static and functional properties.

2   MATHERIALS METHODS

2.1   GENOME ASSEMBLY AND ANNOTATION

The genome assembly was based on the sequencing data. The DNA sequences were obtained using the Oxford Nanopore technology, especially the MinION platform [2]. At first, reads were basecalled using the Guppy basecaller v3.4.4 [3] and quality of the reads was checked by pycoQC v2.5.2 [4]. In the next step, the long-reads assembler Canu [5] was used to assemble individual reads. Finally, the sequence was polished using two tools. Firstly, the Racon [6] was used to create the initially polished assembly, and subsequently, the medaka [7] tool generated the final polished genome. The PAF files necessary for the polishing step were generated using minimap2 [8].

The assembled sequence was stored into a fasta file. Subsequently, the file and the RAST Server [9] were used for the genome annotation, such as identification of protein-coding sequences (CDS), RNAs, as well as genome length or GC content.

2.2 FUNCTIONAL ANNOTATION AND PATHWAYS IDENTIFICATION

Functional categorization of the genome features was done using the BlastKOALA [10] as well as the KO entries assignment. For visualization of the results, the SEED Viewer [11] was used. In a way to identify metabolic pathways in the bacterium, the KEGG Metabolic Analysis was done using the SEED Viewer and the KEGG Mapper [12].

3 RESULTS AND DISCUSSION

3.1 GENOME OVERVIEW

Genome of the *Aneurinibacillus* species H1 was reconstructed from 372,378 reads containing 2.7 billion bases in total. A median read length was 3,320 bases and a median Phred score corresponds to the value $Q \approx 9.6$.

The whole assembly process resulted into the assembly consisting of one circular chromosome with a coverage exceeding 731×. The genome size is slightly above the average of gram-positive bacteria [13] with 3,663,644 bases but it is of the order of magnitude corresponding. The GC content with a value 44.8 corresponds to the gram-positive bacteria average. In the genome, a total number of 129 RNAs was discovered.

The number of CDS is 4,654 of which 1,977 functional parts were identified as identical to the model organism *Bacillus subtilis* subsp. 168. This information will help in next research of the *Aneurinibacillus* sp. H1, as almost half of processes can be studied on the well-described bacterium instead of conducting new laboratory experiments. Genomic properties are shown in the **Figure 1**.
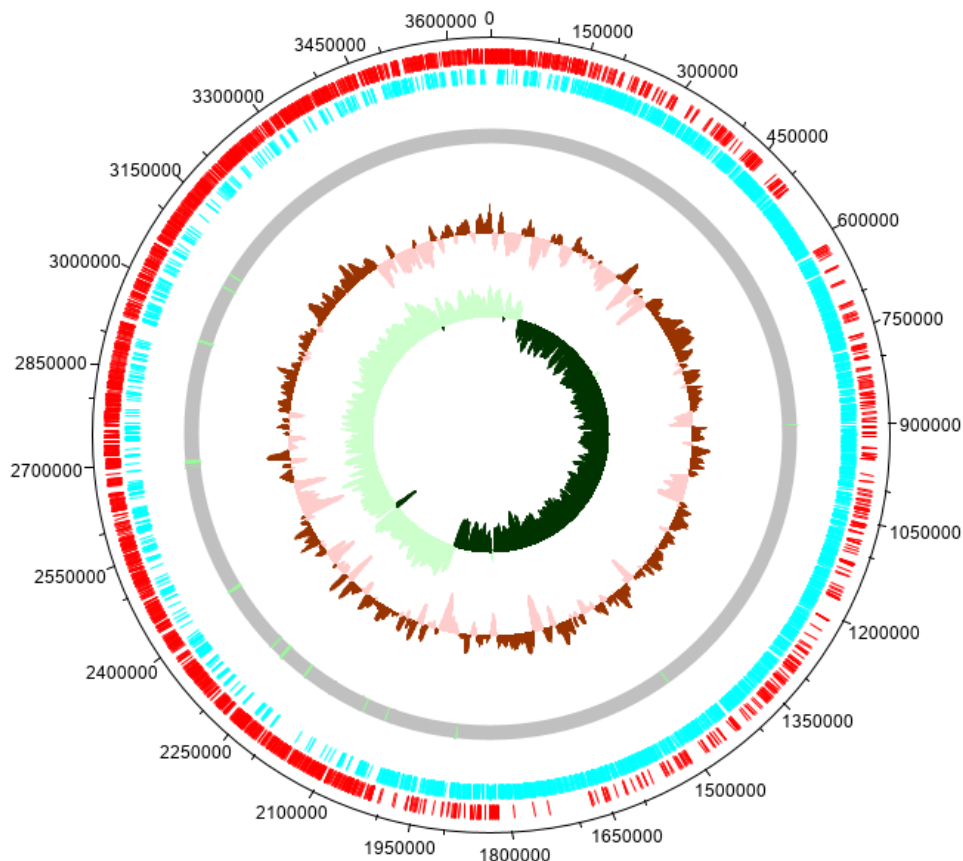
Figure 1: A chromosomal map of the *Aneurinibacillus* sp. H1 genome. The first (red) and second (blue) outermost circles represent CDSs on the forward and backward strands, respectively. The third circle (grey) represents pseudogenes. The inner circles represent the GC content and GC skew.

112

3.2   FUNCTIONAL PROPERTIES AND IDENTIFIED PATHWAYS

Protein-coding sequences were classified into 26 functional systems. In total, 1,006 CDS were assigned to particular categories. The most prevalent group with 21% of assigned CDS is the "Amino Acids and Derivatives" category, followed by the groups "Carbohydrates" with 12%, "Protein Metabolism" corresponding to 11%, and "Cofactors, Vitamins, Prosthetic Groups, Pigments" group with 10% of all assigned protein-coding sequences. The **Figure 2** displays all discovered groups together with the percentage.
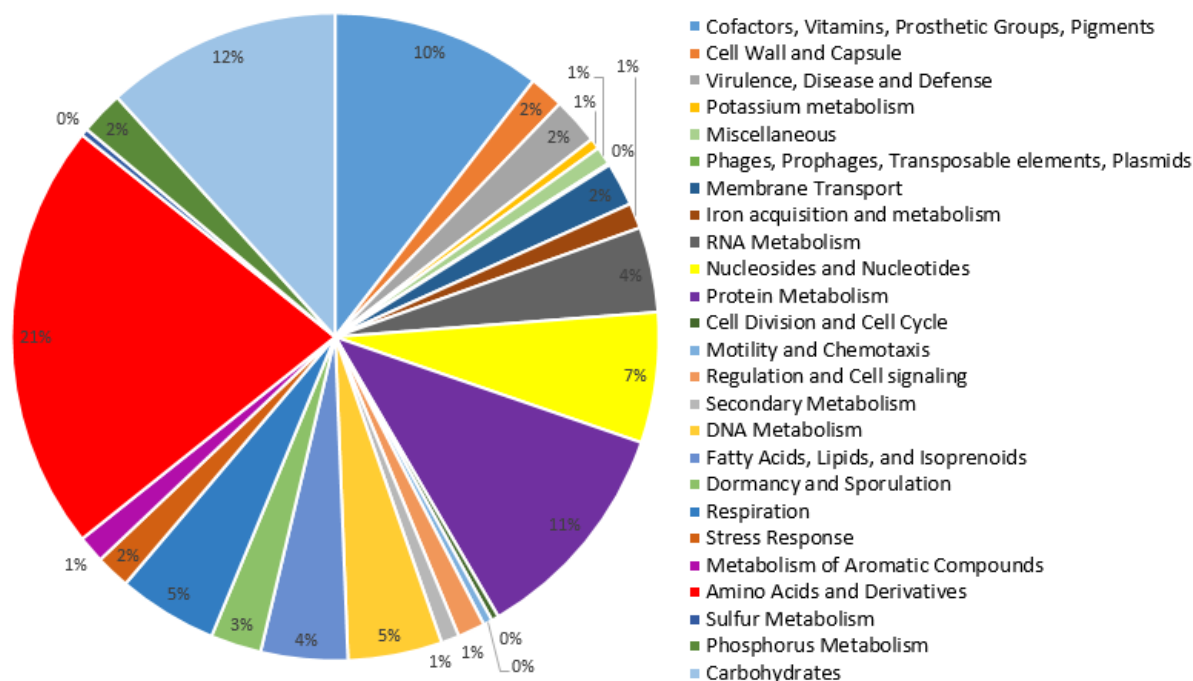
Figure 2: A pie chart of the *Aneurinibacillus* sp. H1 functional properties.

Aneurinibacillus species H1 was associated to 201 pathways according to the KEGG encycledia. The most pathways belong to metabolism. In the metabolism category, 12 groups were discovered, each of them is divided into several subgroups. Individual groups together with a subgroup example (along with the KO entry and the number of associated processes) are shown in the Table 1. Map of bacterium´s metabolic pathways is displayed in the **Figure 3**.

Besides metabolic pathways, bacterial properties were also assigned to the pathways "Genetic Information Processing"; "Environmental Information Processing"; "Cellular Processes"; or "Organismal Systems".

Table 1: Metabolic pathways associated with the *Aneurinibacillus* sp. H1 bacterium according to the KEGG encyclopedia.

Global and overview maps	Metabolism of other amino acids
01100 Metabolic pathways (566)	00410 beta-Alanine metabolism (8)
01110 Biosynthesis of sec. mets. (243)	00430 Taurine and hypotaurine metabolism (5)
Carbohydrate metabolism	**Glycan biosynthesis and metabolism**
00010 Glycolysis / Gluconeogenesis (25)	00540 Lipopolysaccharide biosynthesis (1)
00020 Citrate cycle (TCA cycle) (21)	00541 O-Antigen and sugar biosynthesis (13)
Energy metabolism	**Metabolism of cofactors and vitamins**
00190 Oxidative phosphorylation (41)	00730 Thiamine metabolism (14)
00195 Photosynthesis (8)	00740 Riboflavin metabolism (7)
Lipid metabolism	**Metabolism of terpenoids and polyketides**
00061 Fatty acid biosynthesis (10)	00900 Terpenoid backbone biosynthesis (12)
00071 Fatty acid degradation (9)	00981 Insect hormone biosynthesis (1)

Nucleotide metabolism	Biosynthesis of other secondary metabolites
00230 Purine metabolism (39)	00940 Phenylpropanoid biosynthesis (1)
00240 Pyrimidine metabolism (28)	00950 Isoquinoline alkaloid biosynthesis (1)
Amino acid metabolism	**Xenobiotics biodegradation and metabolism**
00250 Alanine, aspartate and glutamate (26)	00362 Benzoate degradation (14)
00260 Glycine, serine and threonine (26)	00627 Aminobenzoate degradation (6)
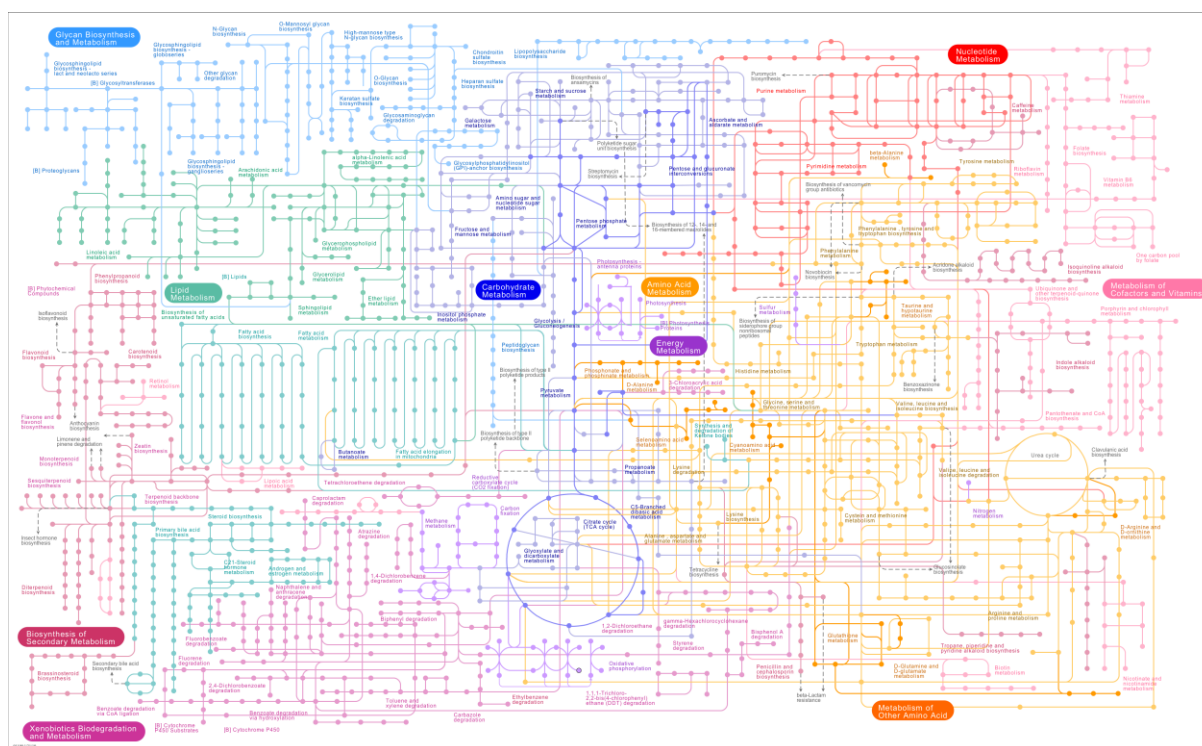
Figure 3: A map of all metabolic pathways discovered in the *Aneurinibacillus* sp. H1.

4 CONCLUSIONS

In this study, we revealed basic genomic and phenotypic properties of the polyhydroxyalkanoates producing bacterium *Aneurinibacillus* species H1 with the aim to discover pathways occurring in the bacterium.

In the first part, we described static bacterium´s properties. As the genome was not published yet, we assembled the complete genome in the first step, for which we used Oxford Nanopore, MinION sequencing data. The assembled genome consists of one circular chromosome with the size of 3,663,644 bases. The genome contains 4,654 protein-coding sequences in total, 129 RNAs and the GC content is 44.8. Comparing to the average of Gram-positive bacteria, the size is slightly above an average 2.53 Mbp, but GC content on average 49.4, which indicates that the genome has been assembled correctly. Nevertheless, average sequencing quality of Nanopore sequencing is around 90 % and polishing the sequence with Illumina sequencing data is needed before the genome can be reliably compared to other strains.

Description of functional properties of the *Aneurinibacillus* sp. H1 bacterium was the second part of this study. We classified in total 1,066 protein-coding sequences into 26 groups with most preva- lent group Amino Acids and Derivatives. In the last step, we identified pathways occurring in the bacterium according to the KEGG encyclopedia. The pathways inference revealed 201 pathways with the most represented metabolic pathways divided into 12 groups.

REFERENCES

[1] SEDLACEK P, I PERNICOVA, I NOVACKOVA, X KOURILOVA, M KALINA, A KOVALCIK, M KOLLER, J NEBESAROVA, V KRZYZANEK, K HRUBANOVA, J MASILKO, E SLANINOVA, M TRUDICOVA and S OBRUCA. Introducing the Newly Isolated Bacterium Aneurinibacillus sp. H1 as an Auspicious Thermophilic Producer of Various Polyhydroxyalkanoates (PHA) Copolymers–2. Material Study on the Produced Copolymers. *Polymers*, 2020, **12**(6), 1298. doi:10.3390/polym12061298

[2] JAIN M, HE OLSEN, B PATEN and M AKESON. The Oxford Nanopore MinION: delivery of nanopore sequencing to the genomics community. *Genome Biology*, 2016, **17**(1), 239. doi:10.1186/s13059-016-1103-0

[3] *Community - Protocol - Guppy* [online, accessed 2021-03-06]. Available at: https://community.nanoporetech.com/protocols/Guppy-protocol/v/GPB_2003_v1_revU_14Dec2018

[4] LEGER A and T LEONARDI. pycoQC, interactive quality control for Oxford Nanopore Sequencing. *Journal of Open Source Software*, 2019, **4**(34), 1236. doi:10.21105/joss.01236

[5] KOREN S, BP WALENZ, K BERLIN, JR MILLER, NH BERGMAN and AM PHILLIPPY. Canu: Scalable and accurate long-read assembly via adaptive κ-mer weighting and repeat separation. *Genome Research*, 2017, **27**(5), 722–736. doi:10.1101/gr.215087.116

[6] VASER R, I SOVIĆ, N NAGARAJAN and M ŠIKIĆ. Fast and accurate de novo genome assembly from long uncorrected reads. *Genome Research*, 2017, **27**(5), 737–746. doi:10.1101/gr.214270.116

[7] *GitHub - nanoporetech/medaka: Sequence correction provided by ONT Research* [online, accessed 2021-03-06]. Available at: https://github.com/nanoporetech/medaka

[8] HENG LI. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics*, 2018, **34**(18), 3094–3100. doi:10.1093/bioinformatics/bty191

[9] AZIZ RK, D BARTELS, A BEST, M DEJONGH, T DISZ, RA EDWARDS, K FORMSMA, S GERDES, EM GLASS, M KUBAL, F MEYER, GJ OLSEN, R OLSON, AL OSTERMAN, et. al. The RAST Server: Rapid annotations using subsystems technology. *BMC Genomics*, 2008, **9**, 75. doi:10.1186/1471-2164-9-75

[10] KANEHISA M, Y SATO and K MORISHIMA. BlastKOALA and GhostKOALA: KEGG Tools for Functional Characterization of Genome and Metagenome Sequences. *Journal of Molecular Biology,* 2016, **428**(4), 726–731. doi:10.1016/j.jmb.2015.11.006

[11] DAVIS JJ, AR WATTAM, RK AZIZ, T BRETTIN, R BUTLER, RM BUTLER, P CHLENSKI, N CONRAD, A DICKERMAN, EM DIETRICH, JL GABBARD, S GERDES, A GUARD, RW KENYON, et. al. The PATRIC Bioinformatics Resource Center: Expanding data and analysis capabilities. *Nucleic Acids Research*, 2020, **48**(D1), D606–D612. doi:10.1093/nar/gkz943

[12] KANEHISA M and Y SATO. KEGG Mapper for inferring cellular functions from protein sequences. *Protein Science*, 2020, **29**(1), 28–35. doi:10.1002/pro.3711

[13] XIU QING LI and DU DONGLEI. Variation, evolution, and correlation analysis of C+G content and genome or chromosome size in different kingdoms and phyla. *PLoS ONE*, 2014, **9**(2), 88339. doi:10.1371/journal.pone.0088339

OBJECT DETECTION NETWORKS FOR LOCALIZATION AND CLASSIFICATION OF INTRACRANIAL HEMORRHAGES

Jakub Nemcek

Doctoral Degree Programme (1), FEEC BUT

E-mail: nemcek@vutbr.cz

Supervised by: Roman Jakubicek

E-mail: jakubicek@feec.vutbr.cz

Abstract: Intracranial hemorrhages represent life-threatening brain injuries. This paper presents two state-of-the-art object detection systems (Faster R-CNN and YOLO v2) which are trained to localize and classify hemorrhages in axial head CT slices by providing labelled rectangular bounding boxes. Publicly available datasets of head CT data and ground truth bounding boxes are used to evaluate and compare the performance of both detectors. The Faster R-CNN shows better results by achieving an average Jaccard coefficient of 58.7 %.

Keywords: intracranial hemorrhage, convolutional neural network, localization, classification, computed tomography, Faster R-CNN, YOLO v2

1 INTRODUCTION

Intracranial hemorrhages (ICHs) represent critical medical event as they belong to life-threatening brain injuries with high incidence (25 cases per 100,000 persons per year), which can originate in physical trauma or other non-traumatic causes (hemorrhagic stroke) [1]. According to its location, there are five types of ICH: intraparenchymal (IPH), intraventricular (IVH), subdural (SDH), epidural (EDH) and subarachnoid hemorrhage (SAH).

Significance of this pathology is given by the risk of lasting severe consequences and high mortality (40 %), hence early and accurate diagnosis is crucial [2]. The x-ray computed tomography (CT) is the most frequent ICH diagnostic method, however, the traditional examination of 3D CT data by radiologists is either time-consuming and requires high level of concentration. Nowadays, computer-aided diagnosis (CAD) systems are being developed to reduce the examination time and to prevent possible omissions in case of small or indistinct bleedings [3].

Novel deep learning methods have become state-of-the-art in many applications including image analysis. Several approaches for analysis of ICH in brain CT data using convolutional neural networks (CNN) have been recently published. In some papers [4, 5] simple 2D classification CNN is used for ICH detection and classification. Some authors use combination of 2D CNN architecture with other methods to extend the detection and classification results to whole 3D scan. In these cases, random forests [6] and recurrent neural networks (long short-term memory - LSTM) [3, 7] were used for the extension. More precise localization of ICH is done by segmentation. The authors in [8] proposed method using U-Net-like architecture, but the ICH type is not determined. A cascade of classification CNN ahead of a segmentation fully convolutional network was used in [9]. In [2] authors proposed mask R-CNN model for ICH segmentation. Although precise voxel-level segmentation might be favourable for a radiologist, training of such algorithms demands voxel-level annotations. Gaining of such annotations for sufficiently large dataset is, however, very time-consuming. In [10] a set of 2D CNN-based classifiers with cascade-parallel architecture was proposed, that enables ICH classification and localization by labelled 3D bounding box (BB).

In this paper, two automatic methods for the localization and classification of ICHs are proposed and compared. In comparison to other recently published methods, two detection networks are trained to predict labelled 2D BBs for a delineation of the bleedings and for the ICH types determination in axial slices, using publicly available datasets of images and annotations. Such an automatic method might lead to a significant decrease in the duration of the diagnostic process (manual detection can take more than 5 minutes) and false negative diagnostic conclusions.

2 METHODS

The combination of a CNN and a region of interest prediction network has recently proved to be an efficient solution to object detection [13, 14]. Thus two state-of-the-art detection networks are chosen to localize and classify ICHs in axial slices of a CT image.

2.1 EXPERIMENTAL DATA

The head CT data from the publicly available dataset CQ500 [6] was used with respect to several conditions: only non-contrast soft tissue kernel series with thick slices (3 mm or more) containing at least one type of hemorrhage were chosen (191 CT scans). The dataset is, however, imbalanced, considering the amount of data with different ICH types. In case of EDH, the amount is insufficient (only 13 scans). Manually extracted axial slice-level bounding box annotations marked by three neuroradiologists are obtained from BHX dataset [11] publicly available on the PhysioNet [12].

A simple pre-processing is applied on individual axial slices of the 3D image in the form of contrast enhancement with respect to three special radiological windows (i.e., brain, subdural and bone) to form three-channel input into the detection networks. The slices are resized to the size of 224×224 pixels.

2.2 DESIGN OF DETECTORS

Two modern detection network architectures are chosen for the purpose of ICH localization and classification: Faster R-CNN [13] and YOLO v2 [14]. Both networks consist of a feature extraction CNN and a detection part. Faster R-CNN uses a region proposal network (RPN) to generate region proposals from the feature space, that are then refined and classified [13]. YOLO v2 is a single neural network applied to the full image which divides an image into regions and then predicts BBs and classes for them [14]. Both architectures provide results in the form of labelled rectangular bounding boxes denoting the precise positions of the ICHs in the image.

Table 1: Training options for both detection networks: number of epochs, mini-batch size, L2 regulation factor and learning rate drop period with the factor of 0.1

	Epochs	Mini-batch	L2 reg.	lr. drop
Faster R-CNN	24	12	0.0005	15
YOLO v2	35	42	0.0005	30

2.3 IMPLEMENTATION DETAILS

The proposed methods were implemented in Matlab. For both detectors, ResNet-50 [15] pre-trained on the ImageNet database was used as a feature extraction CNN. The training dataset consisted of axial slices from 80 % of the available CT scans. An augmentation of the training images is applied in the form of slight rotation, translation, crop, scaling, reflection, and random noise adding. The Adam optimizer [16] with an initial learning rate of 0.001 ($\beta_1 = 0.9$ and $\beta_2 = 0.999$) was used for training of both networks. Other training parameters are shown in Table 1.
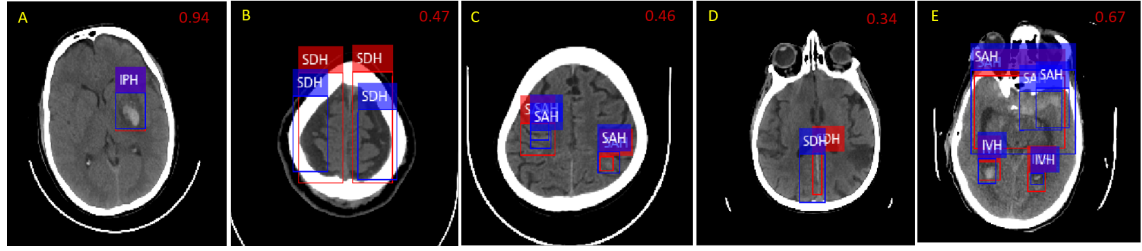
Figure 1: The visual display of predicted labelled BBs (blue) compared to ground truth (red) together with the IoU value (in the right upper corner). A – BB of IPH matching well with the ground truth, B – results of SDH localization on both sides of the head with the total IoU for both bleedings, C – two localized SAH bleedings, D – poorly localized SDH with insufficient overlap of the BB and the ground truth, E – localization of multiple hemorrhage types in one slice with the total IoU for both types jointly.

3   RESULTS AND DISCUSSION

Axial slices from twenty percent of the available CT scans were randomly selected for the evaluation of the detection models regarding the distribution of individual ICH types in the training dataset. The Jaccard coefficient (known as the intersection over union – IoU), which expresses the relative overlap of the ground truth and predicted BB, was chosen as the evaluation metric together with the sensitivity (Se) and the positive predictive value (PPV). To determine the number of true positive (TP), false positive (FP) and false negative (FN) results, threshold value of the IoU was chosen as 0.4 IoU taking into account the properties of the metric – a slight difference in the position of BBs leads to a distinct decrease in the IoU value, although subjectively it seems to be a correct localization (Fig. 1 B C D).

Table 2: Achieved results of the proposed detection networks on the testing dataset and their comparison to recently published 3D BB predicting method. The second row represents relative amount of the ICH type among the testing images – the bold value (multiple types might be present in an image). The sensitivity (Se) and positive predictive value (PPV) are calculated with a threshold value of 0.4 IoU. The bold values in the last column represent the average value through all the testing images.

		IPH	IVH	SDH	SAH	Total ICHs
	cases	32 %	5 %	61 %	29 %	**766**
Faster R-CNN	**IoU [%]**	67.4	60.9	62.4	47.4	**58.7**
	$\text{Se}_{0.4}$ **[%]**	78.0	69.5	74.9	49.6	**69.1**
	$\text{PPV}_{0.4}$ **[%]**	82.2	80.8	79.0	62.5	**76.5**
YOLO v2	**IoU [%]**	59.7	53.8	54.3	41.6	**49.1**
	$\text{Se}_{0.4}$ **[%]**	71.3	60.0	61.6	28.6	**57.3**
	$\text{PPV}_{0.4}$ **[%]**	82.3	60.0	80.7	31.9	**69.9**
Nemcek et al. [10]	**IoU [%]**	59.0	53.5	43.2	52.5	**53.7**
	$\text{Se}_{0.4}$ **[%]**	65.0	62.5	44.4	50.0	**57.3**
	$\text{PPV}_{0.4}$ **[%]**	81.7	83.3	78.0	61.5	**77.1**

Regarding the achieved results (Table 2), the Faster R-CNN shows better ability to correctly locate and classify ICH, than the YOLO v2. However, considering the achieved sensitivities of both networks, it is evident that the FN cases may occur, thus the ICHs are not detected, or more frequently, the localization is inaccurate (Fig. 1 D). Despite the insufficient overlap, such a result would be highly

beneficial for a radiologist as it would call the attention to the bleeding. In comparison to 3D BB-predicting method published in [10], Faster R-CNN generally achieved better results (Table 2).

The detectors can localize and classify various ICH types even in one slice (Fig. 1 E) despite their different sizes, shapes, and positions. The high variability of shapes and possible locations of SAH (Fig. 1 C E) probably affected the detection results (Table 2) of both detectors in comparison to other ICH types.

The trained detectors show an ability to roughly delineate and classify the bleeding, and thus they might minimize the chance of missing an ICH by oversight. The process of manual ICH diagnostics consists in axial slices evaluation, hence the 2D BB-based localization in individual axial slices appears to be convenient for radiologists. In contrast to 3D BB output (as in [10]), 2D BBs may delineate the superior and inferior margins of one bleeding more precisely. Moreover, both real-time object detection systems provide very fast performance. [13, 14] An implementation of such an algorithm in a CAD system might lead to a distinct decrease in diagnostic time, hence, to saving lives or preventing from lasting consequences.

4 CONCLUSION

This paper presents an implementation of two automatic ICH localization and classification systems, that provide labelled 2D rectangular bounding boxes as an output. Two state-of-the-art object detection networks (Faster R-CNN and YOLO v2) were trained, evaluated, and compared using publicly available datasets of head CT image and annotations. An average Jaccard coefficient of 58.7 % was achieved by the Faster R-CNN, that globally showed better results on the available dataset.

REFERENCES

[1] CACERES, J. Alfredo a Joshua N. GOLDSTEIN. Intracranial Hemorrhage. *Emergency Medicine Clinics of North America* [online]. 2012, 30(3), 771-794 [cit. 2021-03-08]. ISSN 07338627. doi:10.1016/j.emc.2012.06.003

[2] CHANG, P.D., E. KUOY, J. GRINBAND, et al. Hybrid 3D/2D Convolutional Neural Network for Hemorrhage Evaluation on Head CT. *American Journal of Neuroradiology* [online]. 2018, 39(9), 1609-1616 [cit. 2021-03-08]. ISSN 0195-6108. doi:10.3174/ajnr.A5742

[3] GREWAL, Monika, Muktabh Mayank SRIVASTAVA, Pulkit KUMAR a Srikrishna VARADARAJAN. RADnet: Radiologist level accuracy using deep learning for hemorrhage detection in CT scans. In: *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)* [online]. IEEE, 2018, 2018, s. 281-284 [cit. 2021-03-08]. ISBN 978-1-5386-3636-7. doi:10.1109/ISBI.2018.8363574

[4] LEE, Hyunkwang, Sehyo YUNE, Mohammad MANSOURI, et al. An explainable deep-learning algorithm for the detection of acute intracranial haemorrhage from small datasets. *Nature Biomedical Engineering* [online]. 2019, 3(3), 173-182 [cit. 2021-03-08]. ISSN 2157-846X. doi:10.1038/s41551-018-0324-9

[5] KARKI, Manohar, Junghwan CHO, Eunmi LEE, et al. CT window trainable neural network for improving intracranial hemorrhage detection by combining multiple settings. *Artificial Intelligence in Medicine* [online]. 2020, 106 [cit. 2021-03-08]. ISSN 09333657. doi:10.1016/j.artmed.2020.101850

[6] CHILAMKURTHY, Sasank, Rohit GHOSH, Swetha TANAMALA, et al. Deep learning algorithms for detection of critical findings in head CT scans: a retrospective study. *The Lancet* [online]. 2018, 392(10162), 2388-2396 [cit. 2021-03-08]. ISSN 01406736. doi:10.1016/S0140-6736(18)31645-3

[7] PATEL, Ajay, Sil. C. VAN DE LEEMPUT, Mathias PROKOP, Bram VAN GINNEKEN a Rashindra MANNIESING. Image Level Training and Prediction: Intracranial Hemorrhage Identification in 3D Non-Contrast CT. *IEEE Access* [online]. 2019, 7, 92355-92364 [cit. 2021-03-08]. ISSN 2169-3536. doi:10.1109/ACCESS.2019.2927792

[8] PATEL, Ajay, Floris H. B. M. SCHREUDER, Catharina J. M. KLIJN, et al. Intracerebral Haemorrhage Segmentation in Non-Contrast CT. *Scientific Reports* [online]. 2019, 9(1) [cit. 2021-03-08]. ISSN 2045-2322. doi:10.1038/s41598-019-54491-6

[9] CHO, Junghwan, Ki-Su PARK, Manohar KARKI, et al. Improving Sensitivity on Identification and Delineation of Intracranial Hemorrhage Lesion Using Cascaded Deep Learning Models. *Journal of Digital Imaging* [online]. 2019, 32(3), 450-461 [cit. 2021-03-08]. ISSN 0897-1889. doi:10.1007/s10278-018-00172-1

[10] NEMCEK, Jakub, Roman JAKUBICEK, Jiri CHMELIK. Localization and Classification of Intracranial Hemorrhages in CT Data. *8th European Medical and Biological Engineering Conference* [online]. Springer International Publishing, 2021, 2021-11-30, s. 767-773 [cit. 2021-03-08]. IFMBE Proceedings. ISBN 978-3-030-64609-7. doi:10.1007/978-3-030-64610-3_86

[11] REIS, E. P., F. Nascimento, M. Aranha, et al. Brain Hemorrhage Extended (BHX): Bounding box extrapolation from thick to thin slice CT images (version 1.1). *PhysioNet* [online]. 2020, [cit. 2021-03-08]. https://doi.org/10.13026/9cft-hg92

[12] GOLDBERGER, A., L. Amaral, L. Glass, et al. PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals. *Circulation* [Online]. 2000, 101 (23), pp. e215-e220 [cit. 2021-03-08].

[13] REN, Shaoqing, Kaiming HE, Ross GIRSHICK a Jian SUN. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence* [online]. 2017, 39(6), 1137-1149 [cit. 2021-03-12]. ISSN 0162-8828. doi:10.1109/TPAMI.2016.2577031

[14] REDMON, Joseph a Ali FARHADI. YOLO9000: Better, Faster, Stronger. In: *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* [online]. IEEE, 2017, 2017, s. 6517-6525 [cit. 2021-03-12]. ISBN 978-1-5386-0457-1. doi:10.1109/CVPR.2017.690

[15] HE, Kaiming, Xiangyu ZHANG, Shaoqing REN a Jian SUN. Deep Residual Learning for Image Recognition. In: *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* [online]. IEEE, 2016, 2016, s. 770-778 [cit. 2021-03-12]. ISBN 978-1-4673-8851-1. doi:10.1109/CVPR.2016.90

[16] Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. CoRR abs/1412.6980 (2015) [cit. 2021-03-12]

AUTOMATIC ADAPTIVE OPTICS RETINAL IMAGES MONTAGING

Eva Valterova

Doctoral Degree Programme (3.), FEEC BUT

E-mail: valterova@vutbr.cz

Supervised by: Radim Kolar

E-mail: kolarr@feec.vutbr.cz

Abstract: The adaptive optics images capture limited field of view. Nevertheless, the quantitative retina assessment requires analysis over extended areas captured by images on various retina positions. The fully automated method for image registration and montaging is presented. The method utilizes scale invariant feature transform (SIFT) for feature extraction from preprocessed images. The method is tested on 200 images of normal healthy patients. The montages of 20 eyes were created and evaluated by normalized mutual information metrics and the results showed high alignment accuracy.

Keywords: retina, adaptive optics, image registration, SIFT

1 INTRODUCTION

Adaptive optics (AO) technology offers high resolutions retinal imaging. The wavefront sensor and deformable mirror measure and correct optical aberrations present in the optic path of the human eye [1]. Thus the resolution near diffraction is reached and individual retinal cell types, such as ganglion cells, photoreceptors and retinal pigment epithelial cells are visible *in vivo* [2, 3]. The structure and function of retinal cell observation are possible to utilize as a relevant marker of pathological retinal changes or disease monitoring [1, 4, 5].

The AO imaging struggle with a limited field of view, typically of size $4°$ by $4°$, which corresponds to approximately 1.1 mm^2. Nevertheless, larger retinal area acquisition is necessary for the comprehensive examination, thus multiple adjacent images must be acquired across overlapping retinal regions and montaged to cover a larger retina area. The montage requires accurate registration of individual adjacent images. Generally, the montage can be created in Photoshop (Adobe Systems, Mountain View, California) [6], ImageJ (National Institutes of Health, Bethesda, Maryland) [7] or other programs based on manual registration. However, this approach is highly time-consuming, requires an experienced evaluator, and is moreover complicated by image noise, blur, and contrast variability across images.

Several automated solutions for adaptive optics scanning laser ophthalmoscopy (AOSLO) image montaging have been published. The registration part of montaging is mostly based on image transformation estimation from detected specific features [8, 9, 10, 11]. *Li et al.* [9] used principal component analysis-scale invariant features (PCA-SIFT) to match horizontally AOSLO adjacent images. The outlier features gained by SIFT are eliminated by the random sample consensus (RANSAC) algorithm. The horizontally montaged images have inhomogeneous illumination, thus were registered in one wide-field montage by cross-correlation. Inspired by this approach, an automated algorithm for AOSLO images was developed by *Chen et al.* [10]. The registration accuracy was evaluated by two metrics - normalized cross-correlation (NCC) and normalized mutual information (NMI). Later, the same group [11] published an approach, where the SIFT features detection is replaced by finding the

center locations of the cones.

This paper describes an automated method for AO image registration into montage upon the method presented by *Chen et al.* [10] for AOSLO image montaging. The method is adapted for AO data, which in comparison with AOSLO images, have a larger field of view and therefore have lower resolution. Also, an AO image is acquired by a flood illumination camera, in contrast to an AOSLO image, which is raster scanned across the retina and thus its contrast varies more than the contrast of the AO image. The method results are evaluated by NMI metric.

2   MATERIALS AND METHODS

2.1   AO IMAGE ACQUISITION

The images were acquired by AO retinal camera with flood illumination (rtx1, Imagine Eyes, Orsay, France). The size of each retinal image is 1500×1500 pixels and field of view $4° \times 4°$. AO images were obtained from both eyes on 10 normal healthy subjects. From each of twenty eyes were acquired 10 images, which correspond to different retinal locations. One image is focused on the center of the fovea, afterward three images in the temporal direction from the fovea area and two images in each meridian (nasal, inferior, and superior) are acquired. The step between images was chosen to 2°for sufficient overlap of adjacent images (see an example of final montage in Figure 3 - b). Based on the acquisition approach, the preliminary position of each image is known, but is not sufficiently accurate for image montage, due to minor eye movements and unstable eye fixation (see Figure 2 - a).

2.2   PROBLEM FRAMEWORK AND PREPROCESSING

The AO image capturing the fovea area defines the origin of the coordinate system for each eye and all other images from an eye are registered to that central image or to one of adjacent overlapping images, which is the closest to the central position and thus is already registered in the previous iteration.

The overlapping parts from reference x_{ref} and moving image x_{mov} are cropped and shown in Figure1. The images are characteristic by inhomogenous blur and also darker and lighter areas, which typically corresponds to vessel areas and adjacent light-reflective photoreceptor groups, respectively. The blur variability across images differs, as is shown in Figure1 - a,b, and thus complicates the registration. Hence the local blur effect in images is suppressed by an averaging filter with a window of size 15. In opposite, the darker and lighter areas are affected by blur barely and mutually correspond in images. Therefore the histogram equalization is used to highlight these areas in x_{ref} and x_{mov}.

In the next step the characteristic image features, similar for both images, are detected to find an alignment transformation between the x_{ref} and x_{mov}.

2.3   SCALE INVARIANT FEATURE TRANSFORMS (SIFT)

Correct and accurate transformation function determination is dependent on the detection of similar key locations as characteristic features in both images. The SIFT algorithm [12] is utilized for identifying these characteristic features, due to its advantage in robustness to illumination and local affine distortions. The features from x_{ref} and x_{mov} are matched to corresponding pairs and the Euclidean distance between two feature locations in each pair is computed. Obtained matching features are shown in Figure 1 - c,d as lines connecting the green points, which denotes the feature position. As is shown, numerous points are matched incorrectly. Thus the Euclidean distance of each match is rounded to tens and then the matches with Euclidean distance, which have the highest incidence, are considered as correct and used for computation of transformation matrix. The correct matches are denoted in Figure1 by red lines. In transformation is considered the influence of shift, rotation, and
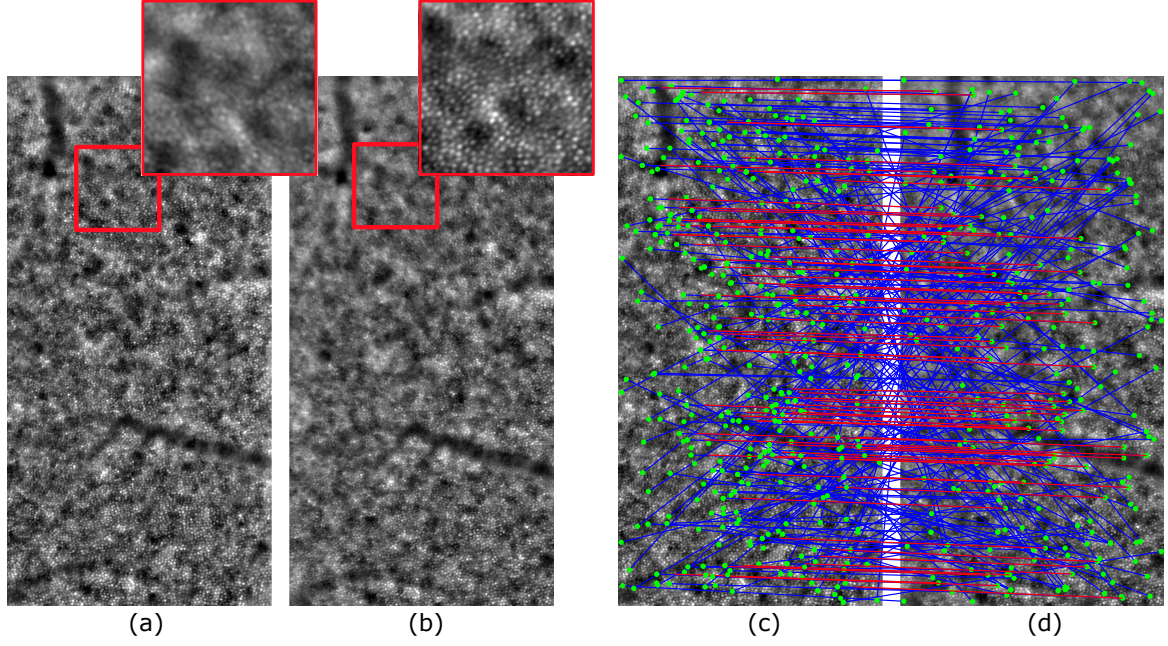
Figure 1: Overlapping parts from (a) reference image x_{ref} and (b) moving image x_{mov} with zoom in areas, which highlight blur variability. Second set of images show detected features, marked by green points determining position and matched by blue lines in (c) x_{ref} and (d) x_{mov}. The red lines denotes correct matches used for transformation matrix computation

scaling. The moving image is transformed by a computed transformation matrix and thus is registered to the coordinate system.

2.4 ALIGNMENT ACCURACY

The registration accuracy assessment is based on image intensities in the overlapping regions. The normalized mutual information (NMI) [13] is used for image similarity evaluation. This metric is robust to systematic intensity variations, therefore is suitable for AO images. The measure is given by

$$NMI = \frac{H(X) + H(Y) - H(X,Y)}{\sqrt{H(X)H(Y)}},$$ (1)

where $H(X)$ and $H(Y)$ are computed from marginal and joint probability density of the possible image intensities in images X and Y. NMI ranges from 0 to 1, where 1 represents perfect alignment.

3 RESULTS AND DISCUSSION

Each of 9 of 10 images from an eye was registered (1 image capturing fovea is allocating origin of the coordinate system). The registration was performed by transformation matrix determined by matched SIFT points. The example of image matching is shown in Figure 2. The images shows x_{ref} and x_{mov} overlaid in different color bands. Gray regions in the composite image show where the two images have the same intensities. Magenta and green regions show where the intensities are different. Figure 2 - a shows the image pair before registration while Figure 2 -b shows the composite image after application of the proposed approach. In the registered image (Figure2 - b) photoreceptors from x_{ref} fit to photoreceptors in x_{mov} and the colored parts denoting the inter-image differences corresponds

to intensity variability between images. The overlaid is averaged and shown in Figure 3 - a. The example of 10 image montage is shown in 3 - b.

(a)　　　　　　　　　　　　　　　　　(b)
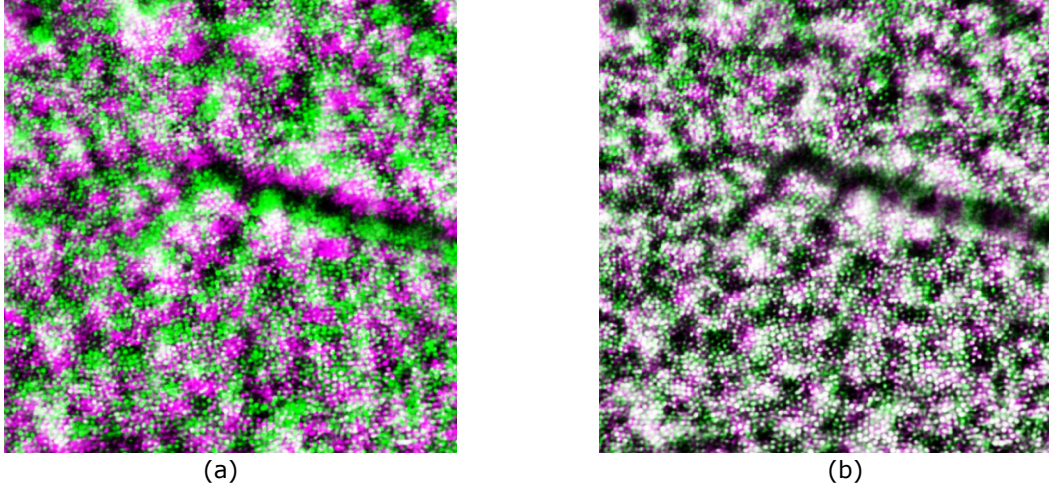
Figure 2:　(a) Overlap of x_{ref} and x_{mov} on preliminary defined position without registration. (b) overlap of x_{ref} and x_{mov} after registration. Magneta and green regions denotes the different images in reference and moving image, respectively.

Chen et al. measured the registration accuracy via NMI. The measure on AOSLO data of normal healthy patients was in the range 0.062-0.075. Our computed NMI, evaluated over all overlapping regions, is 0.08 ± 0.02, which is in comparison with results in [10] exceedingly sufficient.

(a)　　　　　　　　　　　　　　　　　(b)
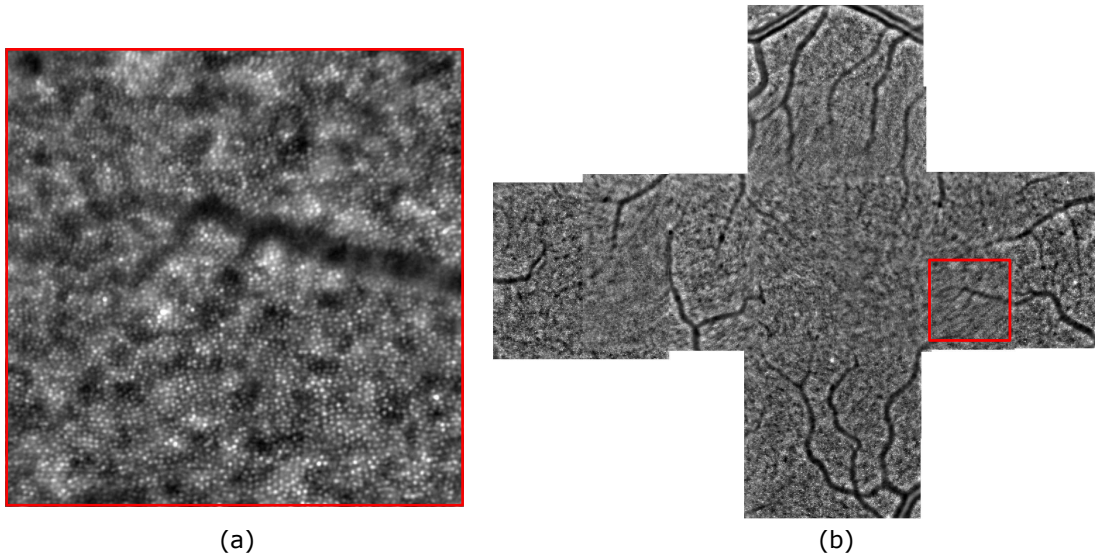
Figure 3: (a) Averaged image from x_{ref} and x_{mov} marked in (b) resultant montage of 10 images from right eye.

4　CONCLUSION

An approach for fully automated AO image montaging is presented. It utilizes SIFT method applied on pre-processed images. This method is fast, but typically provides a lot of incorrect matches, which is its main drawback. Therefore, we proposed an efficient approach solving this problem by incorrectly matches elimination based on feature Euclidean distance highest occurrence. After the

incorrectly matched points elimination, the transformation matrix for each image is obtained and alignment is performed. From 200 images of ten patients were created 20 montage images. The efficiency over all overlapping regions is measured by NMI metric and is 0.08 ± 0.02.

REFERENCES

[1] J. S. Gill, M. Moosajee, and A. M. Dubis, "Cellular imaging of inherited retinal diseases using adaptive optics," *Eye*, vol. 33, no. 11, pp. 1683–1698, 2019.

[2] D. R. Williams, "Imaging single cells in the living retina," *Vision Research*, vol. 51, no. 13, pp. 1379–1396, 2011.

[3] M. Prasse, F. G. Rauscher, P. Wiedemann, A. Reichenbach, and M. Francke, "Optical properties of retinal tissue and the potential of adaptive optics to visualize retinal ganglion cells in vivo," *Cell and Tissue Research*, vol. 353, no. 2, pp. 269–278, 2013.

[4] M. Georgiou, A. Kalitzeos, E. J. Patterson, A. Dubra, J. Carroll, and M. Michaelides, "Adaptive optics imaging of inherited retinal diseases," *British Journal of Ophthalmology*, vol. 102, no. 8, pp. 1028–1035, 2018.

[5] M. Paques, S. Meimon, F. Rossant, D. Rosenbaum, S. Mrejen, F. Sennlaub, and K. Grieve, "Adaptive optics ophthalmoscopy," *Progress in Retinal and Eye Research*, vol. 66, no. 1, pp. 1–16, 2018.

[6] J. Carroll, M. Neitz, H. Hofer, J. Neitz, and D. R. Williams, "Functional photoreceptor loss revealed with adaptive optics," *Proceedings of the National Academy of Sciences*, vol. 101, no. 22, pp. 8461–8466, 2004.

[7] A. L. Chew, D. M. Sampson, I. Kashani, and F. K. Chen, "Agreement in cone density derived from gaze-directed single images versus wide-field montage using adaptive optics flood illumination ophthalmoscopy," *Translational Vision Science & Technology*, vol. 6, no. 6, pp. 1–13, 2017-11-01.

[8] B. Xue, S. S. Choi, N. Doble, and J. S. Werner, "Photoreceptor counting and montaging of en-face retinal images from an adaptive optics fundus camera," *Journal of the Optical Society of America A*, vol. 24, no. 5, pp. 1364–1372, 2007.

[9] H. Li, J. Lu, G. Shi, and Y. Zhang, "Automatic montage of retinal images in adaptive optics confocal scanning laser ophthalmoscope," *Optical Engineering*, vol. 51, no. 5, pp. 1–6, 2012-5-1.

[10] M. Chen, R. F. Cooper, G. K. Han, J. Gee, D. H. Brainard, and J. I. W. Morgan, "Multi-modal automatic montaging of adaptive optics retinal images," *Biomedical Optics Express*, vol. 7, no. 12, pp. 4899–4918, 2016.

[11] M. Chen, R. F. Cooper, J. C. Gee, D. H. Brainard, and J. I. W. Morgan, "Automatic longitudinal montaging of adaptive optics retinal images using constellation matching," *Biomedical Optics Express*, vol. 10, no. 12, pp. 6476–6496, 2019.

[12] D. G. Lowe, "Object recognition from local scale-invariant features," in *Proceedings of the Seventh IEEE International Conference on Computer Vision*, vol. 2, 1999, pp. 1150–1157 vol.2.

[13] A. Strehl and J. Ghosh, "Cluster ensembles—a knowledge reuse framework for combining multiple partitions," *Journal of machine learning research*, vol. 3, no. Dec, pp. 583–617, 2002.

HARDWARE-ACCELERATED CRYPTOGRAPHY FOR SOFTWARE-DEFINED NETWORKS

Peter Cíbik

Doctoral Degree Programme (1st), FEEC BUT

E-mail: xcibik00@vutbr.cz

Supervised by: Lukáš Malina and Jan Hajný (BUT), Jakub Hrabovský (Netcope Technologies a.s.)

E-mail: malina@feec.vutbr.cz, hajny@feec.vutbr.cz, hrabovsky@netcope.com

Abstract: This paper presents a Software-Defined Network (SDN) cryptographic solution targeted on high-speed smart Network Interface Cards (NICs) with an FPGA chip. This solution provides a fast alternative method to develop network-oriented data processing cryptography applications for an accelerator. A high-level programming language – Programming Protocol-independent Packet Processor (P4) – is used to avoid a complex and time-consuming hardware development. The solution consists of two main parts: a library of mainly used cryptographic primitives written in VHSIC Hardware Description Language (VHDL) i.e. a symmetric cipher (AES-GCM-256), a hash function (SHA-3), a SHA-3-based Hash-based Message Authentication Code (HMAC), a digital signature scheme (EdDSA) and a post-quantum digital signature scheme (Dilithium), and a compiler P4/VHDL with the support for these cryptographic components as external objects of $P4_{16}$.

Keywords: Cryptography, Hardware acceleration, FPGA, VHDL, Software-Defined Networks, P4

1 INTRODUCTION

Nowadays trends of a centralized architecture in Information Communications Technology (ICT) solutions deploy central servers that receive messages from many end nodes. In order to process a huge amount of transactions and still keep high quality of services, many functions require a speed-up of their execution in the form of their offloading into hardware accelerators. The cryptography primitives, addressed in this paper, are a good example of such functions. The main problem of design and development of accelerator applications is their cost, primarily caused by the complexity and time and knowledge demand of the whole process.

This paper presents a solution for fast development of hardware-accelerated cryptographic applications. The fast development is reached by usage of a high-level P4 language for description of a packet-processing cryptographic application in the intelligible form. The P4 code is automatically compiled to a configuration file (*bitstream*) via the Netcope P4/VHDL compiler and directly loadable into an FPGA-based accelerator.

The solution consists of a standalone library of VHDL implementations of the basic cryptographic components, such as AES-GCM-256, SHA-3, HMAC, EdDSA and Dilithium, and an extended compiler P4/VHDL with the support for these cryptographic components as external objects of $P4_{16}$.

2 TECHNOLOGIES

- **VHDL** - Very High Speed Integrated Circuits (VHSIC) Hardware Description Language, the most popular Hardware Description Language (HDL) in Europe, is primarily used for an FPGA programming and simulation [1],

- **FPGA** - Field Programable Gate Array, shortly FPGA, is a programmable logic circuit, which

consists of three main elements: Input and Output Block (IOB), Configurable Logic Block (CLB) and programable horizontal and vertical interconection. Each of these elements is fully reconfigurable (programable) to achieve an expected functionality. High performance of FPGA chips is reached by the usage of Look-Up Tables (LUT) [1].

- **P4** - a high-level language designed for Programming Protocol-independent Packet Processors is mainly used in SDN's. Its key features are: reconfigurability, protocol independence, target independence and in the newest version (P4$_{16}$) external objects. The external objects in P4$_{16}$ bring support for target dependent features, in this solution the cryptographic external objects [2, 3].

3 CHOSEN PLATFORM FOR THE ACCELERATOR

In the presented solution, a high-speed smart NIC – NFB200G2QL, produced by *Netcope Technologies a.s.*, is used in the role of an accelerator. The smart NIC contains a Xilinx FPGA chip providing the transmition speed up to 200 Gb/s. The targeted FPGA frequency is 200 MHz. Primary parameters of the card are summarized in the Table 1.

Table 1: The parameters of NFB-200G2QL card [4]

Network interface	2×QSFP28
PCI Express	Gen 3 x16 + x16 (128 + 128 Gb/s)
FPGA chip	Xilinx Virtex UltraScale+ (xcvu7p-flvb2104-2-i)
On-board memory	3×72Mb QDRIIIe SRAM / 3×288Mb QDRIIIe SRAM

4 CRYPTOGRAPHIC COMPONENTS

Our cryptographic components library consists of five specific implementations. The Table 2 summarizes the resource utilization of each of the components.

HASH FUNCTION SHA-3

The Secure Hash Algorithm 3 (SHA-3) is based on the Keccak algorithm that uses a sponge construction. Our SHA-3 component implementation consists of two subcomponents: the **Padding Block** (the input message alignment) and the **Hash Block** (the Keccak function core). The input (a message) and output data width (a hash) are 512 bits [5, 7].

HMAC BASED ON SHA-3

The Hash-based Message Authentication Code (HMAC) component is implemented as a sequence of two separate SHA-3 instances, described in the previous section. Our HMAC component is designed according to the standard FIPS198-1, has the same input (a message) and output (a HMAC stamp) width – 512 bits. The input key (*KEY_0*) length is 576 bits [6]. The HMAC stamp is calculated like:
`HMAC(K, message) = H((K0 XOR opad)∥H((K0 XOR ipad)∥message).`

DIGITAL SIGNATURE EdDSA

The digital signature algorithm EdDSA is a particular version of the ECDSA that considers curves in a twisted Edwards form. This change positively affects the performance, has smaller key requirements and the better resiliency to side-channel attacks, specifically *Ed25519*, which uses SHA-512 (SHA-2) and *Curve25519*. The implementation incorporates all main EdDSA phases: a public key generation, a signature creation and a signature verification [7].

SYMMETRIC CIPHER AES-GCM-256

The Advanced Encryption Standard (AES) in this implementation has a key length of 256 bits for a higher security level. It operates in the Galoise-Counter Mode (GCM), which provides authenticated encryption. A chosen symmetric cipher, key length and cipher mode are in line with current NIST recommendations. Our implementation of the AES-GCM-256 consists of these components: **Expansion** (used for a key expansion for per-round keys generation), **Encryption** (provides a message encryption), **Decryption** (provides a message decryption) and **GCM** (Galoise-Counter Mode operation mode wrapper) [7].

POST-QUANTUM DIGITAL SIGNATURE DILITHIUM

The Dilithium is a post-quantum (quantum-resistant) lattice-based digital signature algorithm and is one of a few candidates for the NIST Standard. The implementation consists of all underlying functions, such as SHAKE-128, SHAKE-256, $ExpandA_q$, $ExpandMask_q$, $PowerToRound_q$, $MakeHint_q$ and Number-Theoretic Transform (NTT) functions. All of them are integrated into the main Dilithium algorithms: key generation, signing and sign verification [8].

Table 2: The hardware resources utilization of the implemented cryptographic components after an RTL synthesis

Component	LUTs	Flip-Flops	Max. Frequency [MHz]
SHA-3			
SHA-3	19 341	4 829	237
HMAC			
HMAC	17 892	11 499	218
EdDSA [7]			
Public key generation	25 830	12 317	307
Signature	31 024	16 706	307
Verification	45 420	24 762	207
AES-GCM-256 [7]			
AES-256	28 389	2 596	222
AES-GCM-256	37 362	3 834	205
Dilithium [8]			
Key Generation	54 183	25 236	350
Signing	68 461	86 295	333
Verification	50 460	28 354	225

5   THE NETCOPE P4$_{16}$/VHDL COMPILER

The Netcope P4$_{16}$/VHDL (NP4) compiler is a central part of the proposed cryptography acceleration. The NP4 compiler translates a P4 source code to a VHDL code and maps it on an FPGA platform, as shown in the Figure 1. Integration of the support for cryptographic external objects to the NP4 compiler consist of several steps: the NP4 pipeline extension, a generation of templates for various wrappers and support components, and an instantiation of the components based on the content of source *.p4* files. The steps are described in detail in the following subsections.
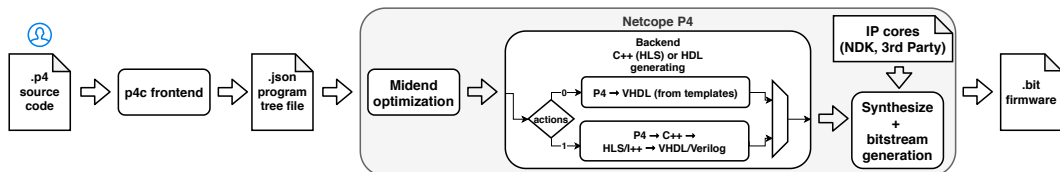
Figure 1: Compilation process

THE NP4 PIPELINE EXTENDED WITH C2 CONTROL BLOCK

The extended NP4 pipeline consists of the Parser (parses packet headers and a payload to separate fields), the Match+Action (implements all the tables and performs the actions, specified in the source code), the **C2 control block** (instantiates cryptographic external objects) and the Deparser (joins all modified headers and a payload back to a resulting packet), as shown in the Figure 2.
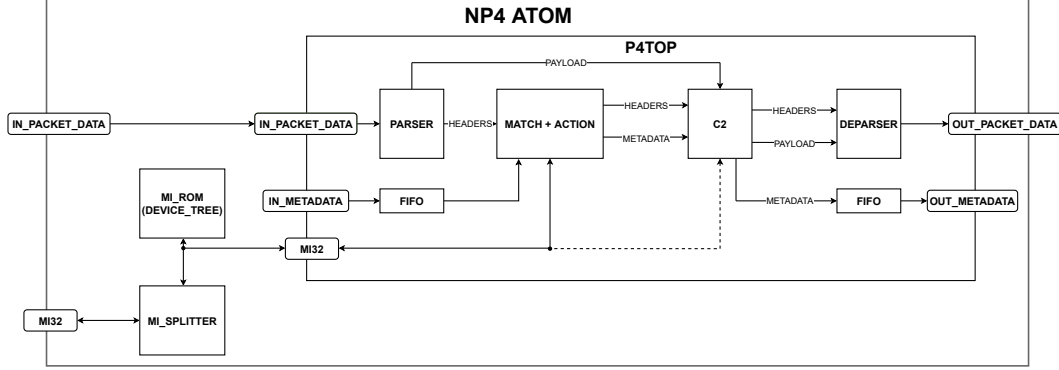
Figure 2: The architecture of the extended NP4 pipeline

WRAPPERS FOR CRYPTOGRAPHIC COMPONENTS

All cryptographic components, integrated into the NP4 pipeline, have their own wrappers to ensure the compatibility with the input and output interface of the pipeline. A general architecture of the wrapper is shown in the Figure 3a. A wrapper integration to the C2 control block is shown in the Figure 3b also with a FLU_APPENDER component, which appends computed results to the end of the packet payload.
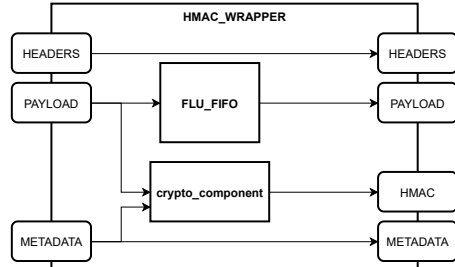
(a) The architecture of wrapper components
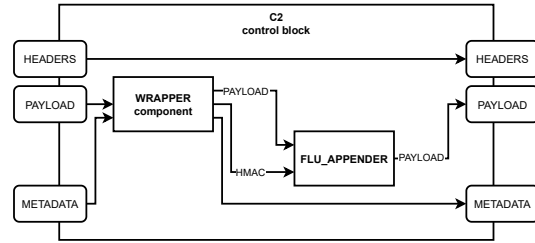
(b) The architecture of the C2 control block with a wrapper component

Figure 3: Architecture schemes

THE NP4 COMPILER EXTENSION

The NP4 Compiler was extended to support cryptographic external objects with these functionalities: processing of a *.json* tree file extracted from a *.p4* source code, generation of wrappers for cryptographic external objects, instantiation of external objects to the C2 control block (the architecture generation) and other supporting functions.

REFERENCE USAGE OF CRYPTOGRAPHIC EXTERNAL OBJECTS IN A .P4 FILE

The NP4 provides a source code `externs.p4` with declaration of currently supported external objects shown in Listing 1.

Listing 1: The declaration of cryptographic external objects `externs.p4`

```
extern HMACOverPayload {
    HMACOverPayload();
    void compute();
}

extern HashOverPayload {
    HashOverPayload();
    void compute();
}
```

6 CONCLUSION

This work describes a hardware-accelerated cryptographic solution based on a smart NIC with an FPGA chip. It employs a high-level P4 language for a fast and flexible development of accelerated cryptographic applications. The VHDL component library contains implementations of AES-GCM-256, SHA-3, HMAC, EdDSA and Dilithium. The NP4 compiler was extended by a set of functions to generate supported cryptographic components and their wrappers, and to place them inside the new C2 control block in the pipeline based on the processed input *.p4* files. Two of these components are already implemented in the NP4 compiler as the cryptographic external objects. In future work, the NP4 compiler will be extended by the support for more cryptographic external objects.

ACKNOWLEDGEMENT

REFERENCES

[1] PINKER, Jiří a Martin POUPA. *Číslicové systémy a jazyk VHDL*. Praha: BEN - technická literatura, 2006. ISBN 80-7300-198-5.

[2] BOSSHART, P.; DALY, D.; GIBB, G.; aj.: *P4: Programming Protocol-Independent Packet Processors. ACM SIGCOMM Computer Communication Review*, year 3, Num. 44, 2014: s. 87–95, ISSN 0146-4833

[3] *P4 16 Language Specification*. Version 1.2.0-rc. The P4 Language Consortium, 2019. Available from: https://p4.org/p4-spec/docs/P4-16-v1.2.0.pdf

[4] Netcope products: *Netcope FPGA boards (NFB)* [online]. [cit. 2019-11-11]. Available from: https://www.netcope.com/en/products/fpga-boards

[5] Bertoni, G., Daemen, J., Peeters, M., Van Assche, G.: *The keccak sha-3 submission. Submission to NIST (Round 3)* 6 (7), 16 (2011)

[6] D. H. Krawczyk, M. Bellare a R. Canetti, *HMAC: Keyed-Hashing for Message Authenti-cation, RFC 2104, feb. 1997*. doi: 10.17487/RFC2104. url: https://rfc-editor.org/rfc/rfc2104.txt

[7] MALINA, Lukáš, David SMÉKAL, Sara RICCI, Peter CÍBIK and Jakub HRABOVSKÝ. *Hardware-Accelerated Cryptography for Software-Defined Networks with P4*. SECITC, 2020.

[8] RICCI, Sara, Lukáš MALINA, Petr JEDLIČKA, Jan HAJNÝ, Peter CÍBIK and Patrik DOBIÁŠ. *Implementing CRYSTALS-Dilithium Signature Scheme on FPGAs*

PHISHING DETECTION USING DEEP LEARNING ATTENTION TECHNIQUES

Yehor Safonov

Doctoral Degree Programme (1), FEEC BUT

E-mail: xsafon00@feec.vutbr.cz

Supervised by: Zdenek Martinasek

E-mail: martinasek@feec.vutbr.cz

Abstract: In the modern world, electronic communication is defined as the most used technology for exchanging messages between users. The growing popularity of emails brings about considerable security risks and transforms them into an universal tool for spreading phishing content. Even though traditional techniques achieve high accuracy during spam filtering, they do not often catch up to the rapid growth and evolution of spam techniques. These approaches are affected by overfitting issues, may converge into a poor local minimum, are inefficient in high-dimensional data processing and have long-term maintainability problems. The main contribution of this paper is to develop and train advanced deep networks which use attention mechanisms for efficient phishing filtering and text understanding. Key aspects of the study lie in a detailed comparison of attention based machine learning methods, their specifics and accuracy during the application to the phishing problem. From a practical point of view, the paper is focused on email data corpus preprocessing. Deep learning attention based models, for instance the BERT and the XLNet, have been successfully implemented and compared using statistical metrics. Obtained results show indisputable advantages of deep attention techniques compared to the common approaches.

Keywords: artificial intelligence, attention mechanism, deep learning, NLP, phishing filtering, text classification, transformers.

1 INTRODUCTION

Nowadays, email is a highly effective method of electronic communication that allows people to exchange information in the online world [1]. A significant advantage of email services is their free availability, efficiency and intuitiveness [2]. All these aspects constitute a significant advantage in the provision of communication services. For a better understanding of the technology's importance, it is necessary to mention that at the end of this year, the count of email messages has exceeded 293 billion and the total number of email users worldwide has exceeded 4,3 billion, which is approximately 55,8 % of Earth's population [3].

Popularity of electronic communication comes with certain security risks, for example, spreading of unwanted emails (*spamming*). The core idea of spamming techniques is based on human imperfection during their decisions-making [3]. The primary motivation of spam is to hide the attacker's intent to obtain sensitive data from the victim (*phishing*), to provide unwanted services and goods, to spread malicious code, etc. Other problems which arise besides those already metioned are unauthorized message reading, content modification, identity theft, etc. According to statistical analysis, the daily count of unsolicited email messages is nearly 165 billion, which corresponds to 56,5 % of the total number of emails [3]. It is quite clear that the popularity of email combined with its mediocre default security enforcement make it an ideal tool for realization of efficient attacks. It is also predicted that the number of phishing attacks is going to grow annually, and therefore issues related to ensuring security and filtering of untrusted emails cannot be underestimated or put aside. [2]

The world of mechanisms for detection and filtering of spam as well as the world of new spam techniques evolves daily. There are a lot of different security techniques aimed to detect unsolicited content [3]. Traditional ones are based on email's reputation, for example, whitelisting, SMTP flow analysis and CR systems [3]. Sometimes these techniques include different statistical methods and provide simple content analysis, e.g., spam word detection, fingerprint based methods, decision trees, SVM, and Naïve Bayes. It is possible to achieve up to 90 % accuracy by using mentioned methods [3]. On the other hand, many of them cannot keep up with the rapid evolution of spam techniques and have long-term maintainability issues [1]. Furthermore, traditional spam detectors do not even try to interpret the context of emails and are often based on trigger detection and application of predefined rules [2]. Most advanced filtering techniques implement deep networks which analyze the lexical context of incoming emails. The main advantage of applying deep learning approaches is that the filtering system is not based on static rules [4]. The system behaves dynamically and is able to successfully perform classification on previously unseen data. As a result, the usage of deep learning techniques requires less time investment for maintenance while preserving a high level of accuracy [4].

2   METHODOLOGY

To apply supervised learning techniques with the motivation to create an efficient deep attention based model for phishing detection, it is important to create a dataset called an *initial corpus* [4]. Final precision of neural networks is proportional to the quality of the dataset which was used during their training process [4]. The higher the quality of the data, the higher the probability of the model. After processing, the initial corpus is divided into separate training, validation and test data frames. Finally, the prepared frames are used for training and testing of selected deep learning networks.

2.1   INITIAL CORPUS PROCESSING

Dataset processing, in other words *corpus preprocessing*, is the first integral part, which should be done before the application of any machine learning algorithms [4]. In the case of this study, the primary motivation is set to create a state-of-the-art spam email dataset containing the latest phishing emails and reflecting the current state of existing obfuscation techniques. For this purpose, two different datasets have been chosen. The first one is called the Enron Email Dataset. It contains **517 401** raw legitimate emails (*ham*). It is an open source set of emails published in 2015. The second one is a TrustPort Spam dataset. This dataset has been provided by the TrustPort company and contains more than **720 000** raw spam emails collected between the years 2015 and 2020. Both corpuses should be processed before the application of attention mechanisms. The data preprocessing process has been divided into the following consecutive steps:

1. **Initial data preprocessing** – the primary motivation was to modify the email corpus into a suitable form allowing effective processing in the future. The main task was to create a set of scripts that are able to load a byte content of emails, determine the character set, decode both the header and the body of the message, parse them and then convert to a JSON object.
2. **Data cleaning** – this process involves removing inconsistent and poor quality data from the dataset. The primary motivation is to improve the quality of the training data and as a result to precise future extraction. Both obfuscation techniques applied by the attackers and the inconsistent emails were removed. Finally, styled text was converted to plain text.
3. **Data selection** – during this stage only relevant emails were selected. In the case of this study, only plaintext data that met the language requirements were relevant. In order to select the most appropriate data, two datasets were analyzed from a lexical point of view. English was chosen as the primary language due to the fact that most of the spam emails was written in it (around 76 %). Also during this stage all multimedia and binary data were excluded. The HTML and plaintext parts of each email were compared and interconnected together.

4. **Data integration** – after the processing of the two datasets, it was necessary to unify their formats and convert them into an appropriate form which allows reading, writing and modifications. Consolidate form of processed data helps to unify functions to access them and, as a result, it allows to work with the data efficiently. The CSV (*Comma-Separated Values*) format was chosen for this purpose. In this context, it should be noted that for efficient parallel processing of input data, it is good practice to divide the dataset into smaller units, i.e. *batches*.

5. **Data transformation** – the data transformation is the final step in the dataset processing. It depends on the targeted task and specifics of the network's input layer. In this case, the transformation process is affected by the embedding techniques. Spam filtering is a classification problem [1]. It is more efficient to apply supervised learning mechanisms to get a better result. To make such a task happen, it is necessary to add specific tags to each email presented in the CSV. For binary classification it is enough to add two different labels ("0" for trusted emails and "1" for spam emails) [4]. In the end, all emails were randomized and put together.

After the completion of the aforementioned steps, the total number of electronic messages appropriate for machine learning was **1 070 186**. The number of spam messages was 553 229 and the number of ham messages was 516 957. Required time for processing of the initial corpus was 32 hours. Average time for processing each email was 0.108 seconds. The final dataset was divided into smaller parts containing only 25 000 entities. As a result, there were 43 CSV documents covering the full range of emails. The structure of a CSV fragment of the dataset is shown in the Figure 1.

```
ID              Subject         Body                    Urls
"13207054898446";"Re: Hi Dear ... ";"Hi Dear, We produce hot  ... ";"http://srv ... "
"13207055055763";"Russian girls ... ";"Visit Bride.ru View photos ... ";""
"13207055058899";"Biggest DIRECT ... ";"Discount -40% Customer ... ";"http://cows ... "
"13207055114514";"Salutations ... ";"I was mistreated in the ... ";"http://corners ... "
      ...             ...             ...                     ...
"13207056012414";"Re: Welcome  ... ";"Hi Dear, We produce hot  ... ";"http://srv ... "
"13207055055432";"Sexy Girls ... ";"Hi, see these amazing photos ... ";"http://se ... "
"13207051211399";"Re: Plastic m ... ";"Discount -80% Go Go ... ";"https://cars/ ... "
"13201211114514";"My dear Roman ... ";"I was waiting for you ... ";"http://horny/ ... "
```

Figure 1: The structure of the generated CSV document of spam emails

2.2 APPLYING DEEP ATTENTION TECHNIQUES

It is important to realize that understanding of language is not a trivial task. It is not possible to encode a text sequence linearly by using numbers that can be directly connected to the input layers of a network. The problem of understanding and encoding of lexical meaning is called NLP (*Natural Language Processing*). A major progress in this area occurred with the RNN (*Recurrent Neural Network*). This approach allowed analysis of long text sequences. Today, there are a large number of derived architectures that bring new features to the classic RNN, for instance LSTM (*Long-Short Term Memory*) networks, GRU (*Gated Recurrent Unit*) networks, etc. [4]

Deep attention techniques, also known as transformers, are the most powerful and advanced concept nowadays that has completely revolutionized the filed of machine learning [5]. Attention mechanisms with the concept of transfer learning achieve better results in the NLP field compared to the traditional approaches. These deep networks were first used in the NLP field in the context of using Seq2Seq technique (in machine translation) and nowadays are also being used in speech recognition and image processing [5]. Compared with the classic LSTM networks, deep transformers allow for processing of a sequential data input not in a sequence order. It enables the network to process the input data faster using parallel processing [6]. Classic transformer architectures are based on Encoder-Decoder models that are based on the principles of cognitive attention [5]. In a nutshell, it means that only

important parts of the input data are enhanced, while others are suppressed. Nowadays, the most powerful deep attention networks are BERT, XLNet, RoBERTa, ALBERT, GPT-2, GPT-3 etc. In this study, two types of transformer artificial networks were tested, BERT [6] and XLNet [7].

The BERT model is a network based on a *bidirectional transformer*. The core of the network consists of 12 layers and contains 110 million optimization parameters. The BERT provides a generalized AELM (*Autoencoder Language Model*). Word tokenization is performed by using the WordPiece technique. The maximum length of the block is limited to 512 tokens, and each tokenized word is represented by a vector in 768 dimensions. Word embedding is achieved by using the TWE technique (*Transferring Word Embeddings*). For a better detection of semantic dependencies, the BERT uses its MLM (*Masked Language Modeling*) and the NSP (*Next Sentence Prediction*) techniques. Training of the generalized model was performed on a large dataset and took approximately 450 GPU days. [6]

The XLNet improves the architecture of transformers using permutation techniques. The core consists of 12 encoder layers and the total number of parameters is 110 million. The XLNet provides a generalized GLMP language model, more specifically the ARLM (*Autoregressive language model*). Tokenization of the input text is achieved using the SentencePiece mechanism. The dimensionality and the maximum block size are identical to the corresponding parameters of the BERT model. In addition, the artificial network implements the PLM (*Permutation Language Modeling*) and the TSSA (*Two-Stream Self-Attention*) techniques to achieve better results. This model used a larger learning dataset than BERT and the training process took 2 000 GPU days. [7]

3 RESULTS AND DISCUSSION

Both models were trained on the selected part of the processed data. All emails were chosen randomly with the motivation to cover the considerable scope of obfuscation techniques. In total, there were three dataframes. The first dataframe was used for the training process and includes 30 000 emails. The second one helped to validate the training process computing validation loss. This dataframe contained 10 000 emails. The third one was not used during the training process, but it was selected for the real testing of both networks. Its size was the same as of the validation dataframe. The final results of the BERT and XLNet deep transformers networks are shown in the Figure 2.

Deep neural network	Learning time	Precision
BERT	78 min	**99,46 %**
XLNet	155 min	**99,50 %**
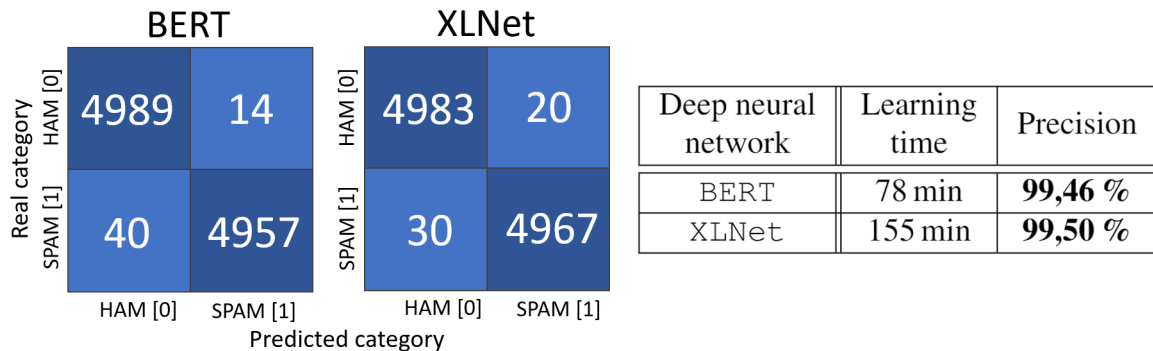
Figure 2: Final precision of BERT and XLNet neural networks

Additionally, two *confusion matrices* were computed for a more precise comparison. They help to identify all possible results of the classification process. The total number of correct answers for the BERT was 9 946, while the number of incorrect predictions was 54. Going deeper into the analysis of correct answers, there were 4 989 legitimate emails (*True Negative*) and 4 957 spam messages were classified as TP (*True Positive*). The set of false predictions contained 14 FP (*False Positive*) messages. The set of undetected spams included 40 entities FN (*False Negative*). The learning time was 78 minutes and the final precision was **99,46 %**. The total amount of XLNet's incorrect

classifications was 50. The number of correct predictions was 9 950. The number of TP, i.e. correctly evaluated spam, was equal to 4 967. Wrong decisions were divided as follows: 20 cases belonged to the FP and 30 to the FN. The learning time was 155 minutes and the precision was **99,50 %**.

4 CONCLUSION

The main motivation of this paper was to apply the most advanced deep learning techniques. For instance, the XLNet and BERT were used to understand the context of incoming emails and provide efficient and precise filtering. The key differentiation of this study, among others, is the usage of the most prominent techniques in the machine learning area, which comes with transformers, multi-head attention mechanisms and positional encoding [6, 7]. Discussed technologies revolutionized the deep learning area and allowed processing of the input text in a parallel manner avoiding recursion [5]. Another strength of the proposed models is their precision. It is achieved by the usage of a large data corpus with the combination of transfer learning techniques. Both models show better results than traditional approaches, such as the SVM, RNN, LSTM and Naïve Bayes, see papers [1, 3]. The final accuracy of the learned BERT was equal to 99,66 % and the XLNet's final accuracy was 99,62 %. In addition to the created models, the output of the study is a set of automated scripts that allow end to end raw email processing, including email preprocessing and cleaning for a more precise evaluation.

ACKNOWLEDGEMENT

REFERENCES

[1] BHOWMICK, Alexy and Shyamanta M. HAZARIKA. 2018. E-Mail Spam Filtering: A Review of Techniques and Trends: A Review of Techniques and Trends. *Advances in Electronics, Communication and Computing* [online]. Singapore: Springer Link, vol. 443, no. 1, pp. 583-590.

[2] AVOINE, Gildas, Pascal JUNOD a Philippe OECHSLIN. *Computer System Security: Basic Concepts and Solved Exercises*. EPFL Press, 2007. ISBN 9781420046205.

[3] DADA, Emmanuel Gbenga, Joseph Stephen BASSI, Haruna CHIROMA, Muhammad Abdulhamid SHAFII, Olusola Adetunmbi ADEBAYO and Emmanuel Ajibuwa OPEYEMI. Machine learning for email spamfiltering: review, approaches and openresearch problems. *Heliyon* [online]. ScienceDirect, vol. 5, no. 6

[4] CHOLLET, Francois. *Deep learning with Python*. Shelter Island, New York: Manning Publications Co., [2018]. ISBN 1617294438.

[5] VASWANI, Ashish, Noam SHAZEER, Niki PARMAR, Jakob USZKOREIT, Llion JONES, Aidan N. GOMEZ, Lukasz KAISER a Illia POLOSUKHIN. *Attention Is All You Need* [online]. [cit. 3. 3. 2021]. Available at: https://arxiv.org/pdf/1706.03762.pdf

[6] DEVLIN, Jacob, Ming-Wei CHANG, Kenton LEE a Kristina TOUTANOVA. *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding* [online]. Google AI Language, 16 str. [cit. 3. 3. 2021]. Available at: https://arxiv.org/abs/1810.04805

[7] YANG, Zhilin, Zihang DAI, Yiming YANG, Jaime CARBONELL, Ruslan SALAKHUTDINOV a Quoc V. LE. *XLNet: Generalized Autoregressive Pretraining for Language Understanding* [online]. Carnegie Mellon University: Google AI Brain Team, 2020 [cit. 3. 3. 2021]. Available at: https://arxiv.org/pdf/1906.08237.pdf

VHDL-BASED IMPLEMENTATION OF NTT ON FPGA

Petr Jedlička

Doctoral Degree Programme (1), FEEC BUT

E-mail: xjedli23@stud.feec.vutbr.cz

Supervised by: Jan Hajný

E-mail: hajny@feec.vutbr.cz

Abstract: This paper is focused on the effective hardware-accelerated implementation of NTT (*Number Theoretic Transform*) and inverse NTT (NTT^{-1}) on FPGA (*Field Programmable Gate Array*). The discussed implementation is intended for the use in the lattice-based cryptography schemes, e.g. CRYSTALS-Dilithium digital signature scheme which is one of the finalists of the third round in the post-quantum standardization process under the auspices of NIST (*The National Institute of Standards and Technology*). The implementation of NTT (NTT^{-1}) requires 1798 (2547) Look-Up Tables (LUTs), 2532 (3889) Flip-Flops (FFs) and 48 (84) Digital Signal Processing blocks (DSPs). The latency of the design is 502 (517) clock cycles at the frequency 637 MHz on Xilinx Virtex UltraScale+ architecture which makes the presented implementation to be currently the fastest one. Regarding the inverse NTT, this is the first implementation at all.

Keywords: NTT, VHDL, FPGA, Dilithium, Montgomery reduction

1 INTRODUCTION

In recent years, there has been a significant progress in the field of quantum computing research. Using special algorithms, quantum computers are able to solve hard problems such as the integer factorization and the discrete logarithm much faster than conventional computers. Many of the widely used cryptosystems with the public key that are considered to be secure are based just on the two problems mentioned above and thus they will become breakable if a quantum computer comprising a sufficient number of qubits is built. As a consequence of these facts, intensive research in the field of so-called post-quantum cryptography (PQC) has started. The PQC represents cryptosystems that are not vulnerable to quantum computer attacks. According to the currently ongoing NIST post-quantum standardization process, ones of the most promising PQC cryptosystems are the lattice-based ones. The algorithms of these cryptosystems perform mathematical operations over polynomials in modular arithmetic such as additions and multiplications that can be accelerated using the NTT algorithm.

This work presents original and optimized VHDL (*Very High Speed Integrated Circuit Hardware Description Language*) implementations of NTT and inverse NTT according to the reference implementation of the CRYSTALS-Dilithium digital signature scheme (shortly Dilithium) in the C programming language [1].

2 RELATED WORK

To the best of the author's knowledge, there are currently only two hardware implementations of the NTT that are targeted to Dilithium. However, neither of them mention the implementation of the inverse NTT. Nguyen *et al.* [2] compare the efficiency of the HDL (*Hardware Description Language*) and HLS (*High Level Synthesis*) implementations on the UltraScale+ architecture. Their implementation does not fully meet the reference implementation because they replaced the Montgomery reductions by so called K-reductions. As a consequence of this modification, it is not possible to verify the correctness of such design by simple comparison with the reference NTT implementation whose

output is in the Montgomery domain. Unfortunately, they do not mention the way how they verified their design. In the second article, Nejatollahi *et al.* [3] present their NTT implementation on the Zynq UltraScale+ architecture using the HLS method which is usually much less effective in terms of speed and pipelined processing.

3 POLYNOMIAL MULTIPLICATION AND NUMBER THEORETIC TRANSFORM

Dilithium performs operations in the polynomial ring $\mathcal{R}_q = Z[x]_q/(x^n + 1)$ where n and q are two integers, i.e. $n = 256$ and $q = 8380417 = 2^{23} - 2^{13} + 1$. Using the textbook definition for the polynomial multiplication, the achieved time complexity is $O(n^2)$. However, the polynomial multiplication accelerated by the properties of the NTT reduces the time complexity to $O(n \log n)$. The multiplication of polynomials $f(x)$ and $g(x)$) using the NTT can be expressed by the following formula [5]:

$$f(x) \times g(x) = NTT^{-1}(NTT(f(x)) \odot NTT(g(x))),$$

where $\odot$ is the point-wise multiplication. The NTT can be used only for the $q = 1 \bmod(2n)$ [5].

3.1 NTT

The NTT is a generalization of the discrete Fourier transform, which is carried out in a finite field instead of complex numbers. Assuming a polynomial $f(x)$ with coefficients $(f_0, f_1, ..., f_{n-1})$ and their representation in the NTT domain $(\hat{f}_0, \hat{f}_1, ..., \hat{f}_{n-1})$, the transformations can be expressed by the following formulas [4]:

$$\hat{f}_i = \sum_{j=0}^{n-1} f_i r^{ij} \bmod(q), i = 0, 1, ..., n-1,$$

$$f_i = \frac{1}{n} \sum_{j=0}^{n-1} \hat{f}_i r^{-ij} \bmod(q), i = 0, 1, ..., n-1,$$

where $r = 1753$ is the $2n$-th primitive root of unity in modulo q arithmetic [5].

3.2 MONTGOMERY REDUCTION

Montgomery reduction is the computation of $xR^{-1} \bmod m$. The computation of $x \bmod m$ is performed in two successive Montgomery reductions and a one intermediate multiplication by $R^2 \bmod m$. The algorithm for one Montgomery reduction $u = xR^{-1} \bmod m$ is following [4]:

$$\tilde{m} = -m^{-1}$$

$$\tilde{q} = x\tilde{m} \bmod R$$

$$u = \frac{x + \tilde{q}m}{R}$$

4 VHDL IMPLEMENTATION

The design has been implemented on a Virtex UltraScale+ FPGA from Xilinx in Vivado 2017.4.1 using the VHDL language. As the selected hardware platform and the series of the mounted FPGA suggests, the main optimization goal of the design is speed rather than area and the chosen implementation strategy corresponds to this fact. To reach as high clock frequency as possible, pipelined processing that reduces the size of combinational logic between two FFs is applied to the individual functional blocks. Furthermore, DSP blocks are used for almost all arithmetical operations, only incrementations or multiplications by a constant power of two are excluded. Due to the speed optimization, DSP blocks are set for the maximum latency whereas the throughput is still one output per one clock cycle. At first, individual functional sub-blocks has been implemented. In next step,

NTT and NTT^{-1} components has been implemented whereas the sub-blocks has been integrated into those two components. All the components follow the reference implementation [1] and are described below.

4.1

The c   uction of 64-bit values
to 32-   it per one clock cycle.
The p   gure 1.
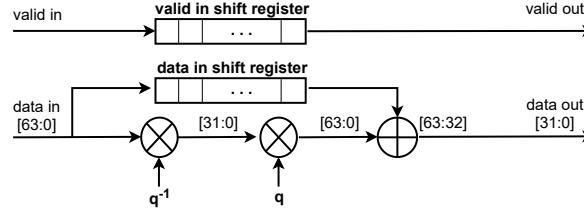
Figure 1: The block scheme of the Montgomery reduce implementation

There are two input signals, `valid_in` and `data_in`, and two corresponding output signals, `valid_out` and `data_out`. The `valid` signals indicates the valid data on the input or output data bus. This signal propagates through a shift register whose length has to be equal to the latency of the component. Whereas there is the maximum possible throughput at data bus there is no need for a backpressure signal. All the arithmetic operations run in parallel, are pipelined and use DSP blocks. There is another shift register used for the input 64-bit value and its length corresponds to the overall latency of the multipliers. Thanks to the appropriately chosen value of the parameter R, the modulo operations and divisions can be done effectively by simple dropping unnecessary bits from the buses.

4.2 BUTTERFLIES

The NTT and NTT^{-1} butterflies has been implemented by applying the same methodology as in the case of Montgomery reduce. The latency is in both cases 24 clock cycles and the throughput is one output per one clock cycle. The algorithm exactly follows the reference implementation [1]. The blo
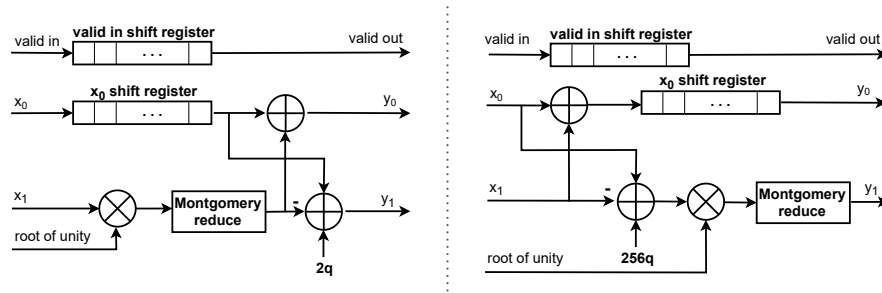
Figure 2: The block schemes of the implementation of the butterflies (left: NTT, right: NTT^{-1})

4.3 NTT AND INVERSE NTT

The described implementation is based on the FFT (*Fast Fourier Transform*) algorithm with decimation in the frequency domain as it is handled in the reference implementation [1]. The transformation takes place over 256 samples, which corresponds to the calculation performed in 8 iterations ($256 = 2^8$). Within each iteration, 128 partial transformations are calculated using butterflies. The difference in the order of the coefficients entering this partial transformation is $\frac{128}{2^{n-1}}$, where n denotes the order of the iteration with indexing starting from 1. To compromise between the speed of calculation and hardware resources, the parallelization of the calculation using 4 butterflies in a 2x2

arrangement is deployed. This means a calculation in two iterations at the same time with two butter-flies in each of them. The butterflies in the first iteration have to be shifted by a value corresponding to the required difference in the coefficients of the following iteration. The reason for spreading the calculation between two iterations is to reduce the number of output coefficients (for one iteration) and thus minimize the number of Block Random Access Memory (BRAM) for storing intermediate results too. Each block memory has only two ports which can be accessed at the same time and there-fore it is needed *N* block memories to store *2N* values in parallel regardless the memory size actually needed. Another reason of the 2x2 arrangement is to reduce the number of interludes between iter-ations. The order of reading the input intermediate results differs depending on the current iteration and therefore each iteration can start after storing all intermediate results from the previous one which is delayed due
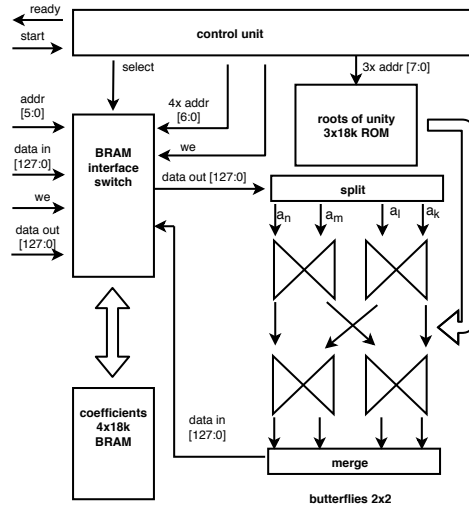
Figure 3: The block scheme of the NTT implementation

The block diagram of the NTT component is shown in Figure 3. Four BRAMs are used to store input values, intermediate results and output values. In the inactive mode, the interface of these memories is switched by the control unit to the component interface, then inputs and outputs can be written or read. During calculating the transformation, the memory interface is made available to the control unit, which sets the addresses, and to the computational structure consisting of butterflies in a 2x2 layout connected to the data buses. Four coefficients are transmitted in parallel over the data bus, which are divided between the inputs of the butterflies and merged again at the output. The last block is the ROM memory with the roots of the unit equation (so-called roots of unity). [5]

The design of the Inverse NTT (NTT^{-1}) component is almost identical to the NTT scheme. Only the iterations are performed in the reverse order, and thus the distribution of the butterflies in the first and second parallel iterations is inverse. In contrast to NTT, the Montgomery reduction of the output coefficients is performed at the end of the whole calculation. [5]

5 IMPLEMENTATION RESULTS AND COMPARISON

Table 1 depicts the implementation results for the presented design and the aforementioned previous works whereas the comparison can be made only for the NTT because the other works do not mention the implementation of the NTT^{-1}. The proposed HDL-based design needs on average 20 times less hardware resources and is 6 times faster than the implementation of Nejatollahi *et al.* [3]. In comparison to the implementations (HLS-based and HDL-based) of Nguyen *et al.* [2], their designs have comparable results of the hardware utilization but they are not so much optimized in terms of the clock frequency and the speed as the presented ones. The presented components are able to run

at the clock frequency 637 MHz while the value reached by Nguyen *et al.* [3] is only 445 MHz. It is remarkable that Nguyen *et al.* [2] reached exactly the same latency for the both implementation strategies.

Table 1: The implementation results and the comparison with the other implementations

	LUT	FF	DSP	BRAM	LUTRAM	Frequency [MHz]	Latency [cycles]
NTT							
This work	1798	2532	48	3.5	438	637	502
Nejatollahi *et al.* [3]	47332	38108	1282	2	-	199	1058
Nguyen *et al.* [2], HDL	1899	2041	8	2	-	445	294
Nguyen *et al.* [2], HLS	1977	2329	8	2	-	434	294
Inverse NTT							
This work	2547	3889	84	3.5	762	637	517

6 CONCLUSION

To the best of the author's knowledge, this work introduced the fastest and most effective hardware implementation of the NTT and the first implementation of the inverse NTT which are the essential and most complex functional blocks of the lattice-based PQC schemes whose implementation on FPGA will be the next step. Using pipelined processing, the components are able to perform over 1,200,000 transformations per second with the clock frequency 637 MHz which is the highest reached value in comparison with the previous implementations whose clock frequencies are in the range between 199 MHz and 445 MHz.

ACKNOWLEDGEMENT

REFERENCES

[1] DUCAS, L., Kiltz, E., Lepoint, T., Lyubashevsky, V., Schwabe, P., Seiler, G., Stehle, D.: Crystals-dilithium: A lattice-based digital signature scheme. IACR Transactions on Cryptographic Hardware and Embedded Systems pp. 238-268 (2018). Available from: doi:10.13154/tches.v2018.i1.238-268

[2] NGUYEN, Duc Tri, Viet B. DANG, Kris GAJ. A High-Level Synthesis Approach to the Software/Hardware Codesign of NTT-Based Post-Quantum Cryptography Algorithms. In: 2019 International Conference on Field-Programmable Technology (ICFPT) [online]. IEEE, 2019, 2019, s. 371-374 [cit. 2021-03-10]. ISBN 978-1-7281-2943-3. Available from: doi:10.1109/ICFPT47387.2019.00070

[3] NEJATOLLAHI, Hamid, Sina SHAHHOSSEINI, Rosario CAMMAROTA, Nikil DUTT. Exploring Energy Efficient Quantum-resistant Signal Processing Using Array Processors. In: ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) [online]. IEEE, 2020, 2020, s. 1539-1543 [cit. 2021-03-10]. ISBN 978-1-5090-6631-5. Available from: doi:10.1109/ICASSP40776.2020.9053653

[4] NEJATOLLAHI, Hamid, Nikil DUTT, Sandip RAY, Francesco REGAZZONI, Indranil BANERJEE, Rosario CAMMAROTA. Post-Quantum Lattice-Based Cryptography Implementations. In: ACM Computing Surveys [online]. 2019, s. 1-41 [cit. 2021-03-10]. ISSN 0360-0300. Available from: doi:10.1145/3292548

[5] RICCI, Sara, Lukas MALINA, Petr JEDLICKA, David SMEKAL, Jan HAJNY, Petr CIBIK, Patrik DOBIAS. Implementing CRYSTALS-Dilithium Signature Scheme on FPGAs [online]. 2021, [cit. 2021-03-10]. Available from: https://eprint.iacr.org/2021/108

SIMULATOR OF NETWORK COMMUNICATION WITH MAP INTERFACE FOR NS3

Jan Benedikt

Doctoral Degree Programme (2.), FEEC BUT

E-mail: xbened03@stud.feec.vutbr.cz

Supervised by: Dan Komosný

E-mail: komosny@feec.vutbr.cz

Abstract: The work deals with the description of creating a network simulator with a map interface for Network Simulator 3. The first part of the work deals with the existing solution Network Animator (NetAnim) and based on the submitted requirements, justifies the intention to create a new simulation interface. The following is a description of the technologies and standards used to create a new interface. The main part of the work is a description of the development of the simulator itself. The last part of the work presents the developed interface for analysis of simulation results from the Network Simulator 3 (NS3).

Keywords: NS3, XML, NetAnim, Python, JavaScript, Folium, Flask

1 INTRODUCTION

One of the many programs for computer network modeling is the program Network Simulator 3. The program itself has no graphical output for visualization of the simulation results. Network Animator (NetAnim) and Python Vizualization (PyViz) are programs intended for visual presentation of simulation results. Although both programs offer many functions, they do not offer the possibility to present the results on a map. Locating network elements according to geographical location can be an advantage for simulating large metropolitan networks. For this reason, a web application was created with the possibility of presenting the results on a map. The first part of this paper deals with the currently available programs NetAnim and PyViz, followed by a description of the function of the newly created application.

2 USED TOOLS

2.1 NETWORK SIMULATOR 3

NS3 is a discrete event network simulator used primarily in research and for educational purposes [1]. The first version of the Network Simulator was released in 2008. Since then, a total of three versions of this simulator have been released. The current version is the third in a row. The software is under GNU's Not Unix! General Public License version 2 (GNU GPLv2) and is free to use, mainly for research and development. The simulation modules are written in C ++ or Python. The whole simulator consists of libraries for specific network functions. Most libraries are available for C ++, the rest are also available for Python. However, using Python for simulation purposes is problematic. Most libraries are for C ++ only, and there is only a basic part of the Python version. The second disadvantage lies in the old version of the libraries, when most of them are for Python version 2.7, which has reached the end of its life in 2020 [2].

2.2 Network Animator

The Network Animator (NetAnim) program is used for graphical interpretation of simulation results generated by the NS3 program. The program graphically shows the behavior of the network and shows the communication between individual network nodes. The current version of the program is 3.108. To display the network, the program uses a grid on which it places network nodes and connects them with lines. If it is necessary to place the device on a map base, for example, for a better context within the topology, the program allows the use of images of a map. However, this is not effective in terms of a single map scale. Furthermore, the user is limited by the size of the grid where the network nodes are located [3]. For this reason, the author decided to create his own alternative.

2.3 PyViz

NS-3 PyViz is a live simulation tool. No trace files are required for simulation with NS3. It is directly connected to the NS3 program and displays the results immediately. PyViz is written in Python, so it is compatible not only with NS3 projects written in Python, but also in C ++. Like NetAnim, PyViz does not allow to plot results on maps [4].

2.4 Python & Flask

The server side of the project was written in the Python language. On the server side, the simulation data are processed and sent to the client side. As a web server, Flask framework is used. The Web Server is serving the webpage with a preprocessed map with all network nodes [5]. Flask is a micro web framework written in Python. Flask is under Berkeley Software Distribution (BSD) license. Development in Flask is straightforward and easy. Therefore, a simple web app is possible to create in minutes. Flask is also using the Jinja template system [5].
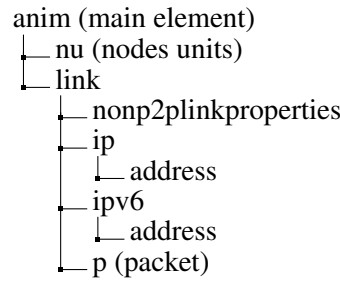
2.5 Folium

Folium is a Python library for visualizing geospatial data. In Python, Folium is as a wrapper for Leaflet. Leaflet is a map engine in the JavaScript language. Within the project, Folium is used as a generator of map data and places individual network nodes on the map. The map also includes city districts as logical units of the network. Folium creates a graphical boundary for these logical units [6].

3 DEVELOPMENT OF WEB APPLICATION

3.1 Simulation data from Network Simulator 3

The author focused on backward compatibility with the NetAnim program and uses an XML file with simulation data, which it processes in a web application. XML files have a tree structure and individual elements are embedded in the parent elements. In the first version of the web application, a total of nine different elements are searched for in the XML file. The structure of XML file is described in Listing 1.

Every element has attributes with information about nodes and the course of the simulation. The root element contains information about the version of NetAnim (*ver*) for which the XML file is intended. An important element is the node. It contains information about the node id (*id*) and Global Positioning System data, which are used to place the node on the map (*locX* for longtitude and *locY* for latitude). The third element of the node unit can be of four types. The first node unit type contains data about the node color. The second type contains data about the position used in the NetAnim grid. The third type contains data about descriptions. A description is used for graphical distinguishing of

```
anim (main element)
├── nu (nodes units)
└── link
    ├── nonp2plinkproperties
    ├── ip
    │   └── address
    ├── ipv6
    │   └── address
    └── p (packet)
```

Listing 1: XML file tree structure.

the type of node (router, access point). And the last type contains data about the size. Next important node is the link. Link contains data about the connection between two nodes. It contains two IDs of nodes that will be connected. It also contains data about the Internet Protocol Address (IP) and the Media Access Control Address (MAC address) of both nodes. The equivalent of the link element is the "nonp2plinkproperties" element. It contains data about wireless connection only for one side (transmitting node). Information about IP version 4 address is in the IP element. IP element contains the sub element's named address. The IP address is stored as the value of an element. Therefore, it is not an attribute, but the value of an element. For IP version 6, there is an element named ipv6. The last element packet contains data about communication across network. The element packet has six attributes. First attributes are ID of the node which is the author of the sent information (*fID*) and the message recipient (*tID*). Shipping times follow as the first bit transmit time of the packet (*fbTx*) and the latest bit time transmit of the packet (*lbTx*). Same timestamps are for receiving bits of the packet (*fbRx* and *lbRx*).

3.2 DESCRIPTION OF THE APPLICATION BACKEND CODE

The server side of the project is consisting of three files. The first file is named *pynsa.py* and it is the main file for the whole project. The second file is named *map_gen.py* and it is used for map generation. The third file is named *xml_parsing.py* and in this file the XML file is parsed. In the main Python file, the Flask server is running. Next, the main class includes methods for routing specific addresses obtained from Hypertext Transfer Protocol (HTTP) requests. Particularly speaking about */index* address and */simulation_data* address. When the address */index* is called, the initial data processing is started for displaying the map with the city districts. However, simulation data is not yet included. These are then uploaded via the website by the user. The process of loading the page is shown in Figure 1.
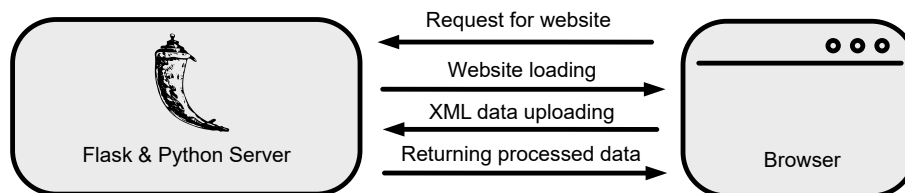

Figure 1: Internal communication in the application between web interface and server backend.

Working with the map is performed in the *map_gen.py* class. The class *map_gen.py* is divided into several methods. The methods are called depending on the input data. In the first loading page, the simulation data is not present yet. Therefore, the map is loaded only with town districts. Creating a town district on the map is done by the method *add_districts()*. The data used for town district zones come from the Czech Office for Surveying, Mapping and Cadastre. Data was converted to

the GeoJSON format. And with the method GeoJson from the Folium library data was applied to the map. In future, the town districts will be used for color highlighting of the lost connection in a specific town district. After user's uploading the XML file, the rest of methods are used. The first method adds the network nodes (*add_network_nodes()*). Using coordinates from XML, a node is placed on the map and the icon is choosen by its description. After that, the nodes are connected via lines representing the physical network connection (*add_links()*). The last thing is to add circles representing the radio connection (*add_wireless_links()*). Class *xml_parsing.py* contains methods for parsing XML with simulation results. Every method is used for specific elements in the XML file. The first method *parse_xml()* is used for obtaining an object of XML data. From this object, other methods get data for specific use (nodes, links, wireless links). The next method is *get_nodes()*. This method is creating a list of nodes for subsequent placing on the map. For information about the physical connection between nodes the method *get_p2p_links()* is used. The equivalent wireless method is *get_non_p2p_links()*. The last method is *get_sim_data()*. This method parses data about communication between nodes. These data are used in the simulation process on the website page. The final simulation data are sent by a Flask data channel right to the website.

3.3 DESCRIPTION OF THE APPLICATION FRONTEND CODE

For the frontend side the Hypertext Markup Language (HTML), JavaScript, and Cascading Style Sheets (CSS) are used. The Leaflet.js file is imported for full support of the Leaflet library. For nice looking of buttons and the whole User Interface (UI) the Bootstrap framework is used. When a user opens the webpage of the simulator, it is possible to see maps across the entire webpage and a small toolbar menu at the top of the page. The webpage is shown in Figure 2. There are several controls on the toolbar. The first is a button for uploading an XML file with simulation results from NS3. When you click the button, a dialog box appears with the option to select the XML file. After confirming the selection, the file is automatically sent to the server and processed. Once the data is processed, the web page is automatically refreshed and the results from the XML file are written to the newly loaded web page.Next to the button for loading the XML file, there is a timeline of the simulation, which in the form of a slider shows the time in which the simulation is currently located. For a more accurate orientation, the simulation time is further listed with the label next to the time axis. To start the simulation, next to the time indicator, there is a "Play / Pause" button to start or pause the simulation. The last item on the toolbar is a dropdown menu with simulation speed settings. In the basic mode, 100 ms is set for the simulation step. To adjust the time, it is possible to use the following options: 10x, 5x, 1x, 0.5x, and 0.1x. The simulation can therefore be speeded up and slowed down if necessary. Dynamically displayed data is controlled using the JavaScript file *simulation.js*. When you click the Play button, the simulation starts. There are a series of elements called "p" (packet) in the XML file, which correspond to the steps in the simulation. The processed simulation data sent from the Flask server are browsed using JavaScript code and any communication is displayed on the map. Displaying results is based on the location of nodes in the backend part. There are two IDs in the "p" element that refer to the nodes located on the map. When processing a new simulation step, the coordinates of the marker (node) are obtained from the markers on the map according to the ID. Using these coordinates, both ends of the line are then positioned. A floating window was created for simulation purposes. Only the last 20 steps of the simulation are displayed, and the others are gradually deleted. The floating window works on the principle of the First In First Out (FIFO) queue. Line objects are placed in a floating window, that is, an array. At each iteration of the simulation, the line object is always deleted first from the map and then from the floating window.
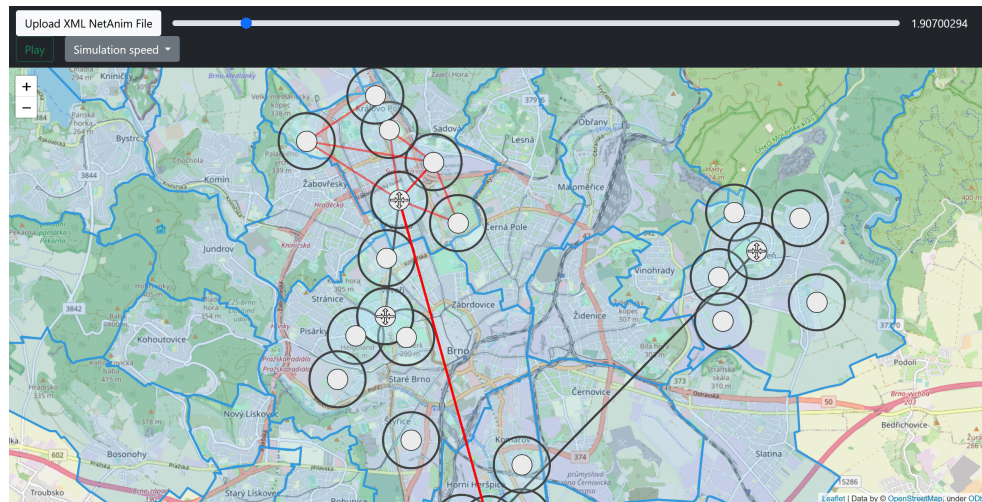
Figure 2: Website interface of created application.

4 CONCLUSION

This paper is dealing with the reasons and the creation of an alternative graphical environment for the presentation of simulation results from the program NS3. The created application allows to present the simulation on a map. The application is an alternative to the NetAnim and PyViz programs, which do not allow to display simulation results on a map. The application is created in Python and is therefore cross-platform. The application needs a Python interpreter and a modern web browser to run. Thanks to the fact that the application has a web interface, it can be run on a remote server and the user can access it remotely.

ACKNOWLEDGEMENT

REFERENCES

[1] About: What is ns-3? Nsnam.org [online]. Berkley: nsnam.org, 2021, 2021 [cit. 2021-03-10]. Available from URL: https://www.nsnam.org/about/

[2] Python 2.7.18, the last release of Python 2. Python Insider [online]. Worldwide: Benjamin Peterson, 2020 [cit. 2021-03-10]. Available from URL: https://pythoninsider.blogspot.com/2020/04/python-2718-last-release-of-python-2.html

[3] NetAnim 3.108. Wiki nsnam.org [online]. Berkley: nsnam.org, 2017 [cit. 2021-03-10]. Available from URL: https://www.nsnam.org/wiki/NetAnim_3.108

[4] PyViz [online]. Berkley: nsnam.org, 2015 [cit. 2021-03-24]. Dostupné z: https://www.nsnam.org/wiki/PyViz

[5] Welcome to Flask: Flask Documentation (1.1.X) [online]. Hermagor, Austria: Armin Ronacher, 2020 [cit. 2021-03-10]. Available from URL: https://flask.palletsprojects.com/en/1.1.x/

[6] Folium: Folium 0.12.1 documentation [online]. Worldwide: Rob Story, 2013 [cit. 2021-03-12]. Available from URL: https://python-visualization.github.io/folium/

UNIVERSAL MULTI-RAT TESTER FOR MMTC APPLICATIONS

[1]**Radovan Juráň**, [2]**Radek Možný**

Doctoral Degree Programme ([1]1st year, [2]2nd year), FEEC BUT

E-mail: [1]xjuran15@vutbr.cz, [2]xmozny03@vutbr.cz

Supervised by: Petr Mlýnek, Pavel Mašek

E-mail: mlynek@feec.vutbr.cz, masekpavel@vutbr.cz

Abstract: While the IoT has made significant progress along the lines of supporting its individual applications, there are many massive Machine-Type Communication (mMTC) scenarios in which the performance offered by any single Radio Access Technology (RAT) available today might be insufficient. To address these use cases, we introduce the concept of multi-RAT tester, which implies the availability and utilization of several RATs within a single IoT device. We begin by offering insights into which use cases could be beneficial and what the key challenges for mMTC implementation are. We continue by discussing the potential technical solutions and employing our own prototype of a multi-RAT device capable of using different Low-Power Wide-Area (LPWA) communication technologies. It is assumed that use of multiple radios simultaneously will lead to overall improvement of communication parameters by leveraging the synergy between RATs. The novel vision enabled by the multi-RAT concept in this work could be impactful across multiple fields and calls for cross-community research efforts in order to adequately design, implement, and deploy future multi-RAT mMTC solutions.

Keywords: LPWA, Multi-RAT, Industrial Internet of Things, SmartGrid, massive MTC

1 INTRODUCTION

The Internet of Things (IoT) growth gave rise to many wireless communication technologies operating in both licensed and unlicensed bands, which are applicable also for the wireless machine-to-machine (M2M) communication. As stated in [1], the M2M is mainly characterized by low data rates, limited bandwidths, and relatively simpler radio devices with very low energy consumption, which contrasts with the needs of human-to-human (H2H) cellular networks, focused on high data rates interchanged between rather complex devices.

In spite of that, the telecommunication providers tend to utilize their existing infrastructures when considering M2M, because not only do they already cover considerably large geographical areas, thus eliminating the need to build new infrastructure, but they also work in the licensed bands, which brings predictable environment properties and especially controlled interference – a parameter not always achievable in unlicensed bands [1].

Reliability of the service goes hand in hand with its security and flexibility – the key issues to address in any general radio communication system, especially one that includes a large number of connected devices, referred to as massive Machine Type Communication (mMTC) [2]. A typical example of the mMTC use case can be any smart-grid application, such as electricity meter data reading, power line monitoring, or demand management [1]. Another example is public transport tracking. One of the ways to enforce the technical performance requirements defined by [2] for reliability-demanding applications can be combining existing technologies in one device together – approach known as Multiple Radio Access Technology (Multi-RAT).

2 UTILIZATION OF LPWAN MULTI-RAT FOR MMTC APPLICATIONS

In general, LPWA technologies have proven to be an ideal and efficient choice to satisfy mMTC applications' needs. Thus sporadic delay-tolerant data transmission for long distances ensures extended coverage and, at the same time, ensures high cost and power efficiency [3, 4, 5].

Each LPWA technology presents unique features, as observed from the comparison Table 1. As was addressed by authors already in, e.g. [3, 4], those technologies utilizing unlicensed spectrum such as Sigfox and LoRaWAN benefits from low operational expenses (OPEX). On the other hand, there is a drawback in limited time-on-air (ToA) due to the duty cycle or need for infrastructure deployment, thus increasing capital expenditures (CAPEX). On the other hand, mobile technologies like NB-IoT or LTE Cat-M provide almost unlimited service with already well-built infrastructure. However, with a trade-off in the form of extended OPEX due to the charge for the amount of transmitted data.

Table 1: Technical parameters comparison of widely used LPWA technologies in Europe [3, 4].

|	**Sigfox**	**LoRaWAN**	**LTE-M**[1]	**NB-IoT**[1]
Coverage (MCL)	159 dB	155 dB	155.7 dB	164 dB
Spectrum	Unlicensed	Unlicensed	Licensed	Licensed
Max. ERP	14 dBm	14 dBm	23 dBm	23 dBm
Modulation	DBPSK (UL) GFSK (DL)	LoRa (CSS) FSK	QPSK, 16QAM	$\pi/2$ - BPSK, $\pi/4$ - QPSK, QPSK (DL)
Module cost	2 $	6 $	20 $	8 $
Restrictions	140 (UL), 4 (DL) Mess. per day	Duty Cycle	Charged data	Charged data
UL datarate	100 bps	$0.3 - 50$ kbps	1 Mbps	$0.3 - 62.5$ kbps
DL datarate	600 bps	$0.3 - 50$ kbps	1 Mbps	$0.5 - 27.2$ kbps
Max. app. payload	12 B / 8 B (UL / DL)	$51 - 242$ B	1600 B[2]	1600 B[2]

[1] According to the 3GPP Release 13.
[2] The maximum size of a Packet Data Convergence Protocol Service Data Unit (PDCP SDU).

Another comparison angle could be signal coverage. As cellular technologies provide almost ubiquitous availability in urban and rural areas, there might be specific cases like mines and ships where mobile technologies' deployment is difficult. Here the provision of continuous connection calls for Multi-RAT usage [3]. It is certain that for many use-cases, a single RAT is sufficient or even required to meet the application requirements. However, there are also cases like those mentioned above that point out that Multi-RAT solutions also have their momentum. Moreover, for certain can be said that there is no single RAT for all mMTC use cases.

In our previous works [3, 6, 7], the large single RAT coverage measurement campaign was thrown with an aim to evaluate coverage for Sigfox, LoRaWAN, and NB-IoT in the Brno, Czech Republic as an example of a mid-sized European city in terms of single RAT or possible Multi-RAT usage. The obtained results are depicted in Fig. 1. The goal was to find out the service availability at the tram stations and overall coverage for possible use cases such as public transport tracking or smart-meter deployments. It is important to note that the NB-IoT map is not included as all the points were served, which indicated omnipresent NB-IoT service availability with average reference signal receive power (RSRP) -85 dBm (good radio conditions). According to the results, it can be said that NB-IoT provides flawless coverage but with charged data trade-off. On the other hand, there were several signal outage areas in the case of Sigfox and LoRaWAN. See more information about the measurement campaign in [3, 6, 7].

The measurement campaign results encouraged authors to work on a Multi-RAT solution, which addresses a few of the mMTC use-cases that are to be widely implemented utilizing the benefits Multi-
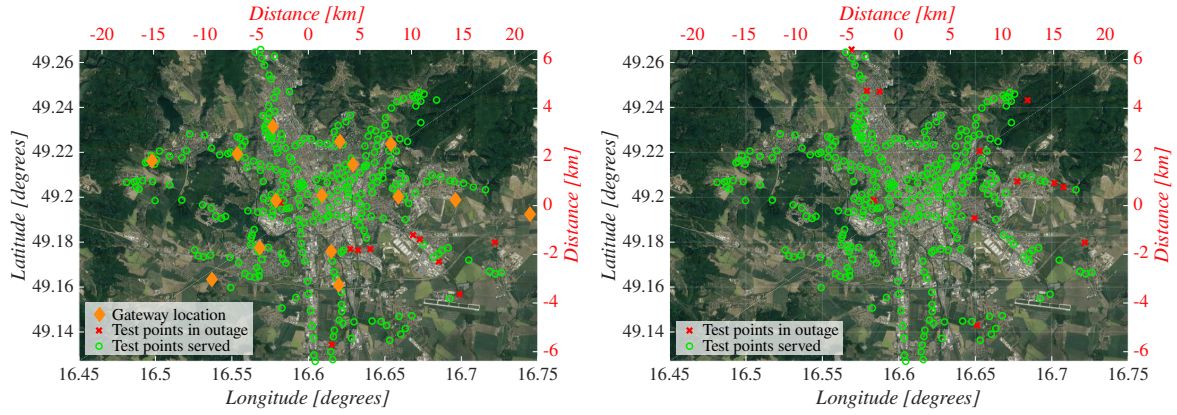
(a) LoRaWAN converage in Brno Czech Republic [7]. **(b)** Sigfox converage in Brno Czech Republic [6].

Figure 1: Coverage measurement campaings of LPWA technologies in Brno Czech Republic 2019.

RAT approach. The first use-case already being implemented is public transport tracking, where LPWA technologies are ideal for data transmission of position to a remote server. Here the challenge relies upon the continuous connection availability as addressed in our previous work [3].

The second and at the same time, the main intended use-case is, as was mentioned, smart-grid, more precisely smart metering. This aim in the case of the Czach Republic is crucial as the selective deployment of electricity smart-meters utilizing NB-IoT and secondary RAT by 2024 is set by legislation and currently deployed meters utilizing power line communication (PLC) and MESH technologies are not suitable for dense deployment. This only enhances the need for a solution such as the one presented in this work. In general, smart the key performance indicator (KPI) for smart meters is not on the side of power efficiency, but ensuring the service availability in remote areas and, e.g., delay tolerance with 10 s [2, 8]. Mentioned remote areas and places under the surface means signal dead-zones for legacy systems. As for the technology most suitable for the scenario, several works addressed this topic (such as [1]) with NB-IoT as the best choice. However, the secondary technologies could extend the versatility for this use case in the Multi-RAT scenario.

With all this in mind, together with the fact that each application has different demands, the need for more universal devices such as the one delivered in this work and addressed in Section 3 is clear. Now one device able to combine all mentioned prerequisites could provide enough versatility in the form of an all-in-one module for the applications mentioned earlier.

However, the Multi-RAT approach also brings certain drawbacks and challenges for sufficient and optimized use. The first downside is the device cost. The use of Multi-RAT requires a more complex device design with multiple communication modules. This inevitably increases the expenses. The second drawback to be addressed is the power efficiency and overall performance considering efficient RAT switching as was addressed in [3]. The machine-learning techniques provide an efficient way how to handle these optimizations [5]. Authors of [9] already elaborated on power-efficiency enhancements utilizing reinforced-learning (LA) that could be implemented even in constrained devices with promising results. Since the proposed design of the device considered batteryless operation, the power efficiency is not KPI as was already mentioned, but the work has proven that LA could be a way to go for the overall Multi-RAT operation optimization.

3 DESIGN OF MODULAR UNIT FOR DATA GATHERING AND TRANSMISSION UTILIZING MULTI-RAT APPROACH

The proposed communication unit is based on Raspberry Pi Compute module 3 (CM). As shown on overall block schematic diagram in Figure 2, the device consists of two main boards – motherboard

for the CM and an extension communication board, which offers two miniPCIe slots for radio communication modules. The whole system seats in a compact housing with standard DIN rail mounting.
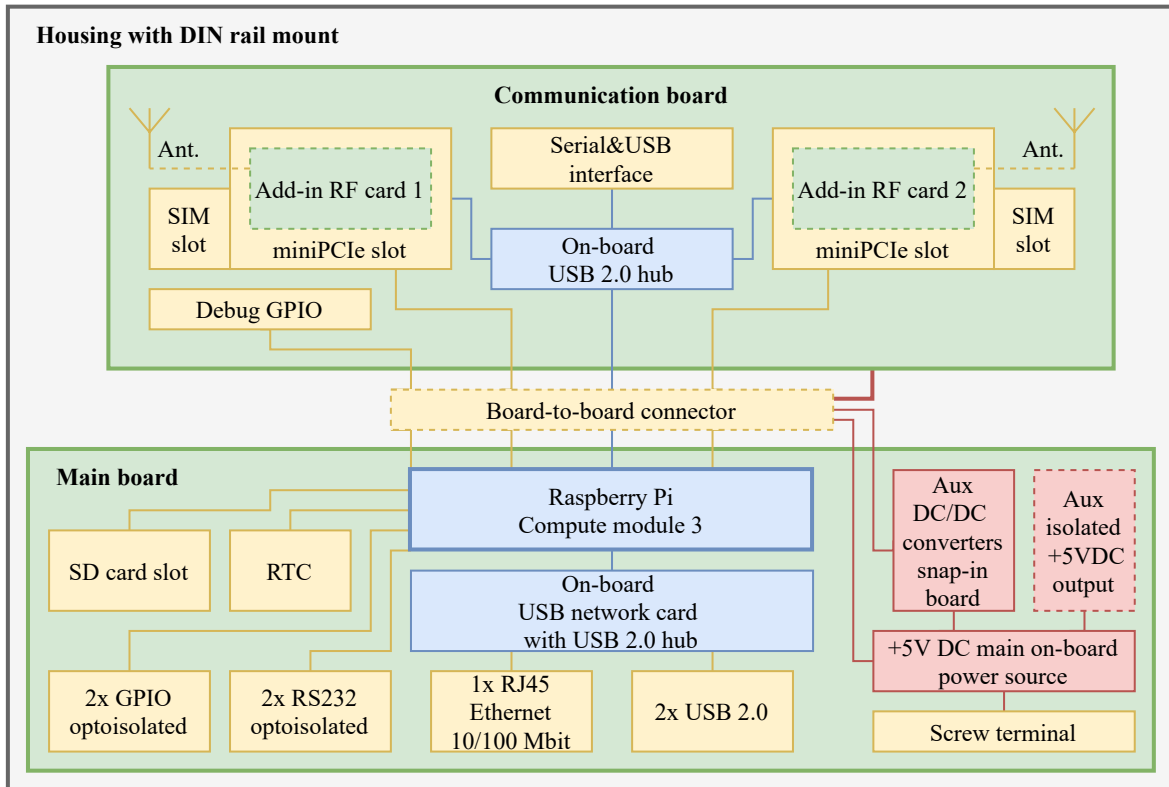

Figure 2: Block schematic diagram of the Multi-RAT communication unit.

The motherboard for the CM is flexibly designed to support mounting options for CM sub versions and to exploit potential of its interfaces. There is a custom-built on-board USB network card with integrated USB 2.0 hub build around Microchip LAN9514, which offers 1x 10/100 Mbit Ethernet and up to 4 USB 2.0 ports, of which two are used for external device connectivity and one is used as board-to-board connection with the RF board. The other interfaces were connected to separate screw terminals, namely two optically isolated RS485, two optically isolated GPIOs, and one +5 V DC supplementary isolated power source. The CM has been extended by an independent RTC with 12 mm coin cell battery holder and micro Secure Digital (SD) card slot.

The power is delivered via screw terminals from external power source (intended to use with auxiliary power source of electricity meter) to the on-board custom designed main power source, which the distributes the power to the main voltage +5 V DC branch on the board directly as well as indirectly via supplementary snap-in power board, which supplies lower voltage levels of +3.3 V DC and +1.8 V DC.

The detachable extension board offers two standalone miniPCIe slots, each supplemented by its own Subscriber Identity Module (SIM) card holder and LED indications as per miniPCIe spec. This brings the true universality to the proposed Multi-RAT solution as it allows to fit any add-in RF cards of various communication standards via one unified interface, either off-the-shelf cards or even custom-built. As the main bus utilized to use with miniPCIe cards is the USB, there is an on-board Texas Instruments TUSB4041I custom USB hub, which interconnects the miniPCIe slots with the main board and on top of that one serial to USB debug interface.

4 CONCLUSION

In this paper, we addressed the idea of Multi-RAT use for mMTC applications such as smart-metering or public transport tracking. It is clear that there is no killer app nor single RAT that could satisfy the demands for such a diverse set of mMTC use cases as might indicate the LPWA measurement campaign in Section 2. Thus the need for modular communication devices, as the one carried out in this work, capable of adopting a Multi-RAT approach is required.

Presented design in Section 3 provides flexible all-in-one solution supporting interfaces for external devices such as electricity meter or other sensors actuators. Modular RAT interfaces in the form of mPCIe slots enable custom setup tailored to the specific mMTC use-case needs. The above-mentioned proved the design suitable for the diverse mMTC applications of public transport tracking smart-meters.

The natural step in future work is to construct the device, verify its functionality with initial measurements, possibly optimize the device and subsequently deploy it with RATs according to the use-case demands. As smart metering is a hot topic now in the 2021 Czech Republic due to its intended deployment of smart electricity meters by 2024, the future test's main focus is in this area. Including possible optimizations (as were addressed in Section 2) such as the implementation of the machine learning technique. This approach might improve overall device performance in the form of predictive transmission according to the signal quality and changes for all RATs or to lower the operational expenses by optimized RAT switching.

REFERENCES

[1] S. Persia and L. Rea, "Next Generation M2M Cellular Networks: LTE-MTC and NB-IoT Capacity Analysis for Smart Grids Applications," in *2016 AEIT International Annual Conference (AEIT)*, pp. 1–6, 2016.

[2] ITU, "Minimum Requirements Related to Technical Performance for IMT-2020 Radio Interface(s)," report ITU-R M.2410-0, November 2017.

[3] R. Mozny, M. Stusek, P. Masek, K. Mikhaylov, and J. Hosek, "Unifying Multi-Radio Communication Technologies to Enable mMTC Applications in B5G Networks," in *2020 2nd 6G Wireless Summit (6G SUMMIT)*, pp. 1–5, 2020.

[4] K. Mekki, E. Bajic, F. Chaxel, and F. Meyer, "A Comparative Study of LPWAN Technologies for Large-scale IoT Deployment," vol. 5, pp. 1–7, 03 2017.

[5] N. Mahmood, O. Lopez, O. S. Park, I. Noerman, K. Mikhaylov, E. Mercier, A. Munari, F. Clazzer, S. Bocker, and H. Bartz, "Massive Machine Type Communication Towards 6G." [White paper], 6 2020. (6G Research Visions, No. 11). University of Oulu.

[6] K. Mikhaylov, M. Stusek, P. Masek, R. Fujdiak, R. Mozny, S. Andreev, and J. Hosek, "Communication Performance of a Real-Life Wide-Area Low-Power Network Based on Sigfox Technology," in *ICC 2020 - 2020 IEEE International Conference on Communications (ICC)*, pp. 1–6, 2020.

[7] K. Mikhaylov, M. Stusek, P. Masek, R. Fujdiak, R. Mozny, S. Andreev, and J. Hosek, "On the Performance of Multi-Gateway LoRaWAN Deployments: An Experimental Study," in *2020 IEEE Wireless Communications and Networking Conference (WCNC)*, pp. 1–6, 2020.

[8] GSMA, "3GPP Low Power Wide Area Technologies," White Paper, May 2018.

[9] M. Stusek, D. Moltchanov, P. Masek, J. Hosek, S. Andreev, and Y. Koucheryavy, "Learning-Aided Multi-RAT Operation for Battery Lifetime Extension in LPWAN Systems," in *2020 12th International Congress on Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT)*, pp. 26–32, 2020.

NETWORK TRAFFIC EVALUATION OF MICROWAVE POINT-TO-POINT LINK PRECIPITATION PREDICTIONS

Petr Musil

Doctoral Degree Programme (3), FEEC BUT

E-mail: xmusil56@stud.feec.vutbr.cz

Supervised by: Petr Mlynek

E-mail: mlynek@feec.vutbr.cz

Abstract: Precipitation prediction derived from attenuation of microwave signal by absorption and scattering of raindrops becomes modern supplementary data input for complex weather predictions. This paper deals with the topic of estimation of data volumes incorporated in the process of data acquisition. Description of network management protocols and its communication with the evaluation of minimum and full data acquisition scenarios is also included. The main goal of this article is to discuss issues related to data acquisition over network, data traffic volumes and storage concerns of the precipitation prediction methods. The conclusion reveals limitations and trade-offs of the process.

Keywords: ICMP, SNMP, rainfall prediction, precipitation observation, microwave attenuation.

1 INTRODUCTION

In the past decades, there has been a significant effort in the research and development of radars for weather analysis and prediction. Radars nowadays have higher resolution, but there are still limitations and obstacles that decrease weather forecasts overall accuracy, i.e., lower accuracy at surface level, issues related to complex terrain profile, and its beam blockage. Perspective idea complementary to weather radars has also been published in research [1], which proposed rain rate prediction based on received signal level values (RSL) attenuated by raindrops within the line-of-sight (LOS). Furthermore, this idea was developed with the focus on various methods and approaches of rain rate estimation of microwave links (MWL) in telecommunication networks. Predictions are usually based on commercial microwave links (CML), which are used in B2B (Business-to-Business) telecommunications or as a backhaul connection of telecommunication and cellular networks [2].

The principle of proposed predictions is based on propagation impairments of radio links caused by weather, in this case, attenuation of precipitation of the signal of point-to-point MWL, as Fig. 1 illustrates. Absorption and scattering of microwaves are caused by raindrops of a size comparable to the wavelength of MWL. Rain attenuation varies with the size, shape and distribution of water droplets [3].

Signal attenuation observed at the microwave unit depends on the intensity of the rain and the operating frequency. In general, higher intensity of rain leads to stronger specific attenuation at the given frequency. Also, a higher operating frequency increases the specific attenuation at the given rain intensity [4]. Research [5] proved that horizontal polarization shows higher attenuation than vertical.

Implementation of rainfall estimation requires tight cooperation of microwave link providers and researchers. Prediction requires data acquisition with a sufficient sampling rate and high availability. This technique is very promising due to already built infrastructure, where no additional costs are required. Prediction based on evaluation of RSL of microwave link should be implemented together with radar and rain gauges systems to increase overall accuracy due to near ground character of rainfall observations [6].
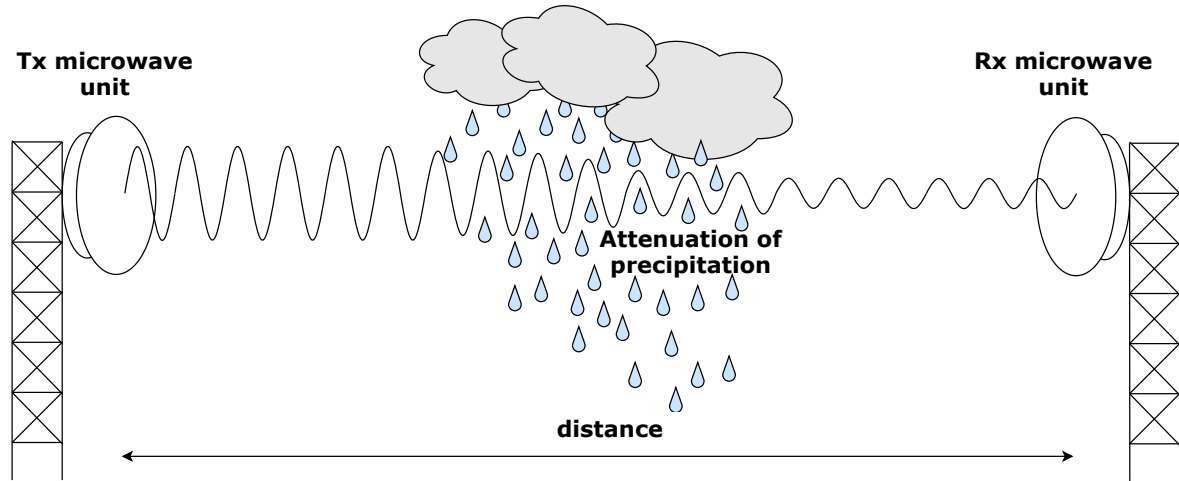
Figure 1: Structure of typical SNMP request/response frame.

Many research groups nowadays focus on the development of methods for more accurate weather prediction models and data classification. However, very few estimates have been done in terms of data load that polling of microwave units over IP network will induce. This article mainly focuses on revealing additional data load on the telecommunication network required for data acquisition.

2   NETWORK MANAGEMENT

Monitoring of devices within the network is usually being performed by a combination of ICMP (Internet Control Message Protocol) and SNMP (Simple Network Management Protocol).

ICMP is being used as a control and diagnostic tool for testing the availability of network devices, error indication, measuring the latency of the network, or determining the number of hops. It runs on a network (3) ISO/OSI layer. Many network utilities are based on the ICMP, i.e., traceroute for determining possible routes and measuring transport delays across the IP network. Utility ping is used for testing the reachability of network devices by commands echo request and echo reply. Ping can provide statistical results about availability, errors, packet loss, and network response over time.

On the other hand, SNMP is classified as application (7) ISO/OSI layer protocol, which is used for collecting information about monitored devices of IP networks and for management of such devices. SNMP uses an organized structure in the form of variables called management information base (MIB), which describes the system status and configuration. SNMP system consists of two key components: manager and agent. The manager, usually the server, performs the monitoring and managing function of the controlled network by polling. The agent is a monitored device that replies to manager requests or sends trap messages (one-way messages from an agent to the manager).

3   MEASUREMENTS

Methods of rainfall prediction require two key parameters of the microwave units of the microwave link: transmitted signal level (TSL) and received signal level (RSL). By evaluating the time-dependent magnitude of TSL and RSL attenuation from the stable conditions, the rainfall can be identified, and consequently, path-average rain rate can be calculated. In general, microwave units can provide more detailed information about their operation, which can be later used as supplementary information for further data analysis.

This article focuses on real-time data access with a relatively short period of data acquisition, which is beneficial in deep and precise rainfall predictions. Offline data which CML operators provide to researchers usually tend to have a very coarse time resolution of approximately 5–15 minutes. Thus, they induce prediction inaccuracies due to the averaging over longer periods.

The network management system used for these measurements runs on the Linux server, connected to the high capacity fiber backbone network to assure high reliability and low delay of monitoring network devices. The traffic of that interface was then captured for network analysis by the Wireshark network protocol analyzer.

3.1 SCENARIO A: MINIMUM DATA RETRIEVAL

This scenario describes the minimum solution for enabling precipitation observations of microwave links. It is assumed that only parameters of TSL and RSL are going to be retrieved. RSL and TSL data requests were being performed via SNMP get-request message. The microwave unit then replied with SNMP get-response. Fig. 2a describes the structure of typical SNMP request/response frame. Data volume structure of communication protocols incorporated in data acquisition is shown in the Fig. 2b, where only SNMP contains desired data, other protocols are considered as communication overhead.
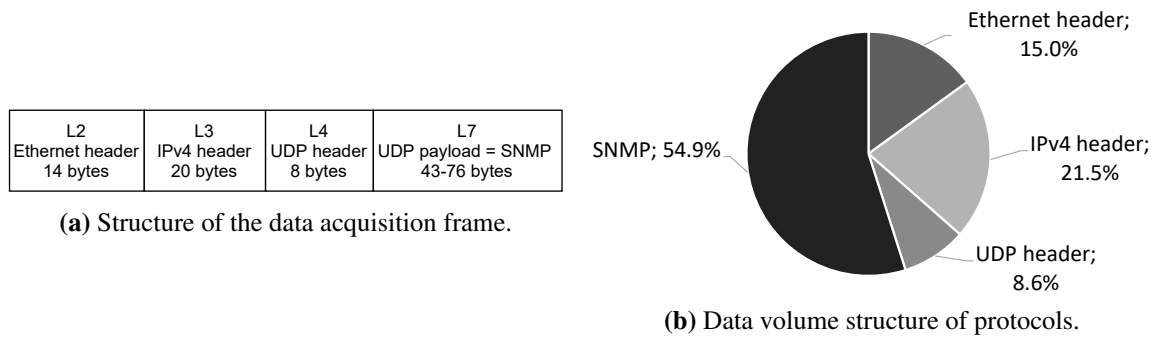

(a) Structure of the data acquisition frame.

| L2 Ethernet header 14 bytes | L3 IPv4 header 20 bytes | L4 UDP header 8 bytes | L7 UDP payload = SNMP 43-76 bytes |

(b) Data volume structure of protocols.

Figure 2: Structure of SNMP request/reply messages.

3.2 SCENARIO B: FULL DATA RETRIEVAL

For deeper and more precise rainfall prediction analysis, it was intended in Scenario B to acquire supplementary data. In this case, both microwave units of MWL were requested for five parameters: TSL, RSL, outdoor unit temperature, signal-to-noise ratio (SNR), and state of ACM (Adaptive Coding and Modulation).

Mentioned parameters could serve as a validation of rainfall events, i.e., when on a hot sunny day, sudden cold rainfall decreases the temperature of the outdoor unit significantly. Parameters of signal-to-noise ratio (SNR) and modulation and ACM state can help determine possible effects of propagation issues (obstacles, blockages, reflections) or noise effects, leading to the influence of RSL values.

SNMP data exchange was accompanied by 5 ICMP request and ICMP reply messages, where the difference of transportation delay and its deviations from the long-term average might point out possible issues on MWL or weather-related impact on overall MWL performance.

In both scenarios, it is assumed that the polarization of antenna and Tx/Rx frequency, required for rain rate calculations, is known and does not change dynamically over time.

4 RESULTS

Data volumes of one data acquisition, measured on real MWL network of CML operator for both scenarios, are summed up in Table 1. Scenario A with minimum data polling uses approximately 24.3 % of traffic volume compared to Scenario B with full data polling for accurate weather prediction.

Table 1: Network data volume of one data retrieval of microwave link.

Message type	Scenario A		Scenario B	
	average size [B]	*number of messages*	*average size [B]*	*number of messages*
SNMP get-request	89	4	89	10
SNMP get-reply	94	4	94	10
ICMP request	-	-	58	10
ICMP reply	-	-	60	10
Total	**732**	**8**	**3010**	**40**

The volume of one acquisition of one microwave link's parameters is seemingly low, but with shorter periods of acquisition, traffic volume steeply increases, as is shown in Table 2. This effect is further multiplied by a high number of operating units of MWL. The number of operating devices in the Czech Republic is estimated in tens of thousands of microwave units. Overall data volumes and the number of monitored units should give a starting point for a precipitation prediction system design.

Table 2: Projection of total data volume over time for different data retrieval periods.

Scenario A					Scenario B				
period [s]	*total volume over time [MB]*				*period [s]*	*total volume over time [MB]*			
	1 hour	**1 day**	**1 week**	**1 year**		**1 hour**	**1 day**	**1 week**	**1 year**
1	2.64	63.2	442.7	23084	**1**	10.8	260.1	1820.4	94923
10	0.26	6.3	44.3	2308	**10**	1.1	26.0	182.0	9492
30	0.09	2.1	14.8	769	**30**	0.4	8.7	60.7	3164
60	0.04	1.05	7.38	385	**60**	0.2	4.3	30.3	1582

5 DISCUSSION

One of the vital precipitation prediction parameters is the interval of data polling. The presented measurement results show that especially polling period of 1 second generates a significant volume of transmitted data. The estimated amount of devices included in the analysis might lead to network congestion and nonnegligible data traffic on the network's management part.

From the storage point of view, analysis has been performed over Scenario B SNMP data for weather prediction. Data for each microwave unit were stored in .csv format with a period of 30 seconds of data polling. 24-hour long file size varied from 96 to 187 kB for chosen microwave units of different vendors. This example gives us an overall overview of possible data volumes for thousands of units that would need to be processed by weather prediction algorithms. It is essential to point out that with a shorter period of data polling, the overall data traffic and volume increase significantly.

As was proven in previously published research, CML operators usually lack the motivation to cooperate with data provision. Their use case for data is different, and there is no standardized way of obtaining data from them. CML operators also tend not to publish their equipment locations due to possible issues related to market competition and safety concerns.

6 CONCLUSION

This paper focuses on the topic of data acquisition for modern methods of observation of precipitation estimated from the real-time values of transmitted signal level (TSL) and received signal level (RSL). Two scenarios of minimum and full data acquisition were presented. The results show that data acquisition, especially at short periods and for many monitored parameters, loads the communication network. Omitting this fact might lead to network congestion and waste of bandwidth. Based on the presented measurements, a trade-off between wasting bandwidth and losing accuracy of precipitation observation should be the subject of further research. Analysis of overhead to transmitted data also reveals the efficiency of data acquisition. The overall volume of data also makes a significant requirement on the data processing hardware, especially with a high number of microwave links.

Implementation of precipitation observation and prediction methods is a highly effective way of increasing the accuracy of the weather predictions. To enable such predictions, it is essential to define a standardized way of obtaining data from CML operators and motivate them for such data provision. On the other hand, data acquisition should be made cautiously to avoid undesired network overload, increased latency, and overall instability.

REFERENCES

[1] MESSER, Hagit, ZINEVICH, Artem and ALPERT, Pinhas. 2006. Environmental monitoring by wireless communication networks. *Science (New York, N.Y.)* [online]. **312**(5774), p. 713–713. Available from: doi:10.1126/science.1120034

[2] CHWALA, Christian, KUNSTMANN, Harald, HIPP, Susanne, SIART Uwe and EIBERT, Thomas. 2012. Precipitation observation using commercial microwave communication links. In: *2012 IEEE International Geoscience and Remote Sensing Symposium* [online]. IEEE, p. 2922–2925. Available from: doi:10.1109/IGARSS.2012.6350714

[3] OVEREEM, Aart, LEIJNSE, Hidde and UIJLENHOET, Remko. 2011. Measuring urban rainfall using microwave links from commercial cellular communication networks. *Water Resources Research* [online]. Washington: Blackwell Publishing, **47**(12). Available from: doi:10.1029/2010WR010350

[4] UIJLENHOET, Remko, OVEREEM, Aart and LEIJNSE, Hidde. 2018. Opportunistic remote sensing of rainfall using microwave links from cellular communication networks. *Wiley Interdisciplinary Reviews: Water* [online]. Hoboken, USA: John Wiley & Sons, 2018, **5**(4), Available from: doi:10.1002/wat2.1289

[5] SHRESTHA, Sujan, DONG-YOU, Choi and FALCONE, Francisco. 2019. Rain Attenuation Study over an 18-GHz Terrestrial Microwave Link in South Korea. *International Journal of Antennas and Propagation* [online]. New York: Hindawi Limited, **2019**. Available from: doi:10.1155/2019/1712791

[6] CHWALA, Christian and KUNSTMANN, Harald. 2019. Commercial microwave link networks for rainfall observation: Assessment of the current status and future challenges. *Wiley Interdisciplinary Reviews* [online]. Hoboken: Wiley Subscription Services, **6**(2). Available from: doi:10.1002/wat2.1337

EVALUATION OF THE NEURAL NETWORK OBJECT DETECTION IN MULTI-MODAL IMAGES

Adam Ligocki

Doctoral Degree Programme (4), FEEC BUT

E-mail: adam.ligocki@vutbr.cz

Supervised by: Luděk Žalud

E-mail: zalud@feec.vutbr.cz

Abstract: This paper studies the information gain of various data domains that are commonly used in the modern Advanced Driving Assistant Systems (ADAS) to develop robust systems that would increase traffic safety. We could see a fast growth of many Deep Convolutional Neural Networks (DCNN) based solutions during the last several years. These methods are state-of-the-art in object detection and semantic scene segmentation. We created a small annotated dataset of synchronized RGB, grayscale, thermal, and depth map images and used the modern DCNN framework tool to evaluate the object detection robustness of different data domains and their information gain process understanding the surrounding environment of the semi-autonomous driving agent.

Keywords: Multi-modal, Object Detection, Convolutional Neural Network, RGB, Grayscale, Thermal, IR, Depth Map

1 INTRODUCTION

These days, we can see the dramatic development of various ADAS systems that improve all traffic members' safety. However, the question of which sensory equipment setup is the best for a fail-safety and robust autonomous agent's orientation stays unanswered. It is also quite clear that in the future, we will need to develop a wide-range of data fusion methods that will combine information from different data sources into a single robust model that will help to understand the situation around the agent.

In this paper, we focus on the several data domains commonly used in automotive. We took the RGB camera, thermal camera, and 3D LiDAR scanners, and we processed these data, so all are represented as a synchronized and unified 2D images of the traffic situations, and we tested the performance of these days state-of-the-art neural network in the task of object detection.

2 RELATED WORKS

In the publically available literature, the topic of combining multimodal data by a single deep neural network model is not new. When we talk about the thermal and visible-spectrum images data fusion methods, the most frequent topic is enhancing pedestrian detection in bad lighting or weather conditions [3] [15].

Another way to improve visible spectrum neural network object detections in traffic is to extend the RGB information with depth map (LiDAR data) [1], [9]. Literature also covers the fusion of all three domains studied by this paper [8], [10]. However, these works usually focus only on pedestrian detections and do not cover the ADAS topic or deep learning methods.

However, non of the papers mentioned above discuss the information gain from the specific data domain (visible spectrum, thermal, and depth map), which is my work's main aim.

3 MATERIALS AND METHODS

3.1 DATA GENERATION

To create the training and validation data for this experiment, we used our existing Brno Urban Dataset [6], and Atlas Fusion Framework [5].

In the first phase, we detected objects using the YOLOv5 on the raw RGB video data. Later, using the principle described in [7], to project detections from RGB images into the thermal ones. This way, we generated annotated therm images. Simultaneously, for every thermal image, the full 3D point cloud model was projected into the camera plane, so the image's corresponding depth map was generated. Last, we estimated the perpendicular plain in the thermal camera's field of view (FoV) in the distance of 50m from the camera. We projected the area of this plain into the RGB camera to approximate the common FoV for both the thermal and RGB cameras. Cutting out this area, we created the RGB image that covers approximately the same scene as the thermal image with minimal epipolar distortion, as both the RGB and the thermal cameras are placed near each other on the sensory framework. We also converted the RGB images into the grayscale to compare the object detection on the images with reduced information. In the case of this paper, we annotated only vehicles. All other classes were left unused.

Figure 1: Tuple of annotated multi-modal data. From left, RGB image, information reduced grayscale image, thermal image, depth map image and annotations visualized on RGB image on the right. Annotations are same for all images.

This way, we generated about 7500 annotated image tuples containing RGB image, grayscale image, thermal image, and the depth map. We split this set of tuples by 4:1 ratio, training to test set.

3.2 NEURAL NETWORK TRAINING

To train a neural network on our dataset, we used the existing YOLOv5 framework [4], which is publically available on Github. The framework provides modified YOLOv3 [13] architecture implemented in the Pytorch framework, mainly updated with training process augmentations and API for training models on custom datasets.

3.2.1 YOLOV3 ARCHITECTURE

The YOLOv3 [13] architecture is an evolution of this end-to-end object detection model's previous versions. Originally introduced as a "You only look once" (YOLO) [11], authors proposed the neural network model that accepts N-channel image on the input and by single inference with the image, it provides an array of detections on the output, without any region proposition phases, compared to the RCNN methods [2], [14]. The RCNN-like methods, on the other hand, separate object detection into two phases, region proposition, and region classification.

The YOLO architecture firstly passes the input image through the backend. The backed is the convolutional neural network (CNN) that produces feature maps on the output. These feature maps are later processed by the output layers that generate the tensor of object detection propositions. These tensors are NxNxM structures, where the NxN is the cell grid that divides the input image into the smaller

areas when the M is a vector of numbers that represents bounding box positions, dimensions, object detection confidence, and the classification score for all classes that model detects. See the Figure 2.
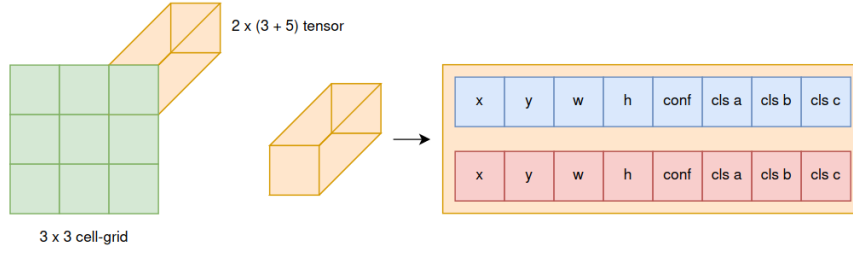

Figure 2: Visualization of the output tensor. In this use-case, the input image is divided into the 3x3 cell grid, and for each cell, there are two bounding boxes proposed. Each bounding box is defined by x and y position w.r.t the cell, width, height, detection confidence, and the classification score for three possible classes.

During the time the YOLO architecture developed, authors introduced many improvements [12], [13], like multi-class classification on the output, batch normalization, multi-scale feature map pyramid, or introducing new backend neural networks. For YOLOv3, the model has three output tensors, when each tensor proposes detections on the different cell grid sizes (20x20, 40x40, 80x80), which corresponds to the different scale and size of the detected objects.

3.2.2 METRICS

To measure the performance of our trained models, we used the commonly used Mean Average Precision score (mAP(x)), where x represents the threshold value intersection over union (IoU) between the ground truth and proposed detection to accept the proposed detection as a true positive sample. More specifically, we used the mAP(0.5) as a basic metric, the mAP(0.5:0;95) to test the model's robustness, and commonly used F1 score to express the precision-recall relation.

3.2.3 TRAINING HYPERPARAMETERS

All trained neural networks are the YOLOv5 version S models with a 640x640 image size and 3-channels per image. The grayscale, thermal, and depth images, which are 1-channel by definition, were extended by two first channel copies.

We trained each model for 100 epochs, and for the performance test, we used the model from the epoch with the best validation score. For training, we used the ADAM optimizer. For testing, we set the IoU threshold to 0.6 and the confidence threshold to 0.1.

4 RESULTS AND DISCUSSION

We tested the object detection performance on each data domain independently. The Table 1 shows the overall results. The best results showed the RGB-based and grayscale-based models, with both mAP(0.5) and mAP(0.5:0.95) on high numbers. That is evidence of the high robustness of the models. Also, the depth-map-based model shows quite good results. Compared to the IR model, it deals quite well with occlusions (Fig 3, column 3) and even with distant objects (Fig 3, column 5).

The grayscale, IR, and depth images were extended from the 1-channel format into the 3-channel by copying the first channel twice. This way, we kept the same number of convolutions applied on each data domain, the same computational requirements, and the similar model capacity.

	mAP(0.5)	mAP(0.5:0.95)	F1	Precision	Recall
RGB	0.942	0.712	0.873	0.847	0.901
Gray	0.937	0.705	0.869	0.840	0.900
Depth	0.839	0.467	0.792	0.832	0.756
IR	0.683	0.219	0.627	0.580	0.683

Table 1: Neural netowrks performance in given RGB, grayscale, depth and thermal domains. Each data domain were tested independently.

We performed experiments using the RGB and grayscale images with perfect lighting and weather conditions in this work. That makes it hard for the depth and IR domain specialized neural networks to reach similar even better results than visible spectrum-focused networks and show their advantages in the more challenging conditions. In future work, it will be interesting to focus on data where lighting conditions are degraded and where the visible spectrum sensors are partially blinded by the night, fog, or by water on the lens during the rain.
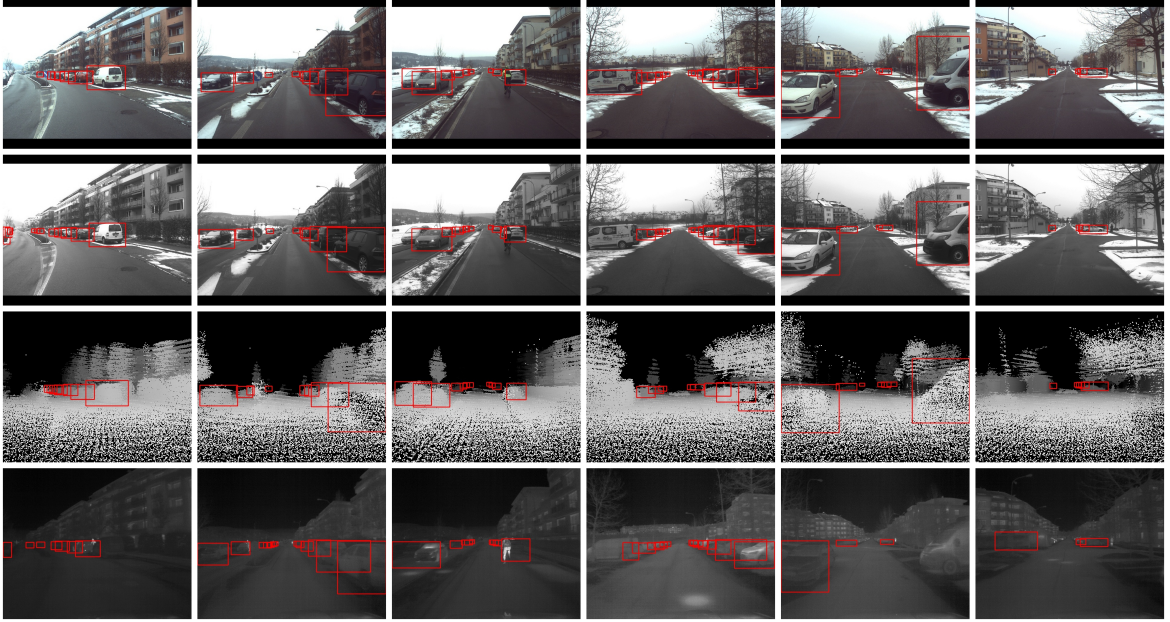
Figure 3: Validation datasets detection visualizations. By the rows (top to bottom) there are the RGB images, grayscale images, depth map images and thermal images.

Overall, there are two interesting discoveries that we did not expect before this study. First, reducing RGB information into the grayscale format did not affect object detection in any significant way. Second, the object detection on a depth map created by projecting the LiDAR data into the camera frame did way better than we originally expected. The results even outperformed the thermal camera-based object detection, see the Table 1 and detection visualization in the Figure 3.

ACKNOWLEDGEMENT

REFERENCES

[1] K. El Madawi, H. Rashed, A. El Sallab, O. Nasr, H. Kamel, and S. Yogamani. Rgb and li-dar fusion based 3d semantic segmentation for autonomous driving. In *2019 IEEE Intelligent Transportation Systems Conference (ITSC)*, pages 7–12. IEEE, 2019.

[2] R. Girshick, J. Donahue, T. Darrell, and J. Malik. Rich feature hierarchies for accurate object detection and semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 580–587, 2014.

[3] D. Guan, Y. Cao, J. Yang, Y. Cao, and M. Y. Yang. Fusion of multispectral data through illumination-aware deep neural networks for pedestrian detection. *Information Fusion*, 50:148–157, 2019.

[4] G. Jocher, A. Stoken, J. Borovec, NanoCode012, ChristopherSTAN, L. Changyu, Laughing, tkianai, A. Hogan, lorenzomammana, yxNONG, AlexWang1900, L. Diaconu, Marc, wang-haoyang0106, ml5ah, Doug, F. Ingham, Frederik, Guilhen, Hatovix, J. Poznanski, J. Fang, L. Yu, changyu98, M. Wang, N. Gupta, O. Akhtar, PetrDvoracek, and P. Rai. ultralytics/yolov5: v3.1 - Bug Fixes and Performance Improvements, Oct. 2020.

[5] A. Ligocki, A. Jelinek, and L. Zalud. Atlas fusion–modern framework for autonomous agent sensor data fusion. *arXiv preprint arXiv:2010.11991*, 2020.

[6] A. Ligocki, A. Jelinek, and L. Zalud. Brno urban dataset-the new data for self-driving agents and mapping tasks. In *2020 IEEE International Conference on Robotics and Automation (ICRA)*, pages 3284–3290. IEEE, 2020.

[7] A. Ligocki, A. Jelinek, L. Zalud, and E. Rahtu. Fully automated dcnn-based thermal images annotation using neural network pretrained on rgb data. *Sensors*, 21(4):1552, 2021.

[8] A. Mogelmose, C. Bahnsen, T. Moeslund, A. Clapés, and S. Escalera. Tri-modal person re-identification with rgb, depth and thermal features. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 301–307, 2013.

[9] T. Ophoff, K. Van Beeck, and T. Goedemé. Exploring rgb+ depth fusion for real-time object detection. *Sensors*, 19(4):866, 2019.

[10] C. Palmero, A. Clapés, C. Bahnsen, A. Møgelmose, T. B. Moeslund, and S. Escalera. Multi-modal rgb–depth–thermal human body segmentation. *International Journal of Computer Vision*, 118(2):217–239, 2016.

[11] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi. You only look once: Unified, real-time object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 779–788, 2016.

[12] J. Redmon and A. Farhadi. Yolo9000: better, faster, stronger. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 7263–7271, 2017.

[13] J. Redmon and A. Farhadi. Yolov3: An incremental improvement. *arXiv preprint arXiv:1804.02767*, 2018.

[14] S. Ren, K. He, R. Girshick, and J. Sun. Faster r-cnn: towards real-time object detection with region proposal networks. *IEEE transactions on pattern analysis and machine intelligence*, 39(6):1137–1149, 2016.

[15] R. Yadav, A. Samir, H. Rashed, S. Yogamani, and R. Dahyot. Cnn based color and thermal image fusion for object detection in automated driving.

MATLAB IMPLEMENTATION OF MULTILAYER PERCEPTRON FOR BEARING FAULTS CLASSIFICATION

Martin Doseděl

Doctoral Degree Programme (3), FEEC BUT

E-mail: xdosed04@stud.feec.vutbr.cz

Supervised by: Zdeněk Havránek

E-mail: havranek@feec.vutbr.cz

Abstract: This paper deals with implementation of multilayer perceptron neural network (NN) for bearing faults classification. Neural network has been created from scratch as an M-script with back propagation learning algorithm also, but without using advanced MATLAB packages. Public available bearing dataset from Case Western Reserve University has been used for both training and testing phase, as well as for the final classification process. Problem with sparse input data for training the network has also been addressed. This relatively simple and small neural network is capable to classify the failures of a bearing with very low error rate.

Keywords: Multilayer perceptron (MLP), deep learning, data classification, back-propagation algorithm, bearing faults

1 INTRODUCTION

Artificial intelligence (AI) is nowadays very expanding area of interest for most of the researchers as well as technicians in the industrial area. Connection of AI and technical diagnostics, especially in the predictive maintenance of machines in the industry [4], is very common and popular topic also thanks to Industry 4.0 and Internet of Things (IoT) [1]. E.g. Yin et al. [8] shows an overview of current data-based techniques for modern industrial applications with big data processing. Since the machine learning methods contain mostly statistical methods [10] like SVM, k-NN, PCA, Mahalanobis-Taguchi strategy etc., strong and powerful neural networks are trained for classification of the vibration signals from the time-domain data, such as very popular convolution neural network (CNN, e.g. [7]) and performed on dedicated powerful hardware or graphics cards. Approach presented in this article uses raw signal preprocessing. A key signal features, extracted from the time domain are used as an input of a simple, small-size, but efficient neural network of the MLP type. Aim of this work is to implement the network completely from scratch. Reason for this solution, even there exist procedures for more simple implementation, mainly using object-oriented programming ([2, 6]), was to understand the internal processes inside the network.

1.1 MULTILAYER PERCEPTRON

The idea of a technical perceptron comes from the similarity with the human's brain and neurons. Single perceptron, which is a base of the network depicted in this work, is shown in Fig. 1.

Formula for output signal considering a bias as the one input is in (1).

$$y = f\left(\sum_{i=1}^{N} w_i x_i\right) \tag{1}$$

where w_i are the individual inputs weights $w_1, w_2, ..., w_N$, x_i are the individual inputs of perceptron $x_1, x_2, ..., x_N$ and $f(\cdot)$ is an activation function. There is also a bias expressed in the equation (1) as
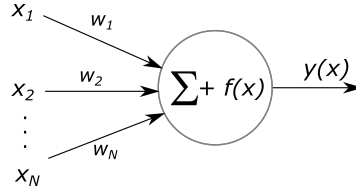
Figure 1: Single perceptron scheme.

the member x_1 and its appropriate weight w_1. It is good to mention, that in some literature, bias is written outside of the sum as a separate member, usually noted as θ. There exists a lot of activation functions [3], e.g. ReLU, binary step, linear function, TanH, Leaky ReLU, Softmax or sigmoid. Last mentioned, sigmoid function, can be expressed by formula (2)

$$f(x) = \frac{1}{1 + e^{-kx}} \tag{2}$$

where x is an input and k is a scale factor.

MLP has been used in this work as a neural network classifier. The topology is shown in Fig. 2 - one input layer (with eight inputs), one hidden layer (with sixteen neurons) and one output layer (with five outputs representing five output classes).
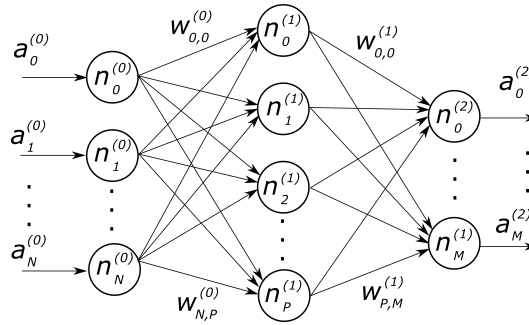
Figure 2: Multilayer perceptron NN topology.

The terminology used in Fig. 2 is as following:

$\mathbf{a}^{(0)}$ is a vector of N inputs

$\mathbf{a}^{(2)}$ is a vector of M outputs

$\mathbf{w}^{(0)}$ is a $[PxN]$ matrix of weights values between input and hidden layer

$\mathbf{w}^{(1)}$ is a $[MxP]$ matrix of weights values between hidden and output layer

N, P and M is a number of the neurons in input, hidden and output layer respectively

Since the first (input) layer does not do any computations (serves only for input signals propagation to other layers), the output of hidden layer can be simply expressed (using formula (1)) in a matrix form as:

$$\mathbf{x}^{(1)} = \mathbf{a}^{(0)} \cdot \mathbf{w}^{(0)} \tag{3}$$

and after applying the sigmoid activation function

$$\mathbf{a}^{(1)} = \frac{1}{1 + e^{-\mathbf{x}^{(1)}}} \tag{4}$$

where $\mathbf{a}^{(1)}$ is the vector of outputs from the hidden layer (or inputs to the output layer). Full matrix equation for hidden layer output can be written in the following form (see formula 5).

$$\begin{bmatrix} a_0^{(1)} \\ a_1^{(1)} \\ \vdots \\ a_M^{(1)} \end{bmatrix} = \begin{bmatrix} a_0^{(0)} \\ a_1^{(0)} \\ \vdots \\ a_N^{(0)} \end{bmatrix}^T \cdot \begin{bmatrix} w_{0,0}^{(0)} & w_{1,0}^{(0)} & \cdots & w_{M,0}^{(0)} \\ w_{0,1}^{(0)} & w_{1,1}^{(0)} & \cdots & w_{M,1}^{(0)} \\ \vdots & \vdots & \ddots & \vdots \\ w_{0,N}^{(0)} & w_{0,N}^{(0)} & \cdots & w_{M,N}^{(0)} \end{bmatrix} \tag{5}$$

The same approach can be applied for the output layer signals $\mathbf{a}^{(2)}$ equation. It is good to note here, that the matrix form of the equation is easy implementable in MATLAB, since it is the matrix-based software.

1.2 BACK PROPAGATION ALGORITHM

To train the network (to minimize the error rate during classification) it is necessary to modify the weights for each neuron's connection. The weights should be modified according to the equation (6)

$$w_j^0(t+1) = w_j^0(t) + \Delta w_j^0 \tag{6}$$

Back-propagation algorithm is very suitable for this procedure (see e.g. [3]). The goal is to minimize the error, expressed by the following equation (7) as a sum of subtracts of calculated (predicted) output values $a_j^{(2)}$ and desired (true) d_j values:

$$E = \frac{1}{2} \sum_j (a_j^{(2)} - d_j)^2 \tag{7}$$

It can be derived, that for back propagation algorithm based on gradient descend method [9] for modification of the weights between hidden and output layer it can be written:

$$\Delta w_j^0 = \frac{\partial E}{\partial w_j^0} = \frac{\partial E_C}{\partial a_j^{(2)}} \cdot \frac{\partial a_j^{(2)}}{\partial z_k} \cdot \frac{\partial z_k}{\partial w_j^0} = \ldots = (a_j^{(2)} - d_j) \cdot a_j^{(2)} (1 - a_j^{(2)}) \cdot a_j^{(0)} \cdot \alpha \tag{8}$$

where $a_j^{(2)}$ is the output of the NN, $a_j^{(0)}$ is the input of NN, d_j is desired output and α is a learning rate.

2 EXPERIMENT DESCRIPTION

2.1 INPUT DATA

As an input dataset, Case Reserve Western University [5] bearing dataset has been used. Data for damaged bearing has been used for both, the training (90 % of the total data number) and verification (remaining 10 % of compleete dataset) phase. Data represents five output categories – normal state and four degrees of the bearing's outer ring fault. Since the dataset was relatively sparse for sufficient classification accuracy (only 100 values for each category), new data has been created by adding Gaussian noise to the original data. Number of the input data (10 times higher) was then sufficient. Vibration values for the same bearing, but acquired within 3x higher load, have been used for validation of the NN.

From this extended data, a set of features has been calculated. Eight time-domain features were used – RMS value, kurtosis, skewness, variance, standard deviation, mean value and min and max value.

2.2 NEURAL NETWORK PARAMETERS

As above mentioned, MLP NN has been used for this classification task. Network has 8 inputs (for eight data features), one hidden layer with 16 neurons and one output layer containing 5 neurons (each neuron for one class). Sigmoid activation function is used and a learning rate $\alpha = 0,25$ was used. The network as well as back-propagation algorithm was implemented as a pure MATLAB m-file code.

2.3 SPARSE INPUT DATA

For achieve the best performance of the NN while using sparse input data, several measures have been done and implemented:

1. Increased number of input data by adding random values in the range of $\pm$ 5 % to the original signal. This significantly improved the number of epochs (10 times).

2. Full randomization of the input training patterns - a special algorithm implementation, which ensures that no input data from the same output class will be applied to the NN's input after each other.

3. Initialization of the weights to the value of 0,01 with additional Gausian noise with the SNR value of 20 dB.

Using these features, the accuracy of NN classification process significantly increased of about 20 %.

3 RESULTS

Minimum square error curve for training process of the resulting NN can be seen in Fig. 3 on the left side. As it can be seen, the error still does not reach its global minimum value – a trend of the curve is still decreasing. This shows to the fact, that the number of training epochs should be higher to reach higher accuracy. Learning ratio was empirically set to the optimal value of $\alpha = 0,25$. Accuracy of
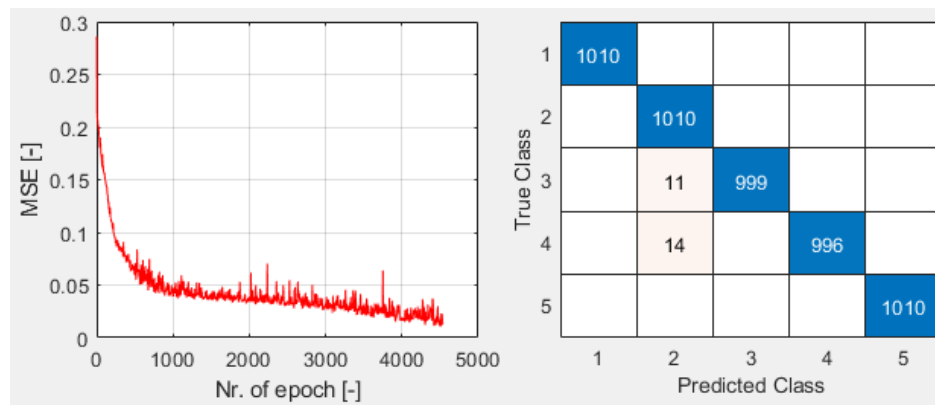

Figure 3: Mean square error during the training process (on the left) and confusion matrix of validation process (on the right).

the trained network for classification is expressed by the confusion matrix (see Fig. 3 on the right). Total accuracy was calculated to 99,5 %, what is an excellent value considering the size of a network and a size of the training group.

4   CONCLUSION

In this paper, simple MLP for bearing faults classification has been implemented in MATLAB environment. Network has been trained using sparse input data and the accuracy has been verified on the different dataset. Overall accuracy of 99,5 % is also a great result considering the fact, that training and validation processes take only 0,35 sec. or 0,05 sec. respectively. This time is with sure strongly dependent on the target hardware.

ACKNOWLEDGEMENT

REFERENCES

[1] Chen, A.; Liu, F. H.; Wang, S. D.: Data reduction for real-time bridge vibration data on edge. In *Proc. - 2019 IEEE Int. Conf. Data Sci. Adv. Anal. DSAA 2019*, Institute of Electrical and Electronics Engineers Inc., oct 2019, ISBN 9781728144931, p. 602–603, doi: 10.1109/DSAA.2019.00077.

[2] Elnozahy, A.; Sayed, K.; Bahyeldin, M.: Artificial Neural Network Based Fault Classification and Location for Transmission Lines. In *2019 IEEE Conf. Power Electron. Renew. Energy*, IEEE, oct 2019, ISBN 978-1-7281-0910-7, p. 140–144, doi: 10.1109/CPERE45374.2019.8980173.
URL https://ieeexplore.ieee.org/document/8980173/

[3] Kriesel, D.: *A Brief Introduction to Neural Networks. dkriesel. com.* 2011, 1–29 p.
URL http://www.dkriesel.com/en/science/neural_networks

[4] Liu, R.; Yang, B.; Zio, E.; et al.: Artificial intelligence for fault diagnosis of rotating machinery: A review. *Mech. Syst. Signal Process.*, vol. 108, aug 2018: p. 33–47, ISSN 10961216, doi:10.1016/j.ymssp.2018.02.016.
URL https://linkinghub.elsevier.com/retrieve/pii/S0888327018300748

[5] Loparo, K. A.: Case Western Reserve University Bearing Data Center. 2012.
URL http://csegroups.case.edu/bearingdatacenter

[6] Mathworks: Deep Learning Toolbox - MATLAB.
URL https://in.mathworks.com/products/deep-learning.html

[7] Qiao, H.; Wang, T.; Wang, P.; et al.: An Adaptive Weighted Multiscale Convolutional Neural Network for Rotating Machinery Fault Diagnosis under Variable Operating Conditions. *IEEE Access*, aug 2019: p. 118954–118964, ISSN 21693536, doi:10.1109/ACCESS.2019.2936625.

[8] Yin, S.; Li, X.; Gao, H.; a kol.: Data-based techniques focused on modern industry: An overview. *IEEE Trans. Ind. Electron.*, vol. 62, ǐssue. 1, 2015: s. 657–667, ISSN 02780046, doi:10.1109/TIE.2014.2308133.
URL https://ieeexplore.ieee.org/document/6748057/

[9] Yu, X.; Efe, M. O.; Kaynak, O.: A general backpropagation algorithm for feedforward neural networks learning. jan 2002, doi:10.1109/72.977323.

[10] Zhang, S.; Zhang, S.; Wang, B.; a kol.: Deep Learning Algorithms for Bearing Fault Diagnosticsx - A Comprehensive Review. *IEEE Access*, vol. 8, jan 2020: p. 29857–29881, ISSN 21693536, doi:10.1109/ACCESS.2020.2972859, 1901.08247.
URL http://arxiv.org/abs/1901.08247

FIBER-COUPLED FABRY-PÉROT INTEROMETRIC SENSOR: ANALYSIS AND MODELLING

Michal Skalský

Doctoral Degree Programme (5), FEEC BUT

E-mail: xskals01@stud.feec.vutbr.cz

Supervised by: Zdeněk Havránek

E-mail: havranek@feec.vutbr.cz

Abstract: This paper presents a theoretical analysis of the Fabry-Pérot interferometric sensor utilizing an extrinsic fiber-optic cavity. Its behavior is modelled and simulated for various conditions, showing the key properties of the different sensor's setups. The simple single-mode sensing tip and the GRIN collimator are compared. Further, new kind of sensing head is proposed exhibiting an inverse reflectivity and enabling a theoretical resolution less than 1 pm.

Keywords: Fiber-optic sensors, Fabry-Pérot, interferometry, displacement sensor, Faraday-rotator.

1 INTRODUCTION

Among various optical sensing principles, the Fabry-Pérot interferometer (FPI) formed by an optical fiber tip and a reflective surface represents an ultimate solution in terms of sensitivity, resolution, and also compactness. Therefore, it has been exploited in many applications. The best-known is the atomic force microscopy [1], where a noise density of several units of $\mathrm{fm}/\sqrt{\mathrm{Hz}}$ was achieved [2]. Nevertheless, the FPI can be used to sense basically any physical quantity which can be transduced to displacement, refractive index, or wavelength. It is thus beeing utilized as vibration sensors, accelerometers, microphones, material properties sensors, etc [3]. Also, its manufacturing process may be very easy since the sensitive part consists only of cleaved fiber tip in the simplest case.

On the other hand, the development of such sensor itself is much less straightforward if particular parameters, such as range and sensitivity, are targeted since the actual behavior of the FPI is strongly dependent on material properties and cavity scales. Therefore, it is very difficult to develop and optimize such sensor unless we perform relevant calculations. From this point of view, simulations play a significant role for basic design and optimizing tasks. There has been proposed several analyses and models involving different physical aspects; one of he most comprehensive descriptions provide Wilkinson [4] and Kilic [5].

Compared to the ideal FPI with parallel plane mirrors, the proper characterization of the extrinsic fiber FPI is more complex since the beam divergency and fiber coupling must be taken into account. In this paper, the behavior of fiber FPI is simulated under different conditions, revealing some notable properties of the sensing capabilities. Further, there are also show some modifications of the fiber FPI, simulated by extending the model so their key advantages can be demonstrated.

2 MODELING OF FPI PROPERTIES

Likewise in any FPI, the fiber-optic arrangements is formed by resonance cavity passed by light beam and allowing its reflections with partial coupling outside the cavity (Fig. 1). In such a way, it can act as a periodic wavelength filter whose shape may be changed by the cavity size or other parameters. The filter selectivity increases with the number of reflections inside a cavity and is referred as finesse

of FPI. Each surface has some reflection and transmission; in case of fiber-air boundary they can differ for each direction. They basically determine the portion of light coupled to the cavity or reflected back inside the fiber. With only one passing inside the cavity ($p = 1$), we have two-beams' interference yielding an output intensity response with a sine shape (low finesse). Conversely, by allowing multiple light passing through the cavity ($p > 1$) we achieve sharpening the response peaks in the output intensity and increasing the sensitivity (Fig. 2). The fiber FPI has typically the reflective response since the zero-order reflection ($p = 0$) interferes with the beams passing the cavity ($p > 0$).
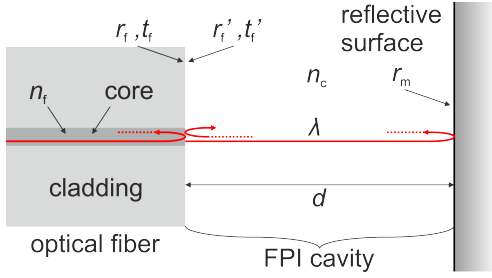
Figure 1: Fiber-optic FPI cavity.
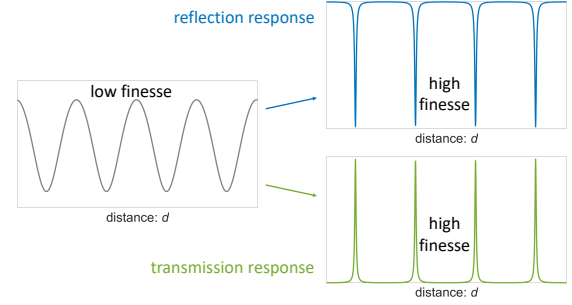
Figure 2: Typical FPI intensity response.

The reflected intensity I_{out} is a linear function of the input intensity I_{in} so we can express it in terms of reflectivity as $R_{\text{FPI}} = I_{\text{out}}/I_{\text{in}}$. However, since the light is a wave, rather than intensity we assume it in the form electric field E, having a magnitude and phase, and lowercase typing of reflection and transmission coefficients, as shown in Fig. 1. The relationship is then $R_{\text{FPI}} = |r_{\text{FPI}}|^2$ where

$$r_{\text{FPI}} = r_{\text{f}} + t_{\text{f}}t_{\text{f}}' \sum_{p=1}^{\infty} r_{\text{m}}^p r_{\text{f}}'^{p-1} e^{-j2kdp}, \tag{1}$$

where k is the wave number. The coefficients r and t are complex in general, reflecting phase shift caused by transmission and reflection. This approximation is valid for the optical fiber assuming much shorter cavity compared to beam mode field diameter (MFD). In case of common single-mode fibers, having MFD of several μm, thus cannot provide sufficient accuracy and more complex approach must be used.

The major problem is the beam divergence causing decreasing the coupling efficiency with every cavity pass, which affects both amplitude and phase of light portion coupled back to the fiber. Further, the model should also incorporate a possible misalignment α between reflective surface plane and fiber cleave plane to determine manufacturing tolerances. Based on the mathematical descriptions in [4, 5], the Matlab model of fiber FPI was created. Compared to [4] where the cavity passing is limited to $p = 5$, here the number of passing beams is only thresholded by carried energy (which can be arbitrarily small) and especially high-finesse FPI behavior can be thus approximated with higher accuracy.

In the following sections, the different structures of fiber FPI sensing head will be shown and compared to the simple single-mode fiber tip. The overview of the structures can be found in Fig. 3. Note that all simulations are performed for air-filled cavity, SMF-28 fiber and wavelength $\lambda = 1550$ nm.

2.1 SINGLE-MODE FIBER (SMF) CAVITY

The simplest case of fiber FPI formed by cleaved single-mode fiber (SMF) is shown in Fig. 3a. The single-mode optical beam can be well approximated by Gaussian beam whose divergence angle θ outside the fiber is inversely proportional to the beam input waist, determined by the MFD. Assuming
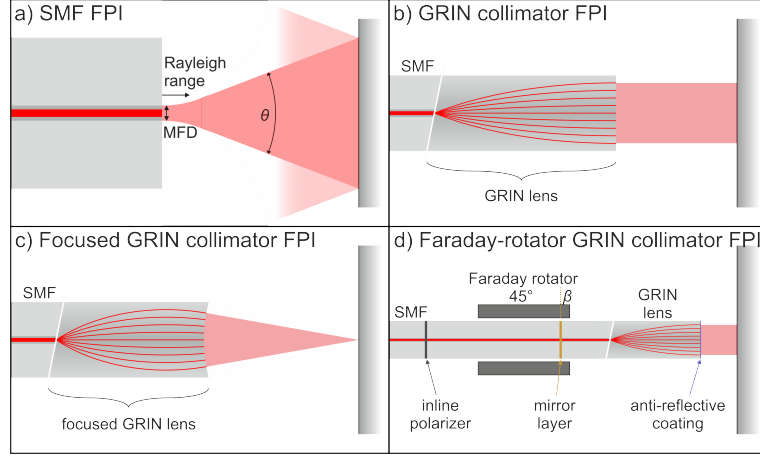
Figure 3: The overview of analyzed fiber FPI sensing structures: a) SMF, b) GRIN collimator, c) Focused GRIN collimator, d) Faraday-rotator GRIN collimator.

a standard telecom SMF-28 and wavelength 1550 nm, the angle corresponds to $11°$. The interesting fact is that the beam diverges as late as it reaches certain distance behind the fiber, until which is approximately collimated. This distance is referred as Rayleigh range and equals to 55 μm in the studied case. The Fig. 4 compares two modeled responses of SMF FPI: with no coating and a dielectric mirror having $R_m = 0.99$, and dielectric fiber coating $R_f = 0.5$ and a metallic mirror with $R_m = 0.95$, both with perfect fiber alignment $\alpha = 0°$. When using FPI as a distance sensor, we usually exploit a small portion of $R_{FPI}(d)$ between positive and negative peaks, which has positive or negative slope. Therefore, more than total reflected power we need to care about an efficient utilization of the detector dynamic range, i.e. minimizing the offset and maximizing the signal component, which is expressed by the visibility. In other cases, rather than SNR and measurement range, the peak sensitivity may be of higher importance, i.e. the slope of R_{FPI} change. The Fig. 4 also shows normalized slopes (scaled on right vertical axis) reflecting the optimal usage of the detector range. The second case exhibits much higher sensitivity due to multiple reflections within the Rayleigh range, which is in accordace with the expectations. Note the positive and negative slopes difference due to Guoy phase shift caused by the beam divergency. Further, the resonant distances are shifted in Fig. 4b with respect to Fig. 4a thanks to metallic mirror absorbing a part of energy (see detailed view).
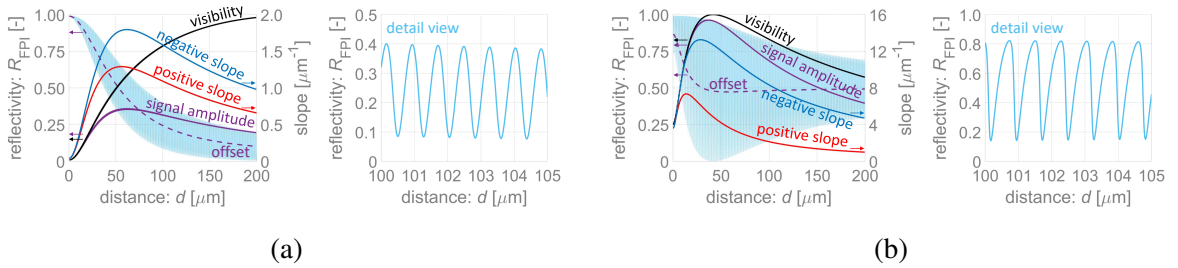

Figure 4: Modelled response of SMF FPI: a) no fiber coating ($R_f = 0.04$), dielectric mirror ($R_m = 0.99$); b) dielectric fiber coating ($R_f = 0.5$), metallic mirror ($R_m = 0.95$).

2.2 GRADED-INDEX (GRIN) COLLIMATOR

The drawback of the SMF FPI is decreasing coupling efficiency due to beam divergence and thus impossibility to reach high finesse since the intensity of the beam inside the cavity decreases rapidly with p. This can be overcome with a graded-index (GRIN) collimator at the fiber tip, as shown in
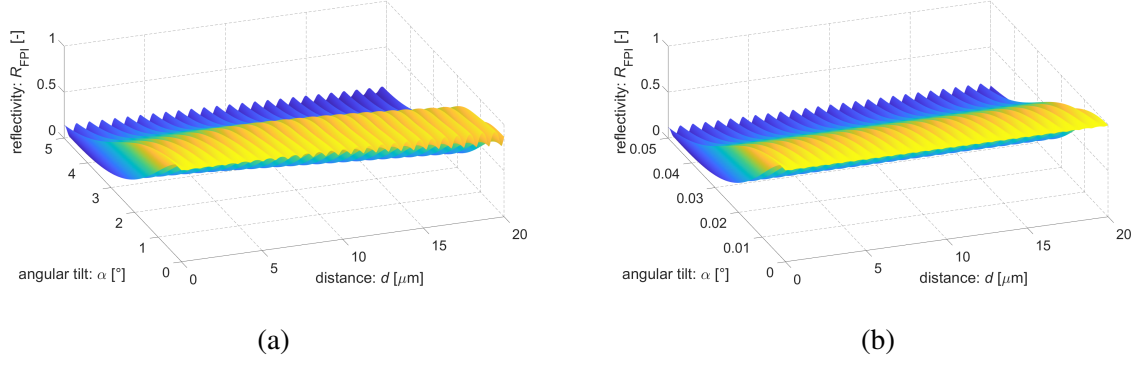
Figure 5: FPI reflectivity ($R_\mathrm{f} = 0.04$, $R_\mathrm{m} = 0.99$): a) simple SMF, b) GRIN collimator.

Fig. 3b. The gradient index fiber works as a lens which substantially extends the Rayleigh range of the output beam. The comparison of the simple SMF and GRIN collimator FPI response is shown in Fig. 5. The modified model was obtained by simple re-calculation of the Gaussian beam according to GRIN collimator performances.

The output reflectivity is here modeled in dependence on cavity size d and also misalignment angle α. Note that the GRIN collimator causes the FPI operates equally within much broader distance range. Conversely, SMF is much less sensitive to angle tilt due to broader beam cone.

In some cases, usage of the GRIN collimator may be limiting due to its bigger spot, which can even exceed 1 mm. This issue can be solved by focused GRIN collimator, shown in Fig. 3c. The simplest implementation may be performed by extending the collimator to form a convergence beam. However, to perform well for $p > 1$, the sensing had needs to have concave shape so the reflected beam did not diverge outside the cavity.

2.3 FARADAY-ROTATOR SENSING HEAD

The fiber FPIs inherently exhibit a reflection response (Fig. 2) which leads to small signal to DC bias contrast (see Fig. 5) if the reflectivities are not balanced. The transmission response does not contain the zero-order reflection and is thus unbiased. In fiber FPI we cannot couple the light from the opposite side of cavity; though, the problem may be overcome by special setup with Faraday polarization rotator (FR), whose structure is in Fig. 3d. Note the FPI cavity is not formed solely by the air gap, but comprises also the GRIN collimator with anti-reflection coating and the fiber portion behind the semitransparent mirror. The nonreciprocal rotation by $45°$ combined with polarizer at
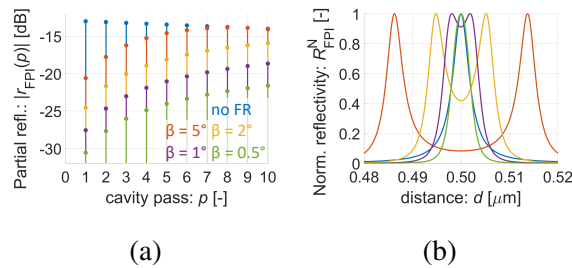

Figure 6: FR FPI head with different rotation β ($R_\mathrm{f} = 0.95$, $R_\mathrm{m} = 1$): a) Back-coupled $|E|$ for separate pass order p, b) Normalized reflectivity peaks (detail).
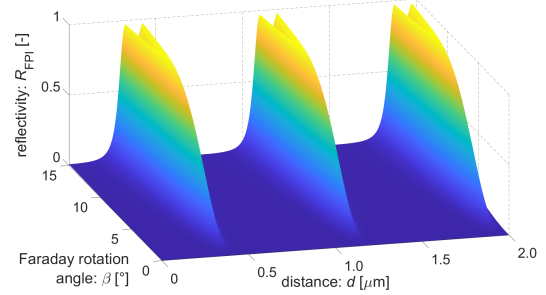
Figure 7: FR FPI reflectivity with β ($R_\mathrm{f} = 0.5$, $R_\mathrm{m} = 0.99$, $h = 10$ mm).

the input eliminates the zero-order reflection. A small rotation of angle β also exist in the cavity causing filtration of p-passed beam by factor $\sin(2\pi\beta)$. The advantage of this filtration also consists in suppressing the low-p reflections and balancing the intensity of high-p beams coupled back to the guiding fiber and thus increasing the sensor's finesse. Fig. 6a shows the theoretical portions of $|E|$ coupled back to the guiding fiber displaying the suppression of the zero-order reflection and balancing the higher-p ones by different angle β. Due to eliminating the low-p beams coupling, we can see the resonance peaks are sharper and with better linearity within the full dynamic range (Fig. 6b). The response of FR FPI with different rotation angle β is shown in Fig. 7. The distance d is the air cavity size, the sensing head length behind the mirror is assumed $h = 10$ mm.

3 CONCLUSION

Based on the mathematical description of fiber FPI, the Matlab model was created to simulate various behavior of different structures of fiber-coupled cavities. Firstly, it was shown at simple SMF FPI that the sensitivity can be greatly increased by fiber coating, albeit the distance d must be maintained small. With GRIN collimator, we can extend it notably, however, the alignment is much more crucial. Nevertheless, we can reach extreme sensitivity with additional coating in that case or modify the GRIN lens to be used as very precise point sensor. Lastly, a novel solution of sensing head was proposed, using a Faraday rotator to achieve an inverse response, yielding maximizing the FPI performance yet more. With the moderate parameters ($\lambda = 1550$ nm, $R_{\mathrm{f}} = 0.95$, $\beta = 1°$), a relative intensity can reach a peak change of 0.377 nm^{-1}. When 12-bit resolution is available, this yields 0.65 pm is the smallest distinguishable distance change. Therefore, such kind of sensor has a high potential for the most accurate measurement applications.

REFERENCES

[1] RUGAR, D., H. J. MAMIN and P. GUETHNER. Improved fiber-optic interferometer for atomic force microscopy. *Applied Physics Letters*. American Institute of Physics, 1989, **55**(25), 2588-2590. DOI: 10.1063/1.101987.

[2] RASOOL, Haider I., Paul R. WILKINSON, Adam Z. STIEG and James K. GIMZEWSKI. A low noise all-fiber interferometer for high resolution frequency modulated atomic force microscopy imaging in liquids. *Review of Scientific Instruments*. 2010, **81**(2), 023703-1-10. DOI: 10.1063/1.3297901.

[3] HOOGENBOOM, B. W., P. L. T. FREDERIX, J. L. YANG, et al. A Fabry–Perot interferometer for micrometer-sized cantilevers. *Applied Physics Letters*. 2005, **86**(7), 074101-1-3. DOI: 10.1063/1.1866229.

[4] WILKINSON, Paul R. and Jon R. PRATT. Analytical model for low finesse, external cavity, fiber Fabry–Perot interferometers including multiple reflections and angular misalignment. *Applied Optics*. 2011, **50**(23), 4671-4680. DOI: 10.1364/AO.50.004671.

[5] KILIC, Onur, Michel J. F. DIGONNET, Gordon S. KINO and Olav SOLGAARD. Asymmetrical Spectral Response in Fiber Fabry-Pérot Interferometers. *Lightwave Technology, Journal of*. USA: IEEE, 0912, **27**(24), 5648-5656. DOI: 10.1109/JLT.2009.2032135.

TRIBOELECTRIC SENSING SYSTEMS

David Riha

Doctoral Degree Programme (2nd year), FEEC BUT

E-mail: xrihad01@stud.feec.vutbr.cz

Supervised by: Pavel Tofel

E-mail: tofel@feec.vutbr.cz

Abstract: Triboelectric phenomenon, historically, is one of many problems which engineers wanted to be avoid. However nowadays this phenomenon seems to be useful source of electrical signal for displacement measurement. In this work all sensing elements are made from pure PVDF by electrospinning manufacturing method. Nano fibres, which are result of the process are placed to aluminium foil naturally without mechanical rectifying. Used material was tested only for triboelectric charge generation, although piezoelectric phenomenon was expected. We were not able to quantify it by standard methods, probably due to chaotic structure and fabric behaviour of the final product. Triboelectric phenomenon is possible to measure directly, and systems are more predictable. Finally, two modes were tested, and potential advantages and disadvantages were qualified.

Keywords: EEICT, triboelectric, sensing, PVDF

1   INTRODUCTION

Triboelectric systems are very popular for energy harvesting purposes. It opens door for investigation of sensing possibilities in wide industrial branch and also for all sensing of natural process where any movement as acceleration are possible to measure. Triboelectric phenomenon could be used also for sensing of position. There are many applications: self-powered human motion sensors [1], self-powered automobile sensors [2]. Example of sensing is possible to see from [3]. Energy harvesting describe flexible Nano generators [4], internet of things [5], environmental monitoring systems [6]. There are also many other applications in biomedicine as non-invasive biomedical monitoring systems [7]. Our research is based on triboelectric active material where big advantage is large surface which is made by electrospinning method. We started with PVDF material, which could be promising also for piezoelectric potential.

2   MANUFACTURING OF ACTIVE MATERIAL

2.1   PVDF

Polyvinylidenfluorid material was made by electrospinning method. This procedure is very common for manufacturing of very fine non-woven fabric. Product of used machine is aluminium foil with dimensions 200x270 mm. We are able to change thickness of final product simply by change of manufacturing time. We used spined material on aluminium foil without removing it from substrate for all sensing elements, because it is very problematic procedure which ideally need very clean environment. We prepared material with compressed thickness from 20 to 100 µm. Fibres have diameter from 600 to 1200 nm, and it could be partially controlled by electric field between needle and substrate and partially also by chemical compounds of input solution.

2.2   SOLUTION

We used PVDF with molar weigh 275,000 g/mod (Sigma Aldrich, St. Louis, MO, USA). As solvents we used dimethylsulfoxide p.a. (DMSO, Sigma Aldrich, St. Louis, MO, USA) and acetone

(Ac, Sigma Aldrich, St. Louis, MO, USA). Solutions were electrostatically spined by machine 4spin (Contipro a.s., Dolni Dobrouc, Czech Republic), see **Figure 1**.

Figure 1: Electrostatic spin device 4spin

Fibres are possible to rectifier to one direction, or it is possible to place them random as non-woven fabric. We used non-woven example, see **Figure 2** for SEM image. We expect isotropic response of triboelectric phenomena from this arrangement for our experiment.
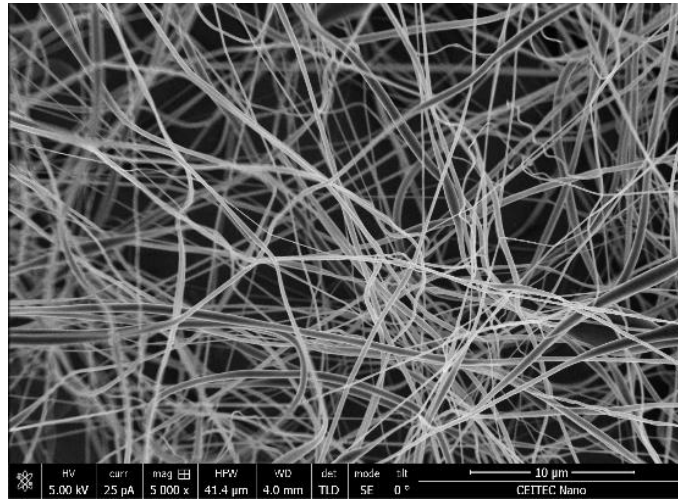

Figure 2: Non-woven fabric SEM image.

3   SENSING SYSTEM

3.1   FLEXIBLE ELECTRODE

We wanted to achieve permanent contact of electrode with PVDF material. The goal was to increase sensitivity of the system and provide good linear response to linear sinus displacement. Sensing element shown on the **Figure 3** is only part of future sensor, is necessary to say that sensing flexible membrane (ground electrode) is very sensitive on shape of exciting tool which is crucial for achieving acceptable distortion factor. We expect that sensing area don't exceed 100 mm^2 area.
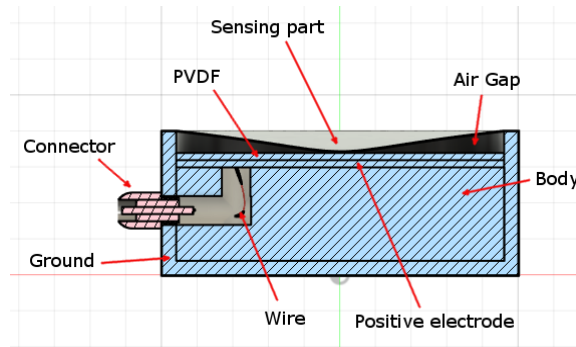
Figure 3: Principle of function of the sensing element "Flexible electrode"

3.2 EXPERIMENT

Original request of the device (sensing element) was designed for impulse input. We tested this sensing element by impulse excitation. Impulse is 5ms wide, repeating frequency was 1 Hz. Maximal applied displacement was 1.2 mm.

Sensitivity (Change of charge vs change of displacement) was measured by electrometer (charge amplifier) Keithley 6517b (Keithley, USA). Mechanical displacement was controlled by vibration testing system TV 50018 (Tira, Germany). Displacement was measured by interferometer ILD 1402-10 (Micro Epsylon, Germany). It is possible to see on **Figure 4**. Capacity in not loaded conditions was 38 pF.
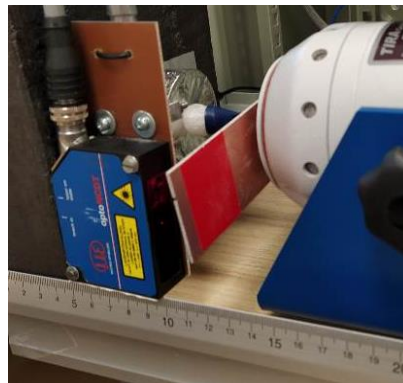

Figure 4: Testing apparatuses of the "Flexible electrode" sensing system

Is possible to see that sensing element indicate very promising sensitivity. For details see **Figure 5** a) which show response in time domain and **Figure 5** b) which report charge response for different amplitudes of impulses applied. Graph **Figure 5** b) shows good linearity of output charge with increasing displacement of impacts.
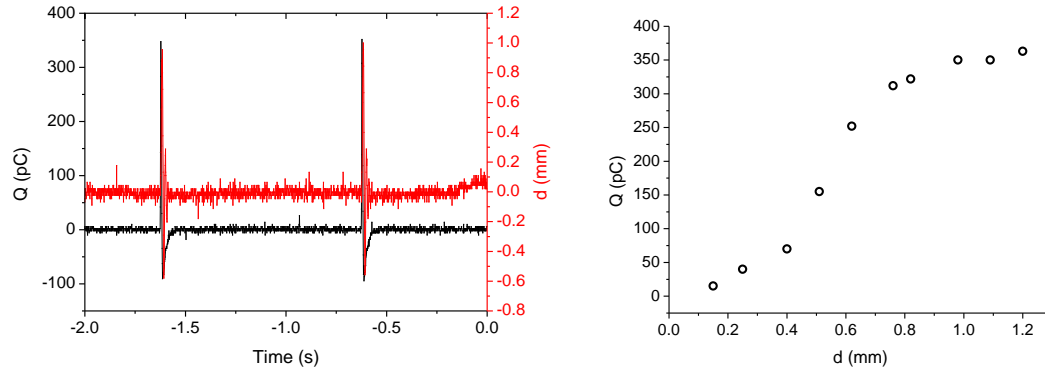
Figure 5: charge response a) in time domain and b) charge response for different amplitude of impulse

The sensing element was exposed to different repeat frequency of impulses, for detail see **Figure 6**. There is possible to see constant response with different frequency applied. This stability is very welcomed feature for future applications for sensing purposes.
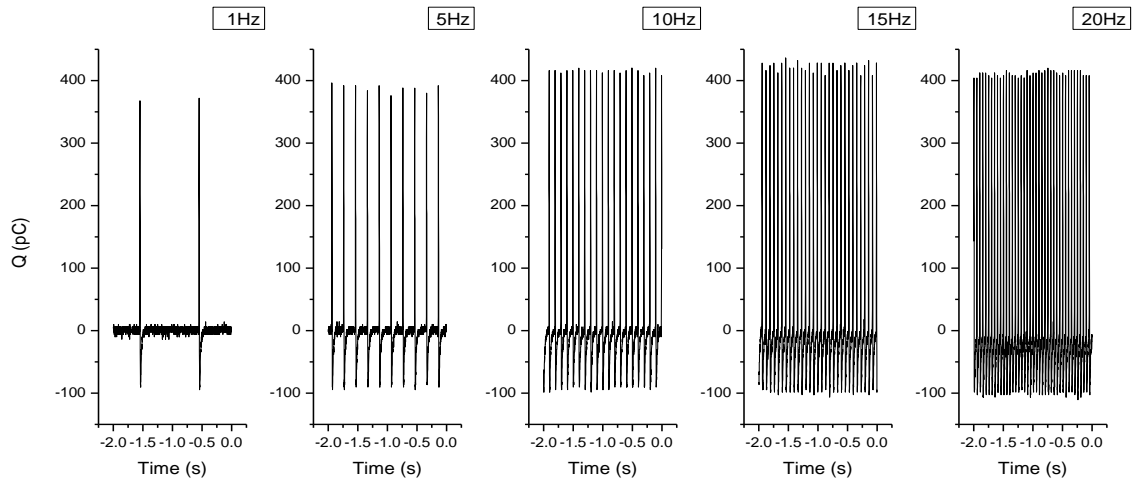

Figure 6: Charge response for different repetition frequency

Sensing element didn´t fulfil request for response to slow linear sinus excitation for very low frequencies, it is possible to see on **Figure 7** a). The reason is probably in not guaranteed contact of electrode to PVDF all time of working process, this is space for improve of the design. Higher frequencies don't show such big problems as is possible to see on **Figure 7** b), but there is still possible to see small distortion of output signal.
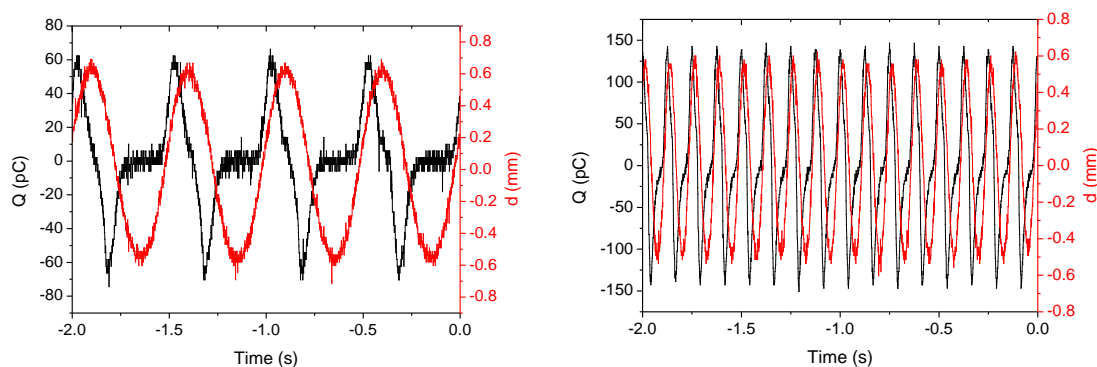
Figure 7: Charge response to linear sinus excitation for frequency: a) 2 Hz and b) 8 Hz

4   CONCLUSION

The "Flexible electrode" sensing system seems very promising due to high sensitivity and finally the element could be tuned for achieving also requested linearity. Original idea is to use vacuum between PVDF fabric and Flexible membrane, this could guarantee permanent contact of electrode which is crucial for linear response and second welcomed benefit could be increasing of sensitivity due missing atmosphere.

System is suitable for impact measurement. Sensing system is very sensitive, and it could be suitable for pressure change monitoring. The system in bigger configuration could be possible to use for self-power applications as energy harvesting unit due to high sensitivity. This kind of design has also potential to be very thin and could be implemented in many applications.

Advantage of this sensing system compared to conventional piezoceramics is friendly, low-cost economic manufacturing process without lead and other toxic substances.

REFERENCES

[1]   P. Zhang, Z. Zhang, J. Cai, Microsyst. Technol. 2021, 5, DOI 10.1007/s00542-020-05199-5.

[2]   D. Heo, J. Chung, B. Kim, H. Yong, G. Shin, J. W. Cho, D. Kim, S. Lee, Nano Energy 2020, 72, 104719.

[3]   T. Chen, Q. Shi, K. Li, Z. Yang, H. Liu, L. Sun, J. A. Dziuban, C. Lee, Nanomaterials 2018, 8, 1.

[4]   A. Ibrahim, A. Ramini, S. Towfighian, Energy Reports 2020, 6, 2490.

[5]   H. Elahi, K. Munir, M. Eugeni, S. Atek, P. Gaudenzi, Energies 2020, 13, 1.

[6]   H. Chen, C. Xing, Y. Li, J. Wang, Y. Xu, Sustain. Energy Fuels 2020, 4, 1063.

[7]   B. Ghatak, S. Banerjee, S. B. Ali, R. Bandyopadhyay, N. Das, D. Mandal, B. Tudu, Nano Energy 2021, 79, 105387.

VERIFICATION OF MODELING APPROACH BY USING MODEL OF CALCULABLE CAPACITOR AND DETERMINATION OF ITS ACCURACY

Vlastimil Mancl

Doctoral Degree Programme 2, FEEC BUT

E-mail: xmancl00@stud.feec.vutbr.cz

Supervised by: Petr Beneš

E-mail: benesp@feec.vutbr.cz

Abstract: During the creation of any sort of model the inevitable question about the model correctness and accuracy, should always be answered. This paper proposes a simple method for modelers to prove their approach to modeling and to determine the accuracy of a simulation. The method consists of creating a model of calculable capacitor based on Thompson-Lampard theorem in Ansys Electronics desktop using Maxwell 3D electronics desktop. The model and its parts are described. The expected results are discussed and compared to data obtained from simulation. Problematic of rounding edges and its influence on simulation accuracy and time is also discussed and shown on two models. A part of the paper is a description of how to interpret the Ansys percentage error.

Keywords: Ansys Electronics Desktop, Calculable capacitor, Fringe effect, Maxwell 2D and 3D, Model accuracy, Simulation, Thompson-Lampard Theorem

1   INTRODUCTION

There are several methods to prove the correctness of a simulation, but most of them are based on a fact, that a creator has a great knowledge about the simulated model and the simulation itself and thus, has no doubt about the fundamentals. The method proposed by this paper aims mainly to beginners or intermediate modelers, who are learning Ansys Maxwell 2D or 3D environment with electrostatics solver. The main advantage of this method is the fact, that it does not require deep knowledge about the electrostatics. So even a learning modeler can, trough this method, prove his modeling approach.

The proposed method is to find a model, which has precise analytic solution, therefore the modeler knows the results beforehand, which makes evaluation more comprehensive and understandable. This paper shows, how to evaluate simulation and how to interpret Ansys errors values and convergence data.

The main motivation for this paper was, for the sake of future research, to prove author's own simulations and approaches.

2   CHOOSING THE MODEL

There are not many calculable capacitors. Most of the analytic solutions are only an approximation. For example, the best known equation (equation 1) for parallel plate capacitor, represents only a simplified solution, that does not include the fringing effect. The greater the distance between the plates, the greater the deviation from reality.

$$C = \varepsilon_0 \cdot \varepsilon_r \cdot \frac{s}{d} \ [F] \tag{1}$$

There is currently no precise analytic solution for a parallel plate capacitor, so it is not suitable for the confirmation of the simulation results. There are a few configurations which have precise analytic solution such as two spheres[1], but such a model is extremely simple and does not test the ability of a model designer to create a model of intermediate level atleast. The solution is to choose a capacitor with electrode configuration, which has precise analytic solution and which is not overly simplified. The calculable capacitor based on Thompson Lampard theorem meets those conditions.
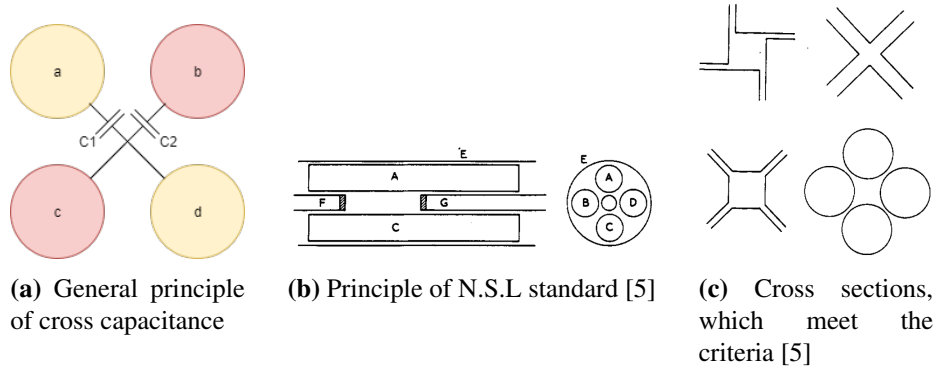

(a) General principle of cross capacitance

(b) Principle of N.S.L standard [5]

(c) Cross sections, which meet the criteria [5]

Figure 1: Principle of theoretical cross capacitor and N.S.L version of cross capacitor.

3   THOMPSON-LAMPARD THEOREM

In 1956 the Thompson, A. M., and Lampard, D. G. published their theorem based on the cross capacitance and its usage in calculable standards. [2] The theorem was an attempt to reduce variables which influenced countable standards. Until then, the countable capacity standards were scarcely used due to their inconsistency. As mentioned before, the problem was that the former standards provided accurate results only if almost ideal conditions were met. Even a slight deviation from those conditions made the results unusable. [3] One of such standard was based on capacitor with parallel plates, whose problems were described in section 2.

This problem was eliminated with a capacitor based on cross capacitance design. Its design consists of four rods/electrodes of arbitrary cross section and with infinite length. The configuration can be seen in figure 1a. If the C_1 equals C_2 then the mean of those capacitances $\overline{C}$ is the same. Lampard has determined that an arbitrary cross capacitor is described by the following equation 2.[4]

$$\exp(-4\pi C_1) + \exp(-4\pi^2 C_2) = 1 \tag{2}$$

So if the C_1 equals C_2 then:

$$C_1 = C_2 = \overline{C} = \frac{\ln 2}{4\pi^2} \ [\text{e.s.u.} \cdot \text{cm}^{-1}] = 1.953548 \ [\text{pF} \cdot \text{m}^{-1}] \tag{3}$$

But for that to be true, two criteria have to be met. Firstly, the electrodes have to be placed in such a manner, that the neighboring electrodes are almost in contact and create a gap in the middle. Secondly, the cross section of the gap has to have one axis of symmetry, which connects two opposite spaces between electrodes. The examples of such cross configurations are shown in figure 1c. Consequence of such theorem is that the shape of electrodes does not matter as long as the cross section of a gap is symmetrical. Or that size of cross section, either of rod or gap, does not matter as long as the insulating space between neighboring rods is considerably small. The most important advantage is that with this configuration only one dimension, the length, needs to be observed to determine the capacitance.

4   REAL CALCULABLE CAPACITOR

Thompson discussed the physical realization of a countable capacitor[5], which resulted into a design of a real calculable capacitor. The theoretical capacitor has infinite length, and is therefore unachievable. With finite length, the fringing effect takes place at each end of the electrodes. Thompson came up with a few designs using shielding technics, which solved the fringe issue. [4]

That design was adopted and modified by many institutions. For purposes of this paper, the model from The National Standards Laboratory (N.S.L), which was renamed to National Measurement Laboratory in 1974, was chosen. The simplified model was used as a template. The real calculable capacitor is much more complicated, because unlike the simulation, the rods have to be somehow attached to the construction. The principle and design of this model can be seen in the figure 1b.

The design consists of one hollow cylinder with two holes on its ends for driving rods. The outer cylinder has shielding function. Inside the cylinder, 6 rods are placed. Two of them (A,C) have shielding function. The other two (B,D) serve as positive and negative electrodes. The last two rods (F,G) are the shielding driving electrodes. They serve to determine the effective length used for determining the capacity. For that purpose, the driving rods are movable. The precision of such calculable standard is dependent on an accuracy of the position measurement. So if a precise measurement principle, such us Fabry-Perot interferometer, is used, then the results can have precision as much as 0.1 ppm. [6]

5   SIMULATION

The main model was made according to previous description. The dimensions are unimportant, because of the mentioned fact, that the shape is more crucial than dimensions, but the length of the outer cylinder should be sufficiently larger than the maximal effective length. Gaps could be tighter, but as of now, they are sufficient. The resulting design can be seen in figure 2a. The materials of electrodes and shield were defined as a perfect conductor and the area was defined as a vacuum. The solution settings were as follows: Maximum number of passes 20, percent error 0.05, refinement per pass 30%. Parametric simulation was performed with parameter l in range <10;200> mm with a step of 10 mm and one simulation with $l = 5$ mm, for acquiring detail at short effective length. This parameter represents the effective length . The parameter was also bound to driving electrodes, thus creating a movement in an opposite direction of each other, so the effective area enlarges. The model was labeled as NR model (NR - Not Rounded).

Another model (R model (Rounded)) was made, but all the sharp corners were rounded. This was done because the sharp corners can cause convergence problem.[7] The second simulation was performed to find out if this models would cause mentioned problem. The settings remained the same.
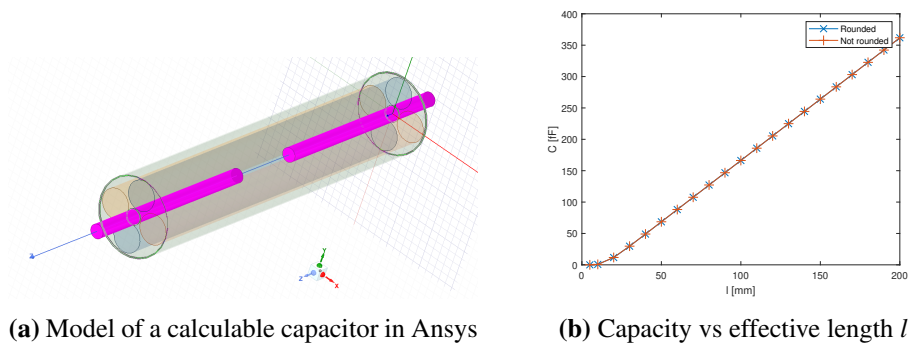

(a) Model of a calculable capacitor in Ansys   **(b)** Capacity vs effective length l

Figure 2: A model created in Ansys and dependence of capacity on effective length l

The driving electrodes, the outer shell and the blue electrodes were excitated with 0 V for simulating the ground shielding. The orange electrodes were excitated with -10 V and 10 V respectively.

6 RESULTS

In theory, the dependence of capacity on effective length should be linear. In reality, the dependence is exponential at the short lengths, due to the fringe effect. The fringe effect makes the electric field non-uniform around the end of the driving rods, but only to certain distance. Beyond said distance, a uniform electric field is generated. The uniformity ensures the stable increment of capacity. This phenomenon is in accordance with W. K. Clothier's research [5]. In the figure 2b, the exponential growth at the short lengths can be seen. Beyond 40 cm the growth is linear and is roughly 19.5 fF per cm. In the same figure, it seems like, that the model with rounded edges gives exactly the same results as the model without the rounded edges.
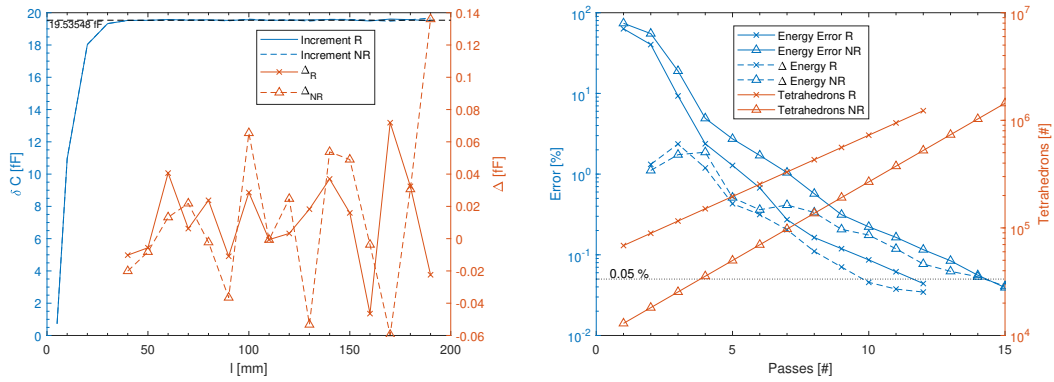
(a) Progress of increment in dependence on l with deviation.

(b) Convergence data for $l=100$

Figure 3: Data obtained by simulation.

7 ACCURACY AND CONVERGENCE

The advantage of used models is the known analytical solution, which is precise even in real environment. That fact presents the possibility to determine the accuracy of a model and to help comprehend simulation settings or errors. Both models converged with total error under the 0.05 criteria. To a direct comparison of both models, the figure 3a has to be observed. There, the increment of capacity for both models and absolute deviations Δ from the theoretical increment, can be seen. The deviations for $l \in \langle 0; 40 \rangle$ were not counted due to aforementioned phenomenon. The deviation with NR model was mostly larger, hence the NR model had overall inferior accuracy. But even in the worst case, the relative deviation is under 1%. The average deviation is 0.0234 pF for R model and 0.0362 pF for NR model. The 1% simulation accuracy could indicate that the 0.05 % error criteria was not met. However those numbers are of different meaning.

The Ansys software uses two error criteria to evaluate convergence. The energy error and the delta error. The first is calculated by sophisticated algorithm and means, that the solution is in the x% vicinity of the simulated value. The former is the difference between last two energy error values in percentage. Be aware, that the simulation can converge into false convergence due to low quality mesh or wrong model. The convergence does not mean that the simulation reached real value.

The percentage error criteria means, that the solver aims to get both of previously mentioned values under that criteria value. The 0.05 % error value may seem too strict, but we have to take in notion,

that with increasing capacity the error increases accordingly. Also the increment is calculated from two values so in a worse scenario, the deviation could be as much as: $\Delta = ((1.0005 \cdot 361.76) - (0.9995 \cdot 342.25)) - 19.53 = 0.33pF$ which is almost 1.7%.

Graph concerning convergence, is presented in figure 3b. There we can observe, that the R model converged faster and with fewer tetrahedrons (the mesh is less dense). Simultaneously, it was found out, that one parametric solution for R model was 30% faster. The slightly rounded edges caused finer mesh at details (edges), so the refinement of the mesh was focused around those details. On the other hand, the mesh of NR model was refined mostly in the whole volume, which caused the need of finer mesh. In this case, the R and NR model are almost equal. The NR model could perform better, by slightly changing the simulation settings. But we could run into the risk of converging into a false solution. Rounding corners makes converging into right solution more certain, but for models with more corners it could also mean an exponential increase in the number of tetrahedrons.

8 CONCLUSION

The calculable capacitor based on Thompson-Lampard theorem was described and model of its N.S.L version was created for simulation. Two simulations were performed. Results show that the model with rounded edges converges faster and provides more precise results. The model with sharp edges is faster to create, but convergence is slower and needs more tetrahedrons (as seen in figure 3b), so more memory is needed. Model designer should be cautious when using sharp edges, because there could be a chance of a false convergence. The verification of a correct convergence was possible due to knowledge of analytic results beforehand. Without that knowledge it would be necessary to make more simulations with different settings. However, the results from both models were approximately the same. The increment deviates only slightly. The knowledge obtained during the work on this paper will shall lay a great foundations for future research about planar capacitive sensing. Simultaneously, reader can make use of this paper as a guide for proofing the reader's modeling approaches by making the same simulation.

REFERENCES

[1] L. K. Baxter. *Capacitive Sensors. Design and Applications.* New York: IEEE Press, 1997. ISBN 0-7803-1130-2.

[2] A. M. Thompson, D. G. Lampard. *A New Theorem in Electrostatics and its Application to Calculable Standards of Capacitance.* Sydney: Nature, 1956.

[3] National Bureau of Standards. *Dimensions Vol.59.* U.S. 1975. No 25-26527.

[4] A. M. Thompson. *The Cylindrical Cross-Capacitor as a calculable standard.* Rev 2. May 1959

[5] W. K. Clothier. *A Calculable Standard of Capacitance.* Metrologia 1 36. Rev 17.9.1964.

[6] G. H. Rayner. *NPL Calculable Capacitor*. IEEE Transactions on Instrumentation and Measurement, Vol. IM-21, No. 4, November 1972.

[7] Stephan D.A. Hannot, Daniel J. Rixen, V. Rochus. *Rounding the Corners in an Electromechanical FEM Model*. Coupled Problems 2007. Barcelona.

PEDESTRIAN DETECTOR DOMAIN SHIFT ROBUSTNESS EVALUATION, AND DOMAIN SHIFT ERROR MITIGATION PROPOSAL

Tomáš Zemčík

Doctoral Degree Programme (2), FEEC BUT

E-mail: zemcikt@feec.vutbr.cz

Supervised by: Karel Horák

E-mail: horak@fit.vutbr.cz

Abstract: This paper evaluates daytime to nighttime traffic image domain shift on Faster R-CNN and SSD based pedestrian and cyclist detectors. Daytime image trained detectors are applied on a newly compiled nighttime image dataset and their performance is evaluated against detectors trained on both daytime and nighttime images. Faster R-CNN based detectors proved relatively robust, but still clearly inferior to the models trained on nighttime images, the SSD based model proved non-competitive. Approaches to the domain shift deterioration mitigation were proposed and future work outlined.

Keywords: Object detection, Pedestrian detection, Cyclist detection, ADAS, AV, Faster R-CNN, SSD, Domain shift, Domain adaptation, Data augmentation

1 INTRODUCTION

In Autonomous Vehicles (AV) and Advanced Driver Assistance Systems (ADAS) one of the critical tasks is collision avoidance. In order to implement any collision detection and avoidance system situational awareness is required. It is the task of machine perception to supply this situational awareness to the higher level planning and decision making system, that being either another machine, or the human element [2].

The machine perception field has been dominated with Convolutional Neural Networks (CNNs) in recent years, and chances are this is going to continue in upcoming years. While CNNs have proven great potential in this field [1], they have a number of weaknesses, in particular they require large amounts of annotated data [3]. Sometimes it is impractical, or outright impossible, to acquire the amount of data needed and in those cases either the performance suffers, or the data has to be created or multiplied via dataset augmentation.

In the field of traffic images one of the examples is nighttime data, the established traffic image datasets are either daytime datasets such as the KITTI and Caltech datasets [8, 9], or the nighttime portion of the dataset is underrepresented (sometimes if not in the number of images, than in the number of class instances) such as the LISA Traffic Light Dataset [10]. Although it has to be noted that the situation is improving with datasets such as the NightOwls dataset [11]. Still instances of certain classes are quite rare in the nighttime by their nature (such as cyclists), yet it is expected of the AV and ADAS to be reliable in these situations. Solution thus has to be found either in architectures generalizing well between day and night, or in generating sufficient data for these rare classes.

In this short paper the current state of the ongoing research is presented. First a short review of current state-of-the-art in domain shift and CNN domain shift robustness is given. Then an experiment to determine robustness of various object detection architectures is presented and solutions to improvements of these detectors are proposed. Lastly future work is outlined and conclusions drawn.

2 PREVIOUS WORK

This paper is concerned with domain shift and its implications on object detection, in particular it is concerned with the domain shift between the daytime traffic image source domain, and the nighttime traffic image target domain. A domain shift is a shift between source and target domains, in other words it is a data remapping where the distributions are different but related [4]. As mentioned above the ways of dealing with this shift is either having data from the target domain - obtained in the usual way, or through data augmentation, or having an architecture robust (architecture robustness has been examined in depth by [16]) to domain shift (or one capable of domain adaptation):

Data augmentation: in recent years a body of work has gone into development of data augmentation methods to increase performance, avoid overfitting and improve reliability of Deep Learning applications. A comprehensive overview of the methods and current trends has been compiled in [14]. The scope of data augmentation reaches from the very simple and naive, such as geometric transformations, colour space augmentations etc. all the way to very complex methods such as Generative Adversarial Networks (GANs) and Neural Style Transfer.

While the simple augmentations has now become a standard in object detection, especially methods like Neural Style Transfer could be very beneficial in domain shifted data generation [3, 15]. GAN based augmentation methods are also perspective.

With augmented data the chosen object CNN can be trained as usual.

Domain adaptation: Domain adaptation can be internal or external to the network architecture (in a way a dataset augmentation in itself is a domain adaptation). A comprehensive survey of methods applicable for domain shifted traffic detection has been conducted by [15, 4]. Noteworthy is also Progressive Domain Adaptation method used in [17].

Domain adapted method usually requires samples of data from both domains, albeit the target domain data might be relatively sparse.

The two object detection architectures this paper is predominantly concerned about are the Single Shot Multibox Detector (SSD), and the Faster R-CNN. Both of the detectors are regarded as state-of-the-art in object detection, and are used in traffic applications [1].

The **Faster R-CNN** architecture was introduced in 2015 in [6] as a performance improvement on the original R-CNN and Fast R-CNN architectures. The detector itself uses a pretrained feature extractor CNN generating a feature map from the input image. A custom trained region proposal network (RPN) then selects 2000 likely locations for objects that are then classified. The output detections are selected based on the classification confidence.

The **SSD** was first published in 2016 by Liu et al. [5] as a real-time object detector that was comparable in performance to the then state-of-the-art Faster R-CNN while being faster. The SSD uses a pretrained feature extractor CNN and a cascade of convolutional layers progressively diminishing the feature map size. features in predefined default bounding boxes are then classified and non-maximally suppressed to only select the strongest detections.

3 THE EXPERIMENT

This paper builds on our previous studies published in [1], where the performance of pedestrian (and concurrent pedestrian and cyclist) detectors was evaluated on real traffic data that was degraded by some of the most common image degradations (motion blur, out-of-focus blur and JPEG compression artifacts) with varying severity. Goal of this evaluation was to determine how various degradations affect the different architectures in order to be able to select the right architecture for further work. The study compared various object detectors based on SSD and Faster R-CNN architectures, and

concluded that in general Faster R-CNN showed better performance and was more robust to degradations at the cost of increased computational cost and thus lower framerate. In this paper the same architectures were applied onto a newly compiled night dataset in order to determine how the day to night domain shift affects their performance.

3.1 DATASET

A new night dataset has been created for verification by combining the Small Pedestrian Night Dataset (SNPD) [7], author's own dashcam images, and some other publicly available images. The entire verification dataset contains over 300 images (none of which have been used in the detector training) that have been manually annotated with classes of pedestrian and cyclist, and contains several hundred instances of the pedestrian class, and tens of cyclists (as cyclists are difficult to come by in nighttime). In keeping with the original SNPD the images were obtained using a medium range dashcam and were shot at nighttime or dawn in all weather. The dataset represents urban and suburban areas. Figure 1 shows typical images of the night dataset.
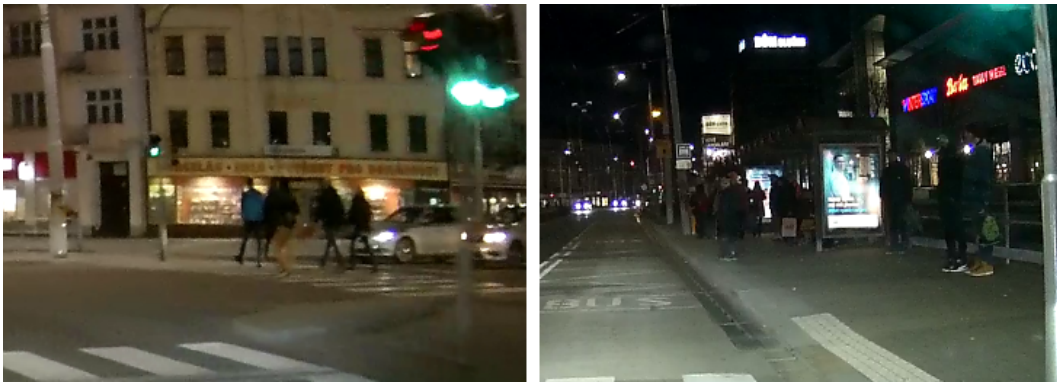

Figure 1:   Representative images from the night dataset

For comparison Small Pedestrian and Cyclist Dataset (SPCD) and Low Quality Traffic Dataset (LQTD) from our previous work [1] were used. The SPCD is a daytime dataset made out of typical dashcam images of urban and suburban traffic situations. The LQTD contains purposely made low quality daytime images using a low resolution, shot camera in all weather and through a dirty windscreen with various reflections. Both datasets were manually annotated with the pedestrian and cyclist classes.

3.2 DETECTORS

In this paper the same SSD Lite and Faster R-CNN architectures were applied onto a newly compiled night dataset in order to determine how the day to night domain shift affects their performance.

Two of the detectors were retrained specially for this paper. These are:

- **FRCNN Zemcik** detector based on the Faster R-CNN architecture

- and the **SSD Zemcik** based on the SSD Lite architecture.

Both trained to detect pedestrians and cyclists. Both detectors were trained on a custom dataset made up of a combination of the KITTI dataset, and author's own data. These detectors were trained purely on daytime images of varying quality.

To be able to compare these detectors further three detectors were evaluated. These were:

- The Faster R-CNN based **FRCNN Kitti zoo** [13] trained to detect pedestrian and cars purely on the KITTI dataset.

- The Faster R-CNN based **FRCNN Tilgner** [7] trained to detect pedestrians predominantly on a mixture of KITTI and SNPD images, thus having some nighttime images in its training.

- The SSD Lite based **SSD Tilgner** [7] being likewise trained on pedestrians from KITTI and SNPD datasets.

3.3 RESULTS

The trained detectors were applied onto the described dataset and detections were obtained. A treshold of 50% was selected. Figure 2 shows the same image with detections by both tested detectors.

Figure 2: Examples of pedestrian detections by FRCNN Zemcik and SSD Zemcik respectively

The results of the experiment were evaluated by the average precision as defined by the Pascal Visual Object Challenge (Pascal VOC) [12]. The precision has been calculated for each class separately, as some of the detectors detect more than one class and the Mean Average Precision (mAP) would not be comparable.

The Pascal VOC AP metric only considers detections as true positives if the Intersection over Union (IoU) measure of the detected and ground truth bounding boxes is over 0.5:

$$IoU = \frac{Area\ of\ Intersection}{Area\ of\ Union} \tag{1}$$

The average class precision of each class is then calculated as the mean of precisions taken at eleven points of the precision / recall curve. The points being $Recall = \{0, 0.1, 0.2, ..., 0.9, 1\}$. The Pascal VOC AP then is:

$$AP = \frac{1}{11} \cdot \sum_{Recall \in \{0, 0.1, ..., 1\}} Precision(Recall) \tag{2}$$

Table 1 compares the class average precisions of the above-described detectors on the night dataset in contrast to daytime SPCD and LQTD datasets. It is readily apparent that results on the night dataset are inferior to those on the daytime datasets. It is also obvious that the SSD based detectors are not competitive at night.

Detector	SPCD	LQTD	Night dataset
FRCNN Kitti zoo (pedestrian)	69.2%	29.3%	40.9%
FRCNN Zemcik (pedestrian)	**84.7%**	**56.7%**	21.6%
FRCNN Zemcik (cyclist)	**75.4%**	**50.6%**	**7.0%**
FRCNN Tilgner (pedestrian)	71.8%	64.6%	**48.4%**
SSD Zemcik (pedestrian)	19.7%	10.6%	0.1%
SSD Zemcik (cyclist)	62.5%	3.1%	1.2%
SSD Tilgner (pedestrian)	53.2%	37.4%	1.1%

Table 1: Class average precision of all detectors on the night dataset compared to daytime datasets. (The best detector for each class highlighted)

Furthermore notice that **FRCNN Tilgner** and **SSD Tilgner** were clearly better with the night dataset due to having been trained on night images even if they were inferior on the daytime validation datasets.

4 PROPOSED SOLUTIONS

In order to improve the performance of the detectors in nighttime several solutions present themselves:

Firstly the **naive approach** would be to avoid the domain shift altogether by training the detector models on an actual nighttime dataset. A suitable dataset now exists in the abovementioned NightOwls dataset [11], which is annotated both for pedestrians and cyclists. The drawback is that should the solution require a further class or a further domain, this approach would require yet another full dataset.

Secondly a solution using some kind of a sophisticated **data augmentation approach** is possible. This approach would involve generating a synthetic nighttime dataset from an actual daytime dataset either using GAN or Neural Style transfer architecture. This approach would however require to train two separate models - first the data augmentation model and only then the object detector itself.

And lastly the **domain adaptation approach**. A promising approach that would allow for a single model to be trainable at once, it would also allow for future extension for more classes and domains.

5 FUTURE WORK AND CONCLUSIONS

In this short paper a short overview of the domain shift problem was presented. An experiment was conducted on currently used pedestrian and cyclist detection models in order to determine their robustness to daytime to nighttime domain shift. It was found - as expected - that their performance is inferior to models outright trained on nighttime images. The Faster R-CNN based models proved to be more robust than the SSD based model. Approaches were then proposed to improve the detectors' performance.

In future work we would like to continue developing the models by firstly retraining them on the NightOwls dataset, and then implementing a domain adaptation version of the model. This way we will be able to compare the performance of a conventionally trained CNN and a CNN using a synthetic data step. Further we plan to apply the general domain adaptation principles onto a model using multi-sensor input.

REFERENCES

[1] Zemčík, T., Kratochvíla, L., Bilík, Š., Boštík, O., Zemčík, P., & Horák, K. (2021). Performance Evaluation of CNN Based Pedestrian and Cyclist Detectors On Degraded Images. *International Journal of Image Processing (IJIP)*, 15(1), 1-13. https://www.cscjournals.org/library/manuscriptinfo.php?mc=IJIP-1213

[2] Zemčík, T. (2020). Overview of approaches to multisensor fusion as used in ADAS and AV. In *Proceedings I of the 26th student EEICT 2020: General papers* (pp. 387 - 391). Brno University of Technology, Faculty of Electrical Engineering and Communications. https://www.fekt.vut.cz/conf/EEICT/archiv/sborniky/EEICT_2020_sbornik_1.pdf

[3] Ji, J., Xu, Z., Yu, H., Fu, L., & Zhou, X. (2020). Domain Adaptation For Vehicle Detection In Traffic Surveillance Images From Daytime To Nighttime. *Transportation Research Board the 99th Annual Meeting*. https://scholarworks.utrgv.edu/cgi/viewcontent.cgi?article=1020&context=cs_fac

[4] Gopalan, R., Li, R., Patel, V. M., & Chellappa, R. (2012). Domain Adaptation for Visual Recognition. *Foundations and Trends® in Computer Graphics and Vision*, 8(4), 285-378. https://doi.org/10.1561/0600000057

[5] Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C. -Y., & Berg, A. C. (2016). SSD: Single Shot MultiBox Detector. *Computer Vision – ECCV 2016*, 21-37. https://doi.org/10.1007/978-3-319-46448-0_2

[6] Ren, S., He, K., Girshick, R., & Sun, J. (2017). Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(6), 1137-1149. https://doi.org/10.1109/TPAMI.2016.2577031

[7] Tilgner, M. (2019). *Pedestrian detection in Traffic Environment by Machine Learning* [master's thesis]. Brno University of Technology.

[8] Geiger, A., Lenz, P., Stiller, C., & Urtasun, R. (2013). Vision meets robotics: The KITTI dataset. *The International Journal of Robotics Research*, 32(11), 1231-1237. https://doi.org/10.1177/0278364913491297

[9] Dollar, P., Wojek, C., Schiele, B., & Perona, P. (2009). Pedestrian detection: A benchmark. *2009 IEEE Conference on Computer Vision and Pattern Recognition*, 304-311. https://doi.org/10.1109/CVPR.2009.5206631

[10] Mogelmose, A., Trivedi, M. M., & Moeslund, T. B. (2012). Vision-Based Traffic Sign Detection and Analysis for Intelligent Driver Assistance Systems: Perspectives and Survey. *IEEE Transactions on Intelligent Transportation Systems*, 13(4), 1484-1497. https://doi.org/10.1109/TITS.2012.2209421

[11] Neumann, L., Karg, M., Zhang, S., Scharfenberger, C., Piegert, E., Mistr, S., Prokofyeva, O., Thiel, R., Vedaldi, A., Zisserman, A., & Schiele, B. (2018). NightOwls: A pedestrians at night dataset. In *Asian Conference on Computer Vision* (pp. 691–705).

[12] Everingham, M., Van Gool, L., Williams, C., Winn, J., & Zisserman, A. (2010). The Pascal Visual Object Classes (VOC) challenge. *International Journal of Computer Vision*, 88, 303-338.

[13] Huang, J., Rathod, V., Sun, C., Zhu, M., Korattikara, A., Fathi, A., Fischer, I., Wojna, Z., Song, Y., Guadarrama, S., & Murphy, K. (2017). Speed/Accuracy Trade-Offs for Modern Convolutional Object Detectors. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 3296-3297. https://doi.org/10.1109/CVPR.2017.351

[14] Shorten, C., & Khoshgoftaar, T. M. (2019). A survey on Image Data Augmentation for Deep Learning. *Journal of Big Data*, 6(1). https://doi.org/10.1186/s40537-019-0197-0

[15] Li, W., Li, F., Luo, Y., Wang, P., & sun, J. (2020). Deep Domain Adaptive Object Detection: a Survey. *2020 IEEE Symposium Series on Computational Intelligence (SSCI)*, 1808-1813. https://doi.org/10.1109/SSCI47803.2020.9308604

[16] Heinze-Deml, C., & Meinshausen, N. (2021). Conditional variance penalties and domain shift robustness. *Machine Learning*, 110(2), 303-348. https://doi.org/10.1007/s10994-020-05924-1

[17] Hsu, H. -K., Yao, C. -H., Tsai, Y. -H., Hung, W. -C., Tseng, H. -Y., Singh, M., & Yang, M. -H. (2020). Progressive Domain Adaptation for Object Detection. *2020 IEEE Winter Conference on Applications of Computer Vision (WACV)*, 738-746. https://doi.org/10.1109/WACV45572.2020.9093358

GEL POLYMER ELECTROLYTES COMPOSED OF ETHYL METHACRYLATE AND METHYL METHACRYLATE MEASURED AT HIGHER TEMPERATURE

Soňa Peterová

Doctoral Degree Programme (1), FEEC BUT

E-mail: xpeter18@stud.feec.vutbr.cz

Supervised by: Marie Sedlaříková

E-mail: sedlara@feec.vutbr.cz

Abstract: This paper is focused on the use of gel polymer electrolytes composed of copolymers of ethyl methacrylate and methyl methacrylate with positive NMC electrode ($LiNi_{0.33}Mn_{0.33}CO_{0.33}O_2$). This article describes the composition of gel polymer electrolyte, the method of preparation and description of the NMC electrode. The experiment is based on the measurement of capacity and impedance at the beginning and at the end of the experiment at room temperature and at higher temperature 40 °C.

Keywords: gel polymer electrolytes, copolymers, NMC electrode

1 INTRODUCTION

With the wider possibilities of using battery-powered devices, the diverse demands on these sources of electricity are growing also. The requirements for the number of cycles, capacity, electrochemical stability, safety, usage of ecological materials and ecological disposal of batteries are increasing. Nowadays, when lithium batteries are widely used, it is being considered to improve the electrolytes used in them too. Liquid electrolytes are used mostly in lithium batteries. The main reason is a high ionic conductivity. However, the problems of liquid electrolytes are for example safety, mainly flammability and toxicity. Gel polymer electrolytes consist of a polymer, an inorganic salt and an organic liquid. The polymer network prevents the liquid part from escaping from the matrix and gives the gel the properties of a solid. The organic liquid serves as a plasticizer and gives the matrix the properties of the liquid. The basic requirements for these electrolytes are high ionic conductivity in a wide range of heat, good mechanical properties, thermal and electrochemical stability and long life span. The conductivity of these electrolytes is in units of mS/cm. [1] [3]

2 EXPERIMENT

The experiment began with the preparation of gel polymer electrolytes (GPE). The gel polymer electrolytes consisted of salt, solvent, initiator of UV polymerization, crosslinking agent and monomers. Initially, the properties of GPE with single monomers like ethyl methacrylate (EMA), butyl methacrylate (BMA), isobutyl methacrylate (IBMA), lauryl methacrylate (LMA), trimetoxysilylpropyl methacrylate (TSPMA) and ethoxyethyl methacrylate (EOEMA) were measured. Combination of two GPE composed of monomers with the highest conductivity and electrochemical stability was chosen. After selecting the two monomers, the composition of the gel polymer electrolyte with a copolymer of EMA and MMA was calculated. The ratio of monomers EMA 20% and MMA 80% suited best to these conditions. Selection of the appropriate GPE composition was followed by the preparation of the Li/GPE/NMC cell. Firstly, it was necessary to mechanically clean the lithium plate from oxides on the surface, then to cut out a GPE of a certain size and transfer it to the cut-out part of the lithium and add an NMC electrode. Followed by setting the parameters of PEIS and GCPL measuring methods. Measurements at room and higher

temperature can not be performed on one sample, therefore two samples with the same composition were prepared for the measurement.

2.1   CHEMICAL COMPOSITION

The method of preparing gel polymer electrolytes is based on mixing monomers with a polymerization initiator, a crosslinker and a salt in a solvent. During the preparation of gel polymer electrolytes, it is necessary to work in an inert atmosphere. Upon contact of the gel polymer electrolyte with the ambient air, the material degrades rapidly.

The GPE consists of the following materials:

- **Salt** Lithium hexafluorophosphate ($LiPF_6$)

- **Solvent** Ethyl Carbonate and Diethyl Carbonate (EC/DEC in weight 1:1)

- **Initiator of UV polymerization** Benzoin Ethyl Ether (BEE)

- **Crosslinking agent** Ethylene Glycol Dimethacrylate (EDMA)

- **Monomer** Ethyl Methacrylate (EMA) and Methyl Methacrylate (MMA)

The exact chemical composition of the gel polymer electrolyte is shown in table 1.

	Chemical substance	Quantity
EMA 20 % MMA 80 %	LiPF6	0.1519 ml
	EC/DEC	1.899 µl
	EMA	137.54 µl
	MMA	470.78 µl
	BEE	0.0139 g
	EDMA	38.2 µl

Table 1: Chemical composition of GPE

2.2   NMC ELECTRODE

NMC-based electrodes suitably combine the properties of nickel and manganese. Nickel has a high specific energy but poor stability. Manganese has a very low internal resistance but low specific energy. The NMC electrode excels in specific energy, specific power, safety, performance and life span. However, the price of these electrodes is also higher. It is used as a material for a positive electrode. There are several types of NMC electrodes, which are divided on the basis of the proportion of individual elements. The most common type is 111 NMC, which means that all elements are represented by the same proportion. For example, other basic elements ratios are 442 NMC or 622 NMC. Besides of the NMC electrodes, there are other materials used for the positive electrode, such as NCA (lithium-nickel-cobalt-aluminium) or LFP (lithium-iron-phosphate). The completed Li/GPE/NMC cell is shown in the figure 1 below. [2] [4]

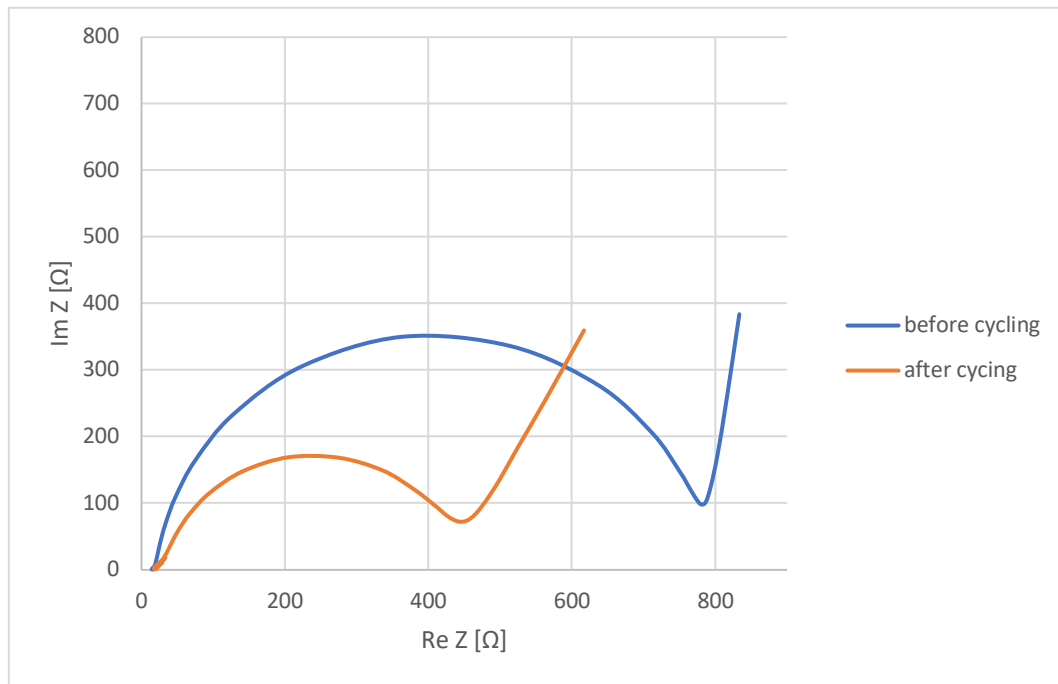**Figure 1:** Complete Li/GPE/NMC cell

2.3   RESULTS

Figure 2: PEIS graph of Li/GPE/NMC cell at room temperature.

The PEIS method (Potentiostatic Electrochemical Impedance Spectroscopy) is an experimental method. The complex impedance is evaluated over a wide range of frequencies. The result of the measurement is the impedance spectrum and it is shown in figure 2. The method was used at the beginning and at the end of the measurement of the Li/GPE/NMC cell. At the beginning of all the measurement, the value of the real component of impedance is 779 Ω and after the measurement of capacity, the value of impedance is 452 Ω. This change can be evaluated as a decrease in impedance of 58%.
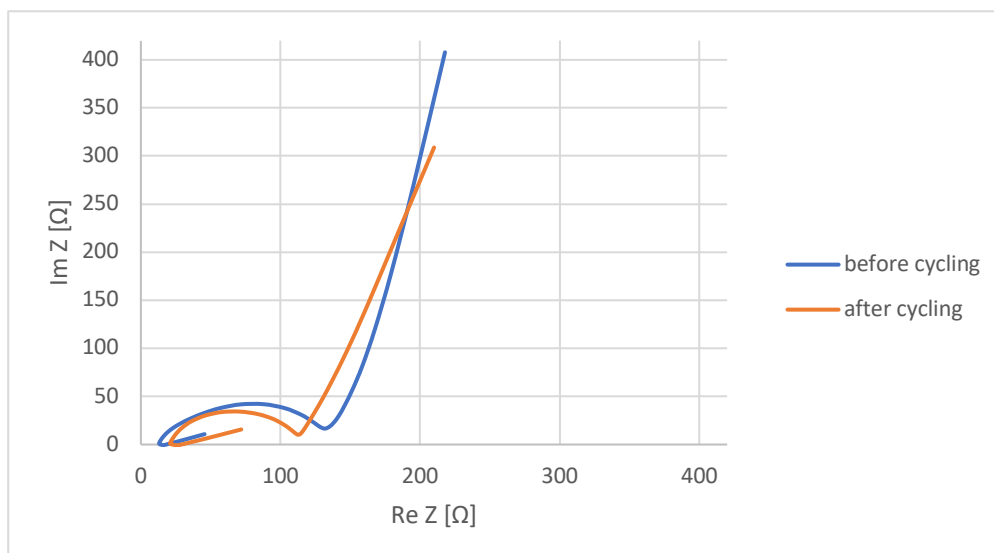
Figure 3: PEIS graph of Li/GPE/NMC cell at higher temperature 40 °C.

Measurement with higher temperatures were done with the same parameters only the temperature was 40 °C. There is graph of measured impedances before and after cycling shown in figure 3. The impedance value at the beginning of the measurement was 134.8 Ω and after the GCPL method measurement the impedance value was 112.6 Ω. The impedance decreased of 16.5 %. When measured with PEIS method at room temperature, the impedance decreased of 58%. This is significant difference from measurement at higher temperatures. Furthemore, it would be useful to detect differences in impedance at different temperatures.
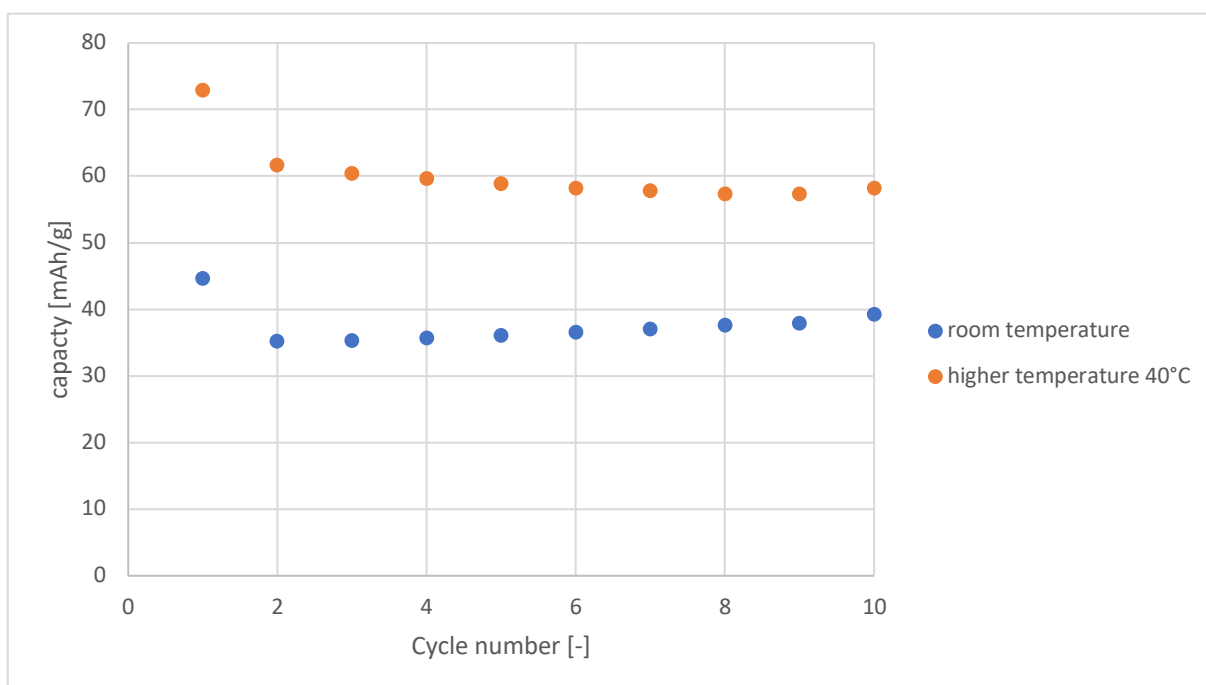
Figure 4: Capacity during cycling of Li/GPE/NMC cell.

Figure 4 shows the dependence of capacity on cycle number. The GCPL (Galvanostatic Cycling with Potential Limitation) was used for the measurement. This method allows cyclic charging and

discharging within the required voltage limits. Ten cycles of charging and discharging with a current of 0.05 C up to 1.2 V were measured. At the beginning of the measurement, the capacity was 44.58 mAh/g, after that the value dropped to 35.15 mAh/g. From the third cycle, the capacity slowly raised again to 39.22 mAh/g. Due to this measurement the drop of capacity is 13.6 %. The average capacitance value is 37.51 mAh/g, while the theoretical value of capacity of the NMC electrode is 160 mAh/g.

As with room temperature and higher temperature measurements, we can see a significant decrease in capacity after the first charge and discharge cycle. Charging and discharging at room temperature took 91 hours, while at higher temperature the cycles took 141 hours. The average value of capacity at higher temperatures was 60.2 mAh/g.

3   CONCLUSION

The aim of this experiment was to determine behavior of a gel polymer electrolyte based on a copolymer of EMA and MMA with NMC electrode. The conductivity of the prepared gel polymer electrolyte was 5.52 mS/cm. Commercial electrodes from Customcell were used for the experiment. The measurement of gel polymer electrolyte and subsequent measurement of system Li/GPE/NMC were done at room temperature. The result of this measurement is that the prepared system of Li/GPE/NMC is stable at room temperature. After these experiments, it was necessary to investigate the behavior even at higher temperatures. The measurement was done at a temperature of 40 °C. The capacity of the cell is higher at higher temperatures and charging and discharging cycles take longer than at room temperature. Furthermore, it would be possible to continue research at higher and lower temperatures and study changes in capacity as a function of temperature. These measurements are important for follow-up research, which could explore the application of GPE in thin layers, because GPE applied in thin layers has a higher conductivity by volume.

REFERENCES

[1]   Ming Zhu, Jiaxin Wu, Yue Wang, Mingming Song, Lei Long, Sajid Hussain Siyal, Xiaoping Yang, Gang Sui, Recent advances in gel polymer electrolyte for high-performance lithium batteries, Journal of Energy Chemistry, 2019, Beijing, ISSN 2095-4956, [https://doi.org/10.1016/j.jechem.2018.12.013.]

[2]   Typy lithium-iontů, DS New Energy [online] Feb 21, 2019. Dostupné z: http://cz.dsnsolar.com/info/types-of-lithium-ion-32657770.html

[3]   Manuel Stephan, A.: Review on gel polymer electrolytes for lithium batteries, European Polymer Journal 42 (2006) 21-42

[4]   Lithium-ion positive electrode technologies [online]. wattalps, 8 Mar 2018n. l. [cit. 2021-02-15]. Dostupné z: https://www.wattalps.com/lithium-ion-positive-electrode-technologies/

CHARACTERIZATION OF ALN NANOLAYERS DEPOSITED ON A SURFACE OF HOPG BY PE-ALD

Rashid Dallaev

Doctoral Degree Programme (4th year), FEEC BUT

E-mail: xdalla03@vutbr.cz

Supervised by: Petr Sedlák

E-mail: sedlakp@feec.vutbr.cz

Abstract: In this study plasma-enhanced atomic layer deposition process of AlN has been performed with the purpose to test the expediency of highly oriented pyrolytic graphite (HOPG) to serve as a substrate in such process. The obtained samples were thoroughly analyzed using various analytical techniques. Atomic force microscopy was employed for studying topographic and morphological features of the surface; x-ray photoelectron spectroscopy (XPS) analysis supported by second ion-mass spectrometry method (SIMS) has been conducted on the obtained sample to investigate the chemical nature of the deposited films as well as elemental distribution. Temperature stability of HOPG makes it a suitable substrate for preparation of AlN films, being a bottom contact for further testing of the films electrical properties. The data gathered from the aforementioned techniques have indicated that HOPG is a viable choice for AlN ALD process.

Keywords: aluminium nitride, atomic layer deposition, highly oriented pyrolytic graphite, x-ray photoelectron spectroscopy, atomic force microscopy, second ion-mass spectrometry.

1   INTRODUCTION

Aluminium nitride (AlN) is a semi-conductive piezoelectric material with various promising physical and piezoelectric properties which attract attention of scientists from all over the world. AlN demonstrates an outstanding performance in a variety of electronics and covers quite a large range of applications which have been abundantly described by many researchers [1-11]. There is a lot of reports of AlN thin films grown using different physical and chemical vapour deposition methods (PVD, CVD) [2, 8, 9]. Atomic layer deposition (ALD) is one of the methods of CVD group. Its main advantages over others stem from the fact that reactions in ALD are sequential and self-limiting which enables a precise control over the film's growth at the atomic level [2, 8, 9, 11]. Further benefit of ALD is that after each reaction cycle the deposition chamber is purged from unreacted components of metalorganic precursors by an inert gas which leads to excellent purity and homogeneity of the film. So far, AlN thin films have been mostly deposited on silicon and sapphire wafers which proved to be reliable substrates to be employed in the process [8, 9, 11]. However, it remains a relevant topic of research to test out different substrates for AlN atomic layer deposition since chemical reactions that take place on the surface of the substrate as well as the nature of film/substrate interface may define the resulting quality of the entire deposited film. Thus, in this paper it has been decided to test out highly oriented pyrolytic graphite (HOPG) as a substrate for AlN ALD process. HOPG appears to be a promising substrate for AlN deposition process due to the carbon's ability to withstand extremely high temperatures, 3550 °C (a potential for high-temperature processing of the sample). Another advantage of HOPG is that its laminated nature allows for easy transferring of films grown on it onto a different foundation should such a need arise [12, 13].

2   EXPERIMENTAL DETAIL

To deposit AlN thin films using plasma enhanced atomic layer deposition (PE-ALD) on highly oriented pyrolytic graphite (HOPG) substrate Ultratech/CambridgeNanoTech Fiji 200 ALD instrument have been used. Obtained films have been analyzed using AFM, XPS and SIMS methods.

In total 1500 ALD cycles have been performed which translates into thickness of approximately 100nm (1cycle ≈ 0,629 Å). The temperature of deposition was 300 °C. The energy of plasma was 300W. The sequence of each cycle consisted of following steps:

1) introduction of TMA (0,06sec),
2) purge 10sec,
3) initiate flow of N2/H2 (20 sccm) and enable plasma (40 sec),
4) purge 5sec.

To ensure to purity of the surface, first layers of HOPG substrate were removed using an adhesive tape right before loading the samples into the ALD chamber.

SIMS analysis was carried out on ION-TOF TOF.SIMS 5 instrument in positive mode. The instrument enables analysis in dual gun mode. Oxygen gun with energy of 2KeV was employed for rough sputtering (material removal only), whereas Bi gun provides much slower sputtering rate and secondary ions produced by it are collected in an analyzer. Area of the crater was chosen to be 200x200nm, area of analysis 80x80 nm. The 3D modelling was done in the native software SurfaceLab 7.1 provided by the manufacturer.

Kratos Analytical Axis Supra instrument with Al Kα excitation source and emission current of 15mA was employed to perform XPS analysis. Wide spectrum was taken at of 80 eV. High resolution spectra were taken at of 20 eV. All spectra were calibrated by shifting major C1s peak (C-C bond) to 284.8 eV. The presented spectra were made in CasaXPS software, SG linear smoothing and background subtraction tools were used.

3   RESULTS AND DISCUSSION

3.1   ATOMIC FORCE MICROSCOPY

Figure 1 shows surface images obtained by AFM. The average roughness for 2x2 and 5x5 μm images are 36,4 Å and 34,7 Å correspondingly. In total, around 15 images have been taken with different resolutions (0,5x0,5 μm, 2x2 μm, 10x10 μm). Several images have been taken at different spots all across the sample and exhibited the similar pattern. The average surface roughness value (Ra) fluctuates around ~35 Å for all of them, this indicates excellent uniformity of the deposited layer and conformal growth.
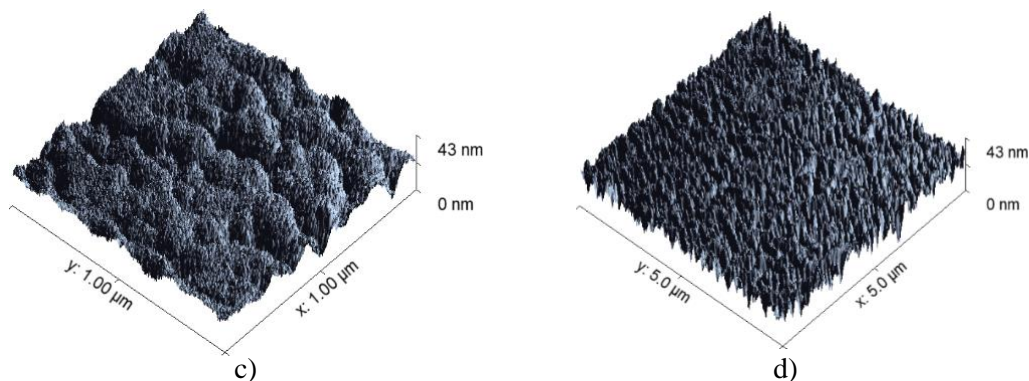

a)   b)
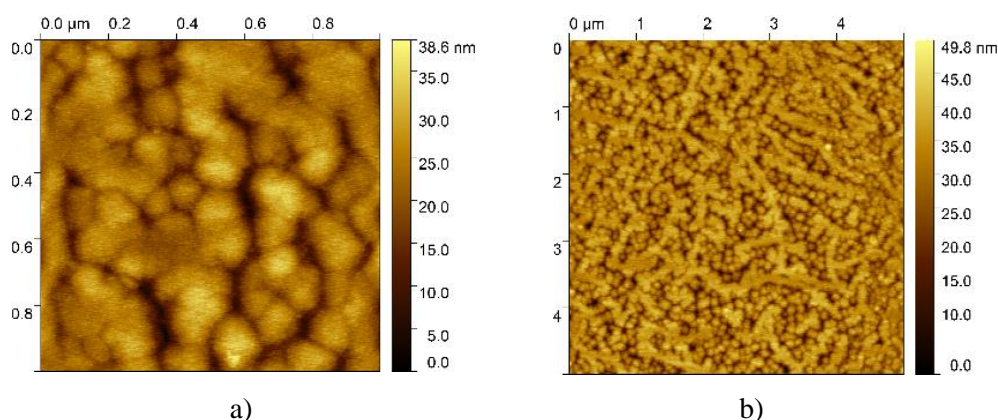
Figure 1. AFM images of: a) 2D 1x1 μm , b) 2D 5x5 μm, c) 3D 1x1 μm , d) 3D 5x5 μm

3.2 SECONDARY ION-MASS SPECTROMETRY

Elemental distribution patterns according to SIMS are provided in Figure 2. During sputtering in SIMS not only individual ions are ejected from a material (C+, N+, O- etc.) but also composite ions (OH-, CH3-, AlO+ etc.) and even molecules. Which is why, in order to get a complete picture about the target compound we need to search for all possible combinations of its constituents, in our case those they are N+, AlN (a, b) and Al (c). By looking at these 3D images of ejected ions distribution we can clearly tell at which points the AlN ends and HOPG substrate begins. 3D profile of carbon (d) is also given to emphasize that shift. Furthermore, some of the carbon can also be noticed in the AlN layer, that is explained by the fact that the precursor employed in the ALD process (TMA) contains carbon some of which might have formed different bonds within the layer, which is also confirmed by XPS results.
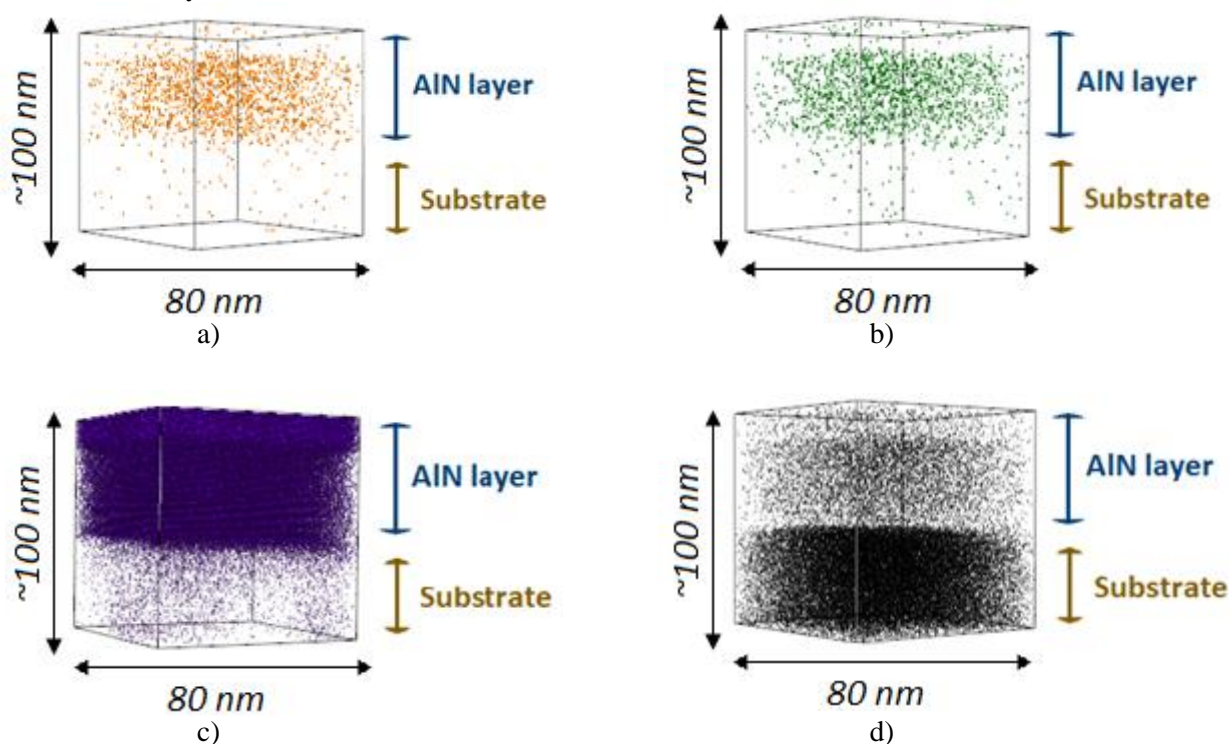

Figure 2. SIMS 3D distribution for AlN on HOPG: a) N+, b) AlN+, c) Al+, d) C+

3.3 X-RAY PHOTOELECTRON SPECTROSCOPY (XPS) DATA

XPS spectra are given in Figure 3. Wide spectum (fig. 3a) shows that the surface of the films is heavily oxidized and some carbon impurities are also present. The presence of carbon and oxygen is the result of the samples exposure to atmosphere after the deposition. Thus, the bulk of the sample is

expected to have superior degrees of purity. The nitrogen peak (fig. 3b) consists of three subpeaks with binding energies of 396,7 eV, 398,3 eV and 399,9 eV. The biggest of them located at 396,7 eV and according to a number of sources [14-18] is the one belonging to N-Al. The remaining two are located at 398,2 eV and 400 eV and attributed to N-O and N-Al-O correspondingly [14, 16].

Aluminum exhibits two subpeaks (fig.3c) the biggest occurs at 74 eV and is assigned as Al-N bond, the same energy values for Al-N bond are reportied in literature [15, 16]. There is also a good correlation with the nitrogen pieak in this regard, which also exhibits the presence of Al-N bond. The second peak has energy of ~73,5 eV and we speculate belongs to Al-Al defects which are though small in percentage but nonetheless present in the film. Given the high affinity of Al/AlN to oxidization [15, 16], both during the deposition process as well as in the open air, Al-O subpeak could also be reasonably anticipated here, however, it occurs at higher binding energies (>75eV) [19, 20] and hasn't been noted in this particular sample. XPS spectra have been taken at several different points on the sample surface and proved to be practically identical.

It appears instead that the bulk of the oxygen is bound to carbon in one way or the other which can be seen from C1s and O1s high-resolution spectra (fig. 3d and fig. 3e). Deconvolution of C1s resulted in 3 subpeaks located at 289,2 eV, 287,4eV and 284,8eV which correspond to O-C=O, C-O-C and C-C bond correspondingly [21]. By looking at O1s spectrum (fig. 3e) it can be inferred that the rest of the oxygen is bound to nitrogen forming N-O bond (530,3 eV). Overall, carbon and oxygen are undesirable but inevitable contaminants in ALD AlN their presence can be minimized by adjustments in the ALD process (temperature, pressure, precursor type) or by post-treatment of deposited samples (annealing, etching). Studying of oxygen impurities in AlN films is important since it is known to have negative effects on the resulting qualities of the film such as optical absorption and thermal conductivity [16].
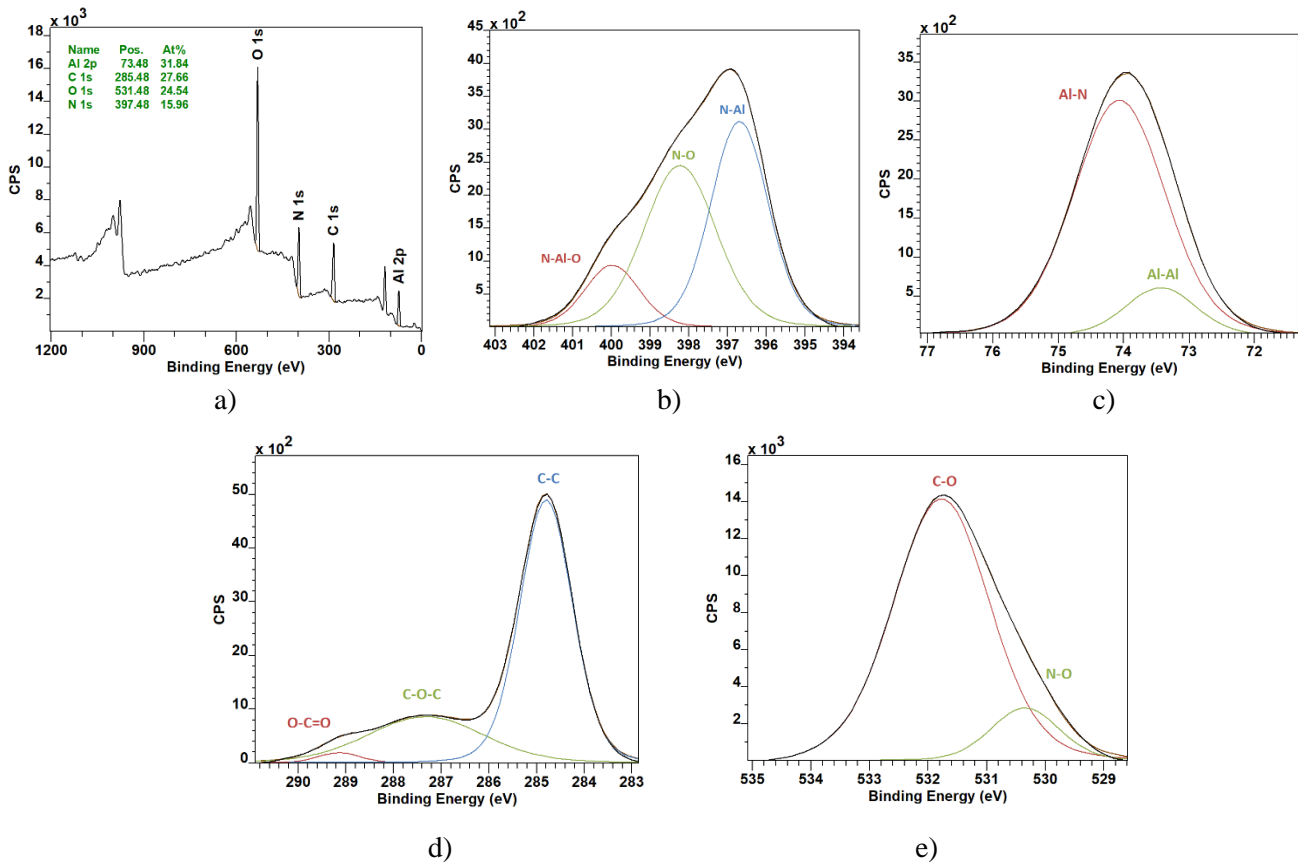

Figure 3. XPS high -resolution spectra for AlN on HOPG: a) wide spectrum, b) N1s, c) Al2p, d) C1s, e) O1s

4 CONCLUSION

In this paper we attempted to deposit AlN thin films using PE-ALD equipment on a previously untested substrate - HOPG. Surface analysis of the samples done by AFM indicates that deposited AlN layer exhibits quite uniform surface topography. Such methods as XPS and SIMS have been employed to investigate the chemical structure of the samples obtained. The data provided by these methods unambiguously confirms the presence of AlN compound, although not in the pure form. We can also observe formation of Al-Al and N-O defects as well as some impregnation of carbon in the XPS spectra which probably occur due to imperfection of the equipment's vacuum system or post-deposition surface oxidization. The removal of the undesired compounds from the films is the implication for future research, high-temperature annealing in nitrogen atmosphere might have a positive effect in this regard which is made possible thanks to carbon extremely high melting point. Presently, however, HOPG seems to be a promising substrate to be utilized in AlN deposition process.

REFERENCES

[1] A. Pandey, J. Kaushik, S. Dutta, A.K. Kapoor, D. Kaur, Electrical and structural characteristics of sputtered c-oriented AlN thin films on Si (100) and Si (110) substrates, Thin Solid Films. 666 (2018) 143–149. doi:10.1016/j.tsf.2018.09.016.

[2] L. Tian, S. Ponton, M. Benz, A. Crisci, R. Reboud, G. Giusti, F. Volpi, L. Rapenne, C. Vallée, M. Pons, A. Mantoux, C. Jiménez, E. Blanquet, Aluminum nitride thin films deposited by hydrogen plasma enhanced and thermal atomic layer deposition, Surf. Coatings Technol. 347 (2018) 181–190. doi:10.1016/j.surfcoat.2018.04.031.

[3] D.L. Ma, H.Y. Liu, Q.Y. Deng, W.M. Yang, K. Silins, N. Huang, Y.X. Leng, Optimal target sputtering mode for aluminum nitride thin film deposition by high power pulsed magnetron sputtering, Vacuum. 160 (2019) 410–417. doi:10.1016/j.vacuum.2018.11.058.

[4] I. Gablech, V. Svatoš, O. Caha, A. Dubroka, J. Pekárek, J. Klempa, P. Neužil, M. Schneider, T. Šikola, Preparation of high-quality stress-free (001) aluminum nitride thin film using a dual Kaufman ion-beam source setup, Thin Solid Films. 670 (2019) 105–112. doi:10.1016/j.tsf.2018.12.035.

[5] Y. Wu, C.H. Jia, W.F. Zhang, Growth of conductive and insulative highly-oriented aluminum nitride thin films using laser molecular beam epitaxy, Diam. Relat. Mater. 25 (2012) 139–143. doi:10.1016/j.diamond.2012.02.022.

[6] Y. Bian, M. Liu, G. Ke, Y. Chen, J. DiBattista, E. Chan, Y. Yang, Aluminum nitride thin film growth and applications for heat dissipation, Surf. Coatings Technol. 267 (2015) 65–69. doi:10.1016/j.surfcoat.2014.11.060.

[7] E. Österlund, J. Kinnunen, V. Rontu, A. Torkkeli, M. Paulasto-Kröckel, Mechanical properties and reliability of aluminum nitride thin films, J. Alloys Compd. 772 (2019) 306–313. doi:10.1016/j.jallcom.2018.09.062.

[8] R. Dallaev S. Stach, Ş. Ţălu, D. Sobola, A. Méndez-Albores, G.T. Córdova, L. Grmela, Stereometric Analysis of Effects of Heat Stressing on Micromorphology of Si Single Crystals, Silicon. (2019). doi:10.1007/s12633-019-0085-4.

[9]   A.I. Abdulagatov, S.M. Ramazanov, R.S. Dallaev, E.K. Murliev, D.K. Palchaev, M.K. Rabadanov, I.M. Abdulagatov, Atomic Layer Deposition of Aluminum Nitride Using Tris(diethylamido)aluminum and Hydrazine or Ammonia, Russ. Microelectron. (2018). doi:10.1134/S1063739718020026.

[10]   H. Zhang, C.L. Marshall, Atomic layer deposition: Catalytic preparation and modification technique for the next generation, Chinese J. Catal. 40 (2019) 1311–1323. doi:10.1016/s1872-2067(19)63321-8.

[11]   Y. Li, C. Zhang, X. Luo, Y. Liang, D. Wuu, C. Tin, X. Lu, Applied Surface Science Surface , structural and optical properties of AlN thin films grown on different face sapphire substrates by metalorganic chemical vapor deposition, Appl. Surf. Sci. 458 (2018) 972–977. doi:10.1016/j.apsusc.2018.07.138.

[12]   R. Dallaev, N. Papež, D. Sobola, S. Ramazanov, P. Sedlák, Investigation of structure of AlN thin films using Fourier-transform infrared spectroscopy. In Procedia Structural Integrity. Procedia Structural Integrity. Elsevier B.V., 2020. s. 601-606. ISSN: 2452-3216.

[13]   P. Kaspar, D. Sobola,; R. Dallaev, S. Ramazanov, A. Nebojsa, S. Rezaee, L. Grmela, Characterization of Fe2O3 thin film on highly oriented pyrolytic graphite by AFM, Ellipsometry and XPS. Applied Surface Science, 2019, roč. 493, č. 1, s. 673-678. ISSN: 0169-4332.

[14]   F. Dwikusuma, T.F. Kuech, X-ray photoelectron spectroscopic study on sapphire nitridation for GaN growth by hydride vapor phase epitaxy: Nitridation mechanism, J. Appl. Phys. 94 (2003) 5656–5664. doi:10.1063/1.1618357.

[15]   P. Motamedi, K. Cadien, XPS analysis of AlN thin films deposited by plasma enhanced atomic layer deposition, Appl. Surf. Sci. 315 (2014) 104–109. doi:10.1016/j.apsusc.2014.07.105.

[16]   L. Rosenberger, R. Baird, E. McCullen, G. Auner, G. Shreve, XPS analysis of aluminum nitride films deposited by plasma source molecular beam epitaxy, Surf. Interface Anal. 40 (2008) 1254–1261. doi:10.1002/sia.2874.

[17]   H. Kim, N. Do, S. Chul, H. Ju, B. Joon, Improved interfacial properties of thermal atomic layer deposited AlN on GaN, Vacuum. 159 (2019) 379–381. doi:10.1016/j.vacuum.2018.10.067.

[18]   M. Zhu, P. Chen, R.K.Y. Fu, W. Liu, C. Lin, P.K. Chu, AlN thin films fabricated by ultra-high vacuum electron-beam evaporation with ammonia for silicon-on-insulator application, Appl. Surf. Sci. 239 (2005) 327–334. doi:10.1016/j.apsusc.2004.05.287.

[19]   O.H. Kim, D. Kim, T. Anderson, Atomic layer deposition of GaN using GaCl3 and NH3, J. Vac. Sci. Technol. A Vacuum, Surfaces, Film. 27 (2009) 923–928. doi:10.1116/1.3106619.

[20]   F. Jose, R. Ramaseshan, S. Dash, S. Bera, A.K. Tyagi, B. Raj, Response of magnetron sputtered AlN films to controlled atmosphere annealing, J. Phys. D. Appl. Phys. 43 (2010) 075304. doi:10.1088/0022-3727/43/7/075304.

[21]   N.C. David, D. Anavi, M. Milanovich, Y. Popowski, L. Frid, E. Amir, Preparation and properties of electro-conductive fabrics based on polypyrrole: Covalent vs. non-covalent attachment, IOP Conf. Ser. Mater. Sci. Eng. 254 (2017). doi:10.1088/1757-899X/254/3/032002.

PROFILOMETRY MEASUREMENTS OF MAGNETIC PLATES MOUNTED ON ROTORS

Radoslav Mach

Doctoral Degree Programme (1), FEEC VUT

E-mail: xmachr00@stud.feec.vutbr.cz

Supervised by: Pavel Škarvada

E-mail: skarvada@feec.vutbr.cz

Abstract: Presented project deals with optical profilometric analysis of magnetic plates position within the rotors surface. The focus of this work is to determine whether measuring by laser profilometry scanning of the surface is suitable for such an issue. Main problem is occasional inaccurate positioning of magnetic plates which might occur in gluing of the plates during manufacturing process. This could lead to inhomogeneous magnetic field causing dysfunctional properties of the device. Results demonstrate that profilometric measurements are sufficient option for solving the problem, nevertheless there is need for special equipped profilometric system to measure and evaluate data properly.

Keywords: laser profilometry, rotor, optical measurements, flatness analysis, magnetic plates

1   INTRODUCTION

Currently, research aimed at measuring and recognizing the three-dimensional shape of objects and surfaces is advancing by leaps and bounds. In many industries and applications of laboratory or industrial measurement, it is necessary to accurately evaluate even the most subtle deviations in the shape of profiles of various surfaces. The use of modern ways to study the properties of materials can shed new light on what has already been explored and open up new horizons.

Based on the measurement requirements and working principles, real-time 3D shape measurement techniques could be classified into three categories: time of flight technique, stereovision and structured light technique. [1] [2]

In case of this project we worked with structured light generated by non-contact optical profilometric system which uses laser as a light source. The measured object is magnetic rotor. Surface of the component is covered by several magnets which mutual flatness level is crucial during operation. Magnets have to be aligned to each other for creating accurate magnetic flux and correct functionality.

As rotors are manufactured, magnets are gradually glued around the unit. Without following optical control of the magnet plates mounting quality level this process may lead to higher probability of failure of the products.

Main asset of this project is to demonstrate option of laser profilometry as an optical control to detect wrong rotors surface anomalies created during manufacturing.

2   LASER PROFILOMETRY

In terms of the use of modern technologies in production process, 3D scanning technologies are currently gaining ground. These technologies are mainly used in the control of products, surfaces and evaluation of their parameters, reverse production process or quality control.

Laser profilometry is a form of non-contact optical profilometric measurement of the surface of a scanned object with an option of the subsequent execution of its three-dimensional model. The composition of such a profilometric system is based on the presence of a laser beam emitted by an optical source of electromagnetic radiation and a scanning device in the form of a camera, usually based on a CCD or CMOS scanning chip.

Laser generates beam in shape of a dot which is subsequently formed into a narrow line by using of optical projection lens. The light trail is captured at an angle by a digital camera. The principle of the function consists in the use of the technique of active triangulation where the light source together with the sensor and the examined object form a so-called triangulation triangle. In this way, it is possible to create a real 3D profile from each scanned image. Such a contactless profilometer cooperates with a classic computer with the Windows operating system via a pre-supplied profilometric program with a user environment. The IEEE 1394 (FireWire) interface is used to transmit and communicate with the computer. [3]

Application of this form of measurement can be found practically in any sphere of industry where there is an emphasis on the quality of the surface of objects and their shape. This technology allows us to control in real time, which is a very important advantage in terms of production efficiency.
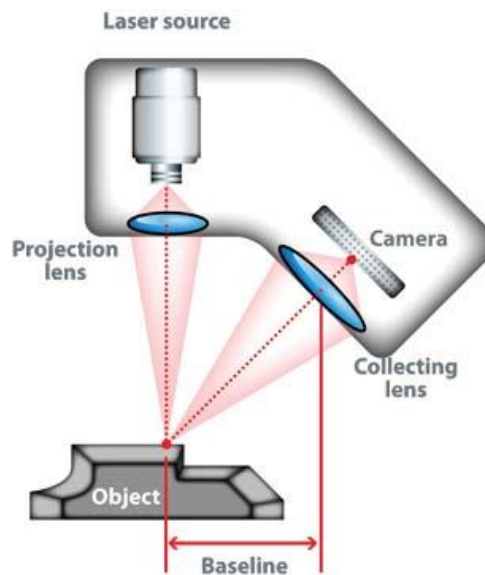

Figure 1: Laser profilometry principle illustration [4]

3   EXPERIMENT

The experiment used a Microepsilon Scancontrol laser line profilometer with a catalogue resolution of 10 μm. [5] The principle of measurement is an active triangulation performed by laser profilometric system.

In this case, several factors and settings affect the achievable resolution. Surface roughness higher than 5 μm can lead to laser beam interference and surface noise. In addition, the reflectivity and colour variation of the surface can affect, but also the associated inappropriate exposure time settings. The arrangement of the experiment or the attachment of the sample is one of the key factors when it is necessary to minimize vibrations.

The measurement equipment did not dispose of a holder with possibility of computer-controlled rotation, therefore it was not possible to create a 3D image of the surface, which would allow to obtain the most accurate desired results. Instead, the rotors with magnets were placed in wooden V-grooves and the rotation was done manually. This arrangement is limited to low resolution results and can lead to a reduction in measurement repeatability.

Figure 2: Photo of a rotor during performing of an experiment

Another problem that arises from the arrangement is the fact that it is not possible to reconstruct the planes of the glued magnets but only their inclination in one axis (along the longer side), and only in the middle of the magnet. There is possibility that the corner of the magnet may be significantly higher or lower than measured because of the tilt in the second axis, that is not considered in the performed experiment. Thus, it is possible that essential problematic pieces of magnets were not captured in the performed measurement.

Figure 3: Workplace arrangement
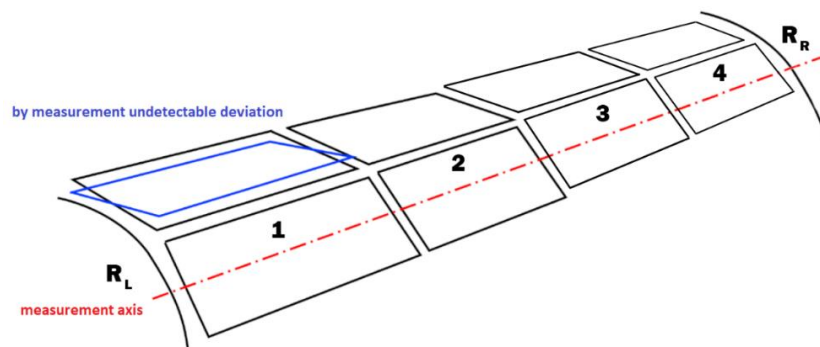
Figure 4: Axis measurement

Experiment was executed for two rotors sample coils – A and B. Each of them consisting of several rows of glued magnetic plates.

Since magnets 1 and 2 are offset from magnets 3 and 4 as shown in Figure 4, the measurement was performed specially for each magnet. Each magnets level deviation is presented in Figure 5. Measurements were performed 40x for each measurement axis to minimize vibration influence. As part of the evaluation, a linear function interpolation was performed for individual magnets 1 - 4 and the inclination of the individual magnets was subtracted in proportion to the inclination of the reference level formed by the sections R_L and R_R. Furthermore, the distance of its centre from the reference level was determined for each magnet. These deviations were plotted (on the x-axis the proportion of guidelines, on the y-axis the deviation of the position). From the above, it can be seen from the graph that the deviations of the centres of the magnets from the reference axis are in the range up to $\pm$ 150 μm. The guidelines differ by a coefficient of 0.7-1.5. If we consider approximately the normal distribution of values in both axes, we can determine the standard deviations from the arithmetic mean and define the boundary of the criterion 3-sigma.
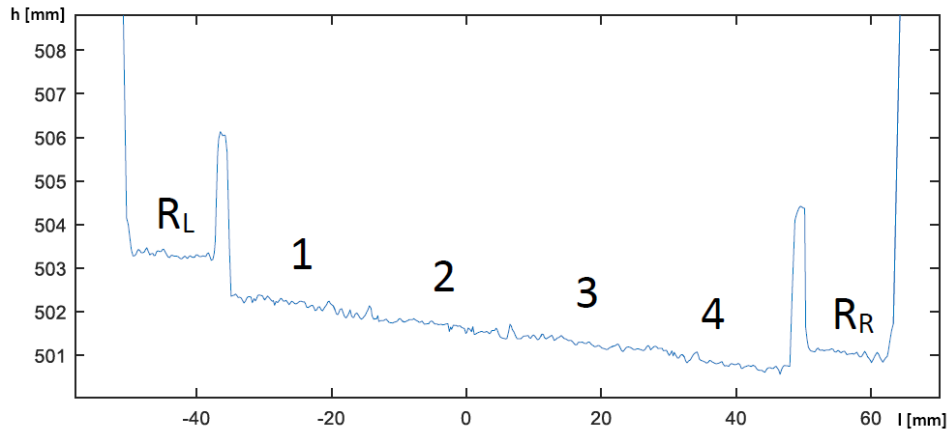

Figure 5: R_L, R_R - reference level ; 1, 2, 3, 4 - row of magnetic plates

The y-axis of the Figure 5 describes distance of the measured profile from a measuring device and it is used to define k and q parameters which are specified by the slope equation $y = kx + q$.
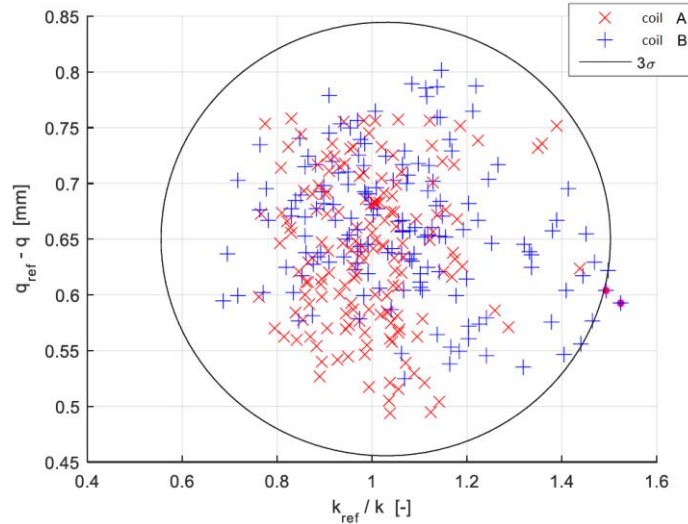

Figure 6: Variance of values for A and B sample [6]

Sample B, which was exceeded in two measurements, shows a larger variance of values defined boundaries and at least three other measurements are boundary.

4 CONCLUSION

To confirm the experiment and find problem areas, it is essential to create a sample holder with the possibility of computer-controlled rotation with minimal deviations. Only this arrangement would make it possible to measure the true position of the entire magnet in space relative to the rotor axis and to reliably determine the measurement uncertainty. Due to the absence of such a holder, the performed measurement was performed by a reference method against a rotating edge and without assessing the repeatability of the measurement. The sample was placed in a wooden V-groove with manual rotation. This could be a significant source of measurement inaccuracies.

From the evaluation, it appears to be a worse sample B, but it cannot be ruled out that the problematic magnets were not detected because the position / inclination of the magnets was measured in only one axis (along the longer side) and only in the centre of the magnet.

The resolution of optical profilometers seems to be sufficient for identification of defective pieces, but the key factor is the arrangement of measurements. In particular, axially accurate fixation of the sample in a sufficiently rigid holder and controlled rotation of the sample is essential for repeatable and reliable measurements.

REFERENCES

[1] S. V. d. Jeught, J. J. J. Dirckx, "Real-time structured light profilometry: a review", Journal of Optics and Lasers in Engineering, vol. 87, pages 18-31, Dec. 2016

[2] Z. Wong, "Review of real-time three-dimensional shape measurement technique", Journal of Measurement, vol. 156, no. 107624, May. 2020

[3] MACH, R. Meranie drsnosti laserovou profilometriou. Brno: Vysoké učení technické v Brně, Fakulta elektrotechniky a komunikačních technologií. Ústav mikroelektroniky, 2016. 39s.

[4] Accessible from WWW: https://www.pinterest.se/pin/9710955429684354/

[5] Accessible from WWW: http://sentec.jp/products/pdf/LLT2700_2750_2710.pdf

[6] MATLAB. (2020). version 9.9.0 (R2020b)

RECOVERY OF CATHODE MATERIAL
FROM Li-ION BATTERIES

Patrícia Guricová

Doctoral Degree Programme (1), FEEC BUT

E-mail: xguric01@stud.feec.vutbr.cz

Supervised by: Tomáš Kazda

E-mail: kazda@feec.vutbr.cz

Abstract: This article focuses on the possibilities of recycling materials from spent Li-ion batteries, especially on direct recycling. Lithium-ion batteries (LIBs) are gradually replacing other types of accumulators from practical applications and further increase in their production is expected in the future. With high production, bigger amount of waste will have to be deal with. Recycling processes currently in use are not efficient enough and they have been developed mainly for one type of cathode material. Direct recycling is very promising in terms of efficiency, but it needs to be further enhanced. Direct recycling technology is discussed in the text as well as the recovery process for commercial Samsung 18650B-20R aged cell using different kinds of solvents.

Keywords: Li-ion battery, direct recycling, NMC

1   INTRODUCTION

Today, LIBs are used in many fields, from wearable electronics, smart phones, laptops, electromobiles to energy storage power stations. LIBs are in their principle, as well as other battery types, an electrochemical energy source. Due to the fact that they can accumulate energy in their electrodes, they are classified as secondary sources of energy.

LIBs are made of three main parts, cathode, anode and electrolyte. The cathode consist of aluminium collector and active material containing transition metal oxides and lithium, which ensures the battery function. The working principle of LIBs is called intercalation process. During the charging of the battery, lithium ions are deintercalated from the structure of cathode and migrate through the electrolyte to the anode, where they intercalate to the graphite. During discharging the process is reversed. Therefore, Li-ions are also referred to as rocking chair batteries [1], [2].

During the life cycle of the battery, due to the mentioned charging and discharging processes, there is a gradual decrease in capacity. Because of ions drifting between two electrodes, a certain deformation in the material structure occurs, capacity is decreasing leading to the end of life of the battery. Another reason for capacity loss is growing of SEI layer.

2   RECYCLING PROCESS OF LI-ION BATTERIES

Today, the efficiency of recycling is about 50%, and the processes themselves focuses on recycling the cathode material, which contains many important elements like lithium, cobalt, nickel, manganese etc.. Recycling is interesting mainly from the economic point of view, total energy costs and total emissions are reduced [3].

Recycling technologies of LIBs essentially consist of physical and chemical processes, with physical processes mostly used in the pre-treatment of batteries to separate the cathode material from the used LIBs. The chemical processes themselves are then used to recover the cathode material. Nearly all of the recycling techniques were created to recycle LCO ($LiCoO_2$) chemistry concentrating on the metals contained [4].

There are three main technologies we distinguish in recycling, namely pyrometallurgical, hydrometallurgical processes and direct recycling. The main issue in recycling is the diversity of compounds which can be used in LIBs as active material and for which appropriate separation process must be designed.

Pyrometallurgical process is based on high temperature combustion. Energy consumption is very high and the second drawback is the emission of toxic gases which are produced during combustion. The material obtained afterward can be considered as a sort of intermediate product, meaning that he must be refined or otherwise processed. The biggest advantage remains the high recovery factor for cobalt. Second, hydrometallurgical way is a more environmentally friendly, leaching in various acids or bases is used. Leaching process must be adapted to the given battery chemistry. Using this method, we obtain precursors ready for the preparation of cathode material. In a lot of cases companies combine these two approaches to take advantage of both. The third type, direct recycling, is apart from the previous ones not based on separating the material into individual elements but tries to regenerate the active material as a whole [5], [6].

2.1 DIRECT RECYCLING

Effort is made to create a recycling method that would be less harmful to the environment and the obtained material could be directly used for the production of a new battery cell. Direct recycling is reaching this idea. Although it is still at level of research and not yet used commercially, it could theoretically give us high recycling efficiency rate (higher than 80%). It has the least steps to get final product, which can directly be used as a cathode material. In contrast to the other two methods mentioned, direct recycling is based on reactivation of capacity and other properties that were lost during life cycle of a battery. Defects in the active material can be fixed by relithiation process, meaning adding Li_2CO_3 to the final product. Direct recycling cannot yet restore original properties of the materials fully, so the recycled material probably will not be as effective as it was in the beginning. Main advantages remain in less pollution and lower energy consumption. Also almost all of the materials from the accumulator can be restored, including aluminium [6], [7].

Main method used in direct recycling is separating the active material from its current collector by using some sort of solvent, which can dissolve binder that hold everything together. In most of the cases NMP and DMSO are trusted organic solutions, effectively dissolving PVDF, which is commonly used binder in LIBs. One of the approaches uses cyrene as a solvent for separation the cathode powder from collector. It is a solvent derived from cellulose waste, which could be good alternative for toxic NMP and DMF. Another advantage is that cyrene itself can be recycled and used again [4], [8].

3 RECOVERY OF CATHODE MATERIALS

For this work, the battery which we chose to extract the cathode material from, was commercial Samsung INR 18650-20R (2 000 mAh) with NMC ($LiNiMnCoO_2$) chemistry. Rate capability test was performed using a different C-rates between 0.1 C and 1 C before and after long term galvanostatic cycling for 500 cycles at 1 C at 80% depth of discharge (DOD). The capacity of the battery after 500 cycles decreased by 9%.

The degraded cell was firstly discharged, and then dismantled in a hood. All parts of the battery were measured and weighted. We can see bill of materials that made up the battery in the **Figure 1**. Understandably, the biggest part of the battery constitutes of cathode (32.8 wt.%) and then the anode material (21.2 wt.%). Aluminium casing is on the third place taking up 15.8 wt.% of the battery mass and copper collector taking up 12.1 wt.%. The amount of electrolyte was determined as the deviation of the weights of individual dried components together from the weight of the whole battery.
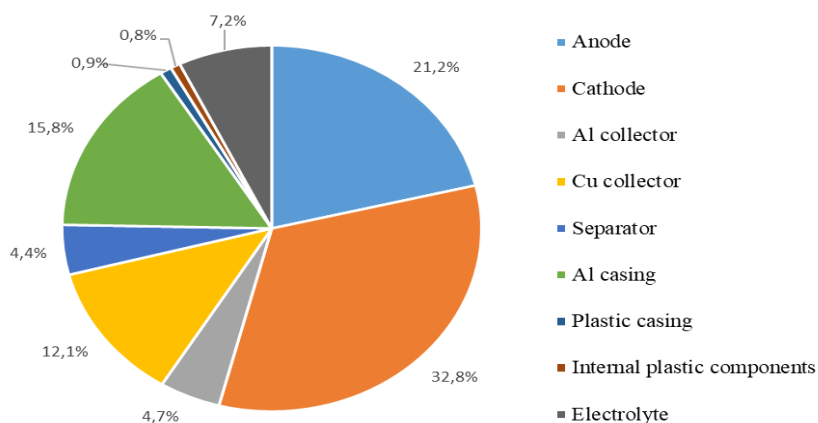
Figure 1: All parts of lithium-ion cell (in wt.%)

3.1 EXTRACTION OF ACTIVE MATERIAL

In this work, we investigated influence of the temperature to the dissolution process of cathode materials in four selected solvents, namely dimethylsulfoxide (DMSO), n-methylpyrrolidone (NMP), cyrene and ethyl acetate. DMSO and NMP were picked because they are widely used in industry. In the effort to find an alternative mainly to the toxic NMP, cyrene and ethyl acetate were picked as a promising candidates.

Cathode scraps were prepared and first test started at room temperature. Samples were continuously checked after 1, 2, 5, 10 and 24 hours. Same procedure was then repeated with the temperature at 50 °C, 75 °C, 100 °C and lastly 120 °C.

Using DMSO as a solvent at room temperature was not efficient, the material was almost unaffected, as it was with the other solvents also. Although when heated up to 50 °C, the material started to separate from the collector after only one hour. Visible waviness was spotted on the surface as the layer was detaching itself. After another hour in the solution, most of the material fell off the foil, and the rest was removed using scalpel.

As for the NMP, it was the only solvent that could be used without using heating. At room temperature small pieces of active substance started to fell into the solution, and after five hours it was possible to remove it entirely from aluminium. With the temperature rising to 50 °C the process speeded, and it took active material only 2 hours to detach. No speeding in the procedure was observed with further raising the temperature. At 75 °C it also took 2 hours, but the substance was more dissolving into the solvent rather than detaching itself as a whole piece. It was necessary to filtrate the solution afterwards.

Cyrene was chosen as a green solvent and so an alternative not only to NMP but also DMSO. The results in the beginning were not promising. Until 75 °C almost nothing was happening with the material and it could not been scraped off, or it was much harder than with the other solvents. The temperature needed to rise up to 75 °C in order to remove the material, although it took 10 hours to release the bonding, 24 hours to completely separate. The big difference was observed at 120 °C, when only after 1 hour the layer started to break. From the places where the surface was little corrupted it was easy to remove, complete removal took 24 hours.

The last chosen solvent was ethyl acetate as it is less harmful for environment. This solvent was found out to be not as effective as the others. It evaporates very quickly and has boiling point of 77 °C. After crossing this temperature a little fall out from the scrap was observed, but it was probably due to boiling of the solvent which caused breaking of the cathode layer. There were visible cracks in the surface, the material was not dissolving in the solvent. Experiment was finished with the temperature of 100 °C, because it was hard to manipulate because of boiling solution.
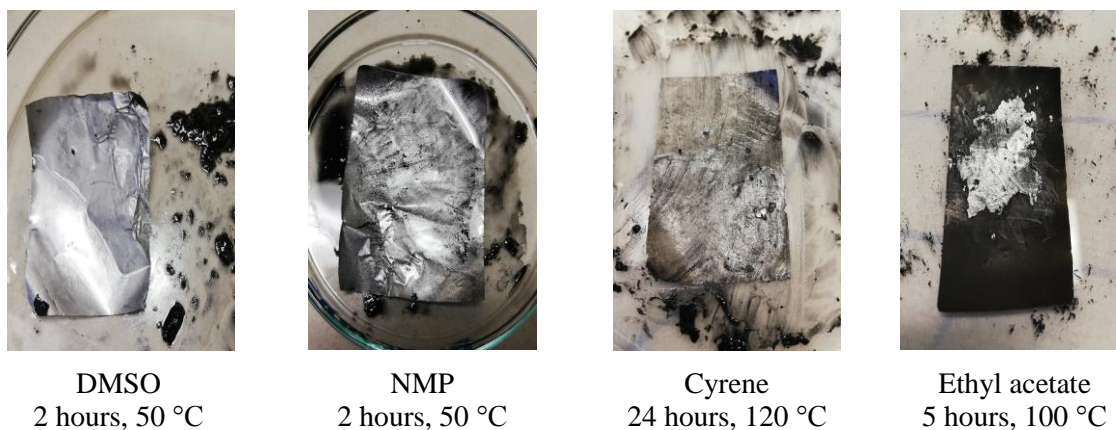
DMSO
2 hours, 50 °C

NMP
2 hours, 50 °C

Cyrene
24 hours, 120 °C

Ethyl acetate
5 hours, 100 °C

Figure 2: Cathode scraps after dissolving in different solvents

In the **Figure 2** we can see photos of the cathode scraps after they were removed from the solvents mentioned above. It captures the samples at the time and temperature at which the best results were obtained for the specific solvent.

The structure of the cathode material was also observed using XRD analysis. The aim was to find out how the material would change when exposed to air for longer period of time and if it will make any structural changes. The studied material was the one obtained from scraps immersed in NMP for two hours at the 50 °C. Due to analysis, the material was identified as NMC111, meaning that molar ratios between the metals are 1:1:1, formula $Li_{0.84}Ni_{0.333}Mn_{0.333}Co_{0.333}O_2$. This material was found to have less than 1% of impurities, which most likely are formed by Co_3O_4 (0.61%) based on structure analysis. Sample was measured during four week period, each measurement took place after one week. In the **Figure 3** we can see the comparison of all four diffractograms. As we can see, there are no considerable changes in the structure during long time measurement. This information is important in the sense of storage of the recycled material. Also, if the battery would undergo some damage, there would be no changes in the material itself, as it is not reactive and do not degrade.
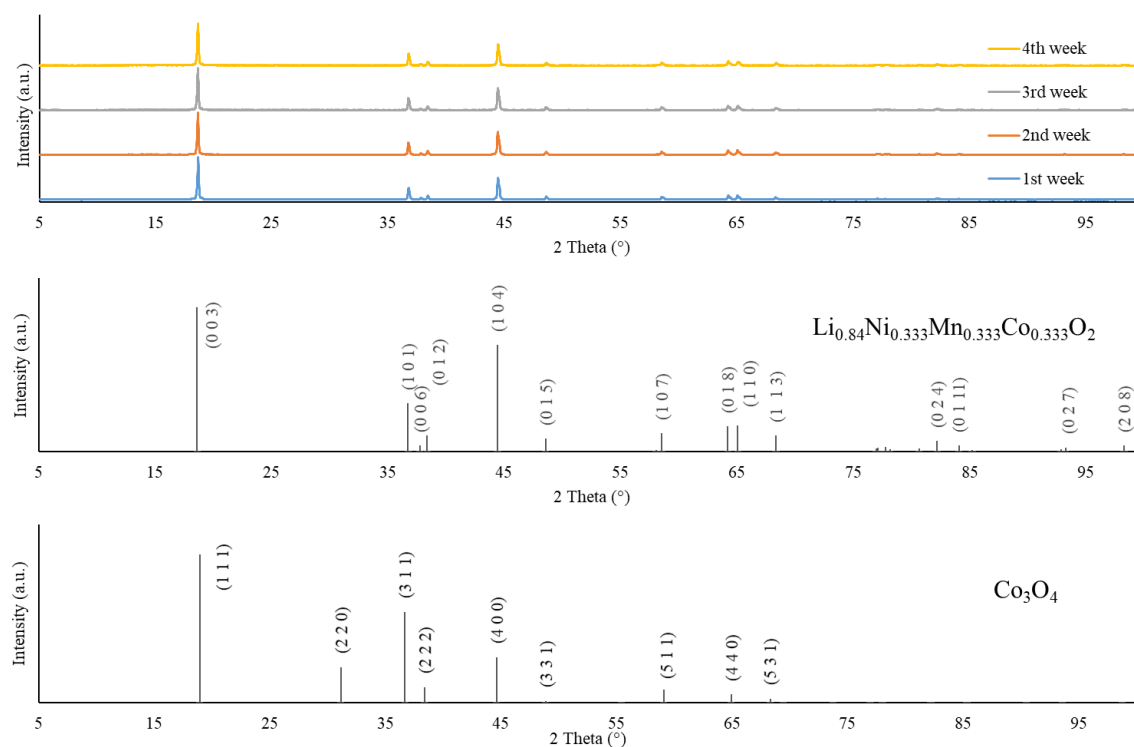

Figure 3: Diffractograms for NMC material

4 CONCLUSION

Direct recycling as one of the approaches for recycling of spent Li-ion cells is discussed. With high recycling rate and lower ecological impact could replace economically and energetically demanding pyro and hydrometallurgical processes. Four different solvents were demonstrated to separate active cathode material from current collector. NMP and DMSO were most sufficient, NMP working without using heat. Since NMP has many disadvantages related to environment and health, using DMSO would be a better option. It breaks the binder similarly to NMP solvent, although slightly higher temperature must have been used (50 °C). Cyrene performed well and successfully separated active layer, but the lowest temperature was 120 °C. Material structure was analysed using X-ray diffraction and was defined as NMC111. The material appeared to be stable, does not react with air and no considerable degradation was observed, which is favourable material feature.

REFERENCES

[1] Lithium-ion battery overview. KORTHAUER, Reiner. *Lithium-Ion Batteries: Basics and Applications*. 1. Berlin: Springer Nature, 2018, s. 13-19. ISBN 978-3-662-53071-9.

[2] KAZDA, Tomáš. *Modifikace materiálů pro kladné elektrody lithno-iontových akumulátorů*. Vysoké učení technické v Brně. Fakulta elektrotechniky a komunikačních technologií, 2015.

[3] PAGLIARO, Mario a Francesco MENEGUZZO. Lithium battery reusing and recycling: A circular economy insight. *Heliyon* [online]. Elsevier, 2019, **5**(6), 1-7. ISSN 2405-8440. Dostupné z: doi:10.1016/j.heliyon.2019.e01866

[4] ZHANG, Xihua, Yongbing XIE, Hongbin CAO, Faheem NAWAZ a Yi ZHANG. A novel process for recycling and resynthesizing LiNi1/3Co1/3Mn1/3O2 from the cathode scraps intended for lithium-ion batteries. *Waste management (Elmsford)* [online]. Elsevier, 2014, **34**(9), 1715-1724. ISSN 0956-053X. Dostupné z: doi:10.1016/j.wasman.2014.05.023

[5] WANG, Yuqing, Ning AN, Lei WEN, Lei WANG, Xiaotong JIANG, Feng HOU, Yuxin YIN a Ji LIANG. Recent progress on the recycling technology of Li-ion batteries. *Journal of energy chemistry* [online]. Elsevier B.V, 2021, **55**, 391-419. ISSN 2095-4956. Dostupné z: doi:10.1016/j.jechem.2020.05.008

[6] PIĄTEK, Jędrzej, Semih AFYON, Tetyana M. BUDNYAK, Serhiy BUDNYK, Mika H. SIPPONEN a Adam SLABON. Sustainable Li- Ion Batteries: Chemistry and Recycling. *Adv. Energy Mater.* 2020, 2003456. https://doi.org/10.1002/aenm.202003456

[7] KIM, Hee-je, Kamran ZEB, Vinodh RAJANGAM, Sangaraju SAMBASIVAM a Ihab OBAIDAT. A Comprehensive Review of Li-Ion Battery Materials and Their Recycling Techniques. *Electronics* [online]. Basel: MDPI, 2020, **9**(7), 1161. Dostupné z: doi:10.3390/electronics9071161

[8] BAI, Yaocai, W. Blake HAWLEY, Charl J JAFTA, Nitin MURALIDHARAN, Bryant J POLZIN a Ilias BELHAROUAK. Sustainable recycling of cathode scraps via Cyrene-based separation. Sustainable Materials and Technologies [online]. Elsevier B.V, 2020, 25, 1-7. ISSN 2214-9937. Dostupné z: doi:10.1016/j.susmat.2020.e00202

ELECTROHYDRODYNAMIC MODEL OF ELECTRON MICROSCOPE

Martin Mačák

Doctoral Degree Programme (3), FEEC BUT

E-mail: xmacak00@stud.feec.vutbr.cz

Supervised by: Petr Vyroubal

E-mail: vyroubal@feec.vutbr.cz

Abstract: This work presents a complex multiphysics model of an electron microscope in ANSYS Fluent software. A custom electromagnetic model was used to describe the relativistic transport of electrons. The model consisted of a description of an electromagnetic field, relativistic transport of charged particles and interactions between charged particles and solid materials. Presented results suggest, that the custom model can be coupled with the CFD module, which results in a possibility of coupled simultaneous simulations of hydrodynamics, electromagnetics, and a transport of charged particles.

Keywords: Electron microscope, numerical modelling, ANSYS Fluent, relativistic transport

1 INTRODUCTION

An Electron microscope is a device used for studying a micro-structure, surface, and a chemical composition of specimens by an electron beam. Nowadays, the technology is well established in many industrial and scientific areas. Still, the design of an electron microscope is a very complex process, which combines many areas of physics. Experimental studies, which could help with designing and optimising electron microscopes, are often very difficult due to extreme operational conditions. On the other hand, numerical simulations are often able to study these processes, as they only require knowledge of the geometry and general operating and boundary conditions. Currently, the numerical simulations in electron microscopy are often focused only on one area of physics (electromagnetics, particle-matter interactions, or hydrodynamics) [1, 2, 3]. Usually, the individual investigation of one specific area will not bring any inconsistencies in the results, however in some cases the multiphysics nature of these processes might be very important. This is the case for environmental scanning electron microscopy (ESEM), which uses a higher pressure in the specimen chamber, which allows for the scanning of organic materials. For these microscopes, it is necessary to consider electron transport in the vacuum as well as its interactions with residual gas and the specimen. A huge amount of research is focused on the area of gas-particle interactions as it is a very complex phenomenon. However, the transport of electrons from the electron gun and the gas flow in the differentially pumped chamber is usually ignored [2, 3].

This paper presents a custom numerical model implemented into a CFD software ANSYS Fluent, which was shown to be able to successfully model the rarefied gas flow in an ESEM [3]. The custom model consists of a description of an electromagnetic field and a description of a relativistic electron transport. This model was used for simulation of a simplified electron microscope, which included generation of electrons, their focusing and their interaction with a specimen as well as the rarefied gas flow.

2 NUMERICAL MODEL

The presented numerical model consists of an in-built CFD module and a custom model describing electromagnetic field with relativistic electron transport. This model was used to study processes in

a simplified electron microscope, which is described in Figure 1. The microscope consisted of a thermionic electron source, two focusing coils, differentially pumped chamber, scintillation detector and a specimen. The radius of the filament tip was 0.25 mm. The internal diameter of the coils was 26 mm, the external diameter was 78 mm, and the height was 65 mm. The diameter of all apertures was 0.5 mm. The distance between the filament tip and the specimen was 250 mm.
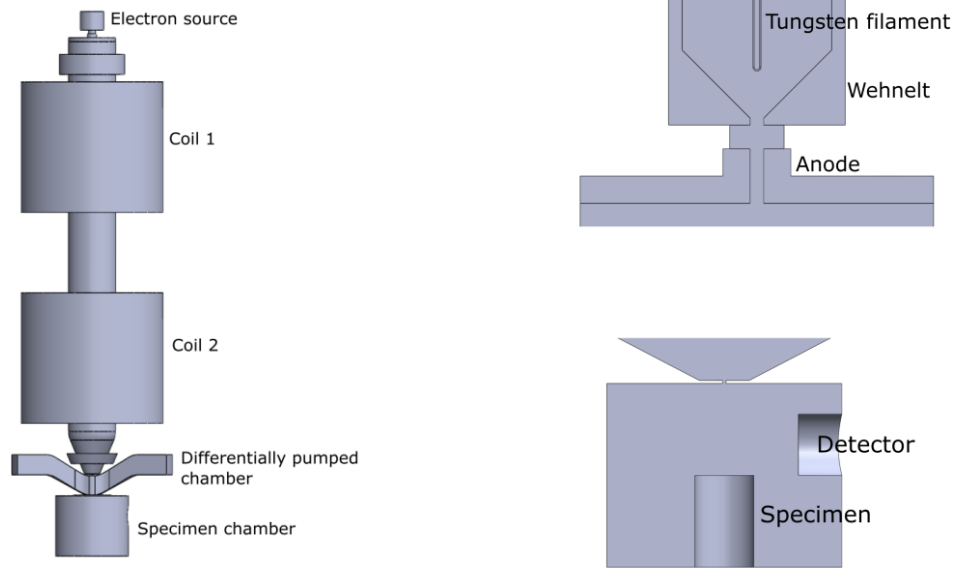

Figure 1: Simplified geometry of an electron microscope

2.1 GAS FLOW

Generally, any CFD software is based on the continuum assumption, which uses Navier-Stokes equations to describe a general fluid flow [4, 5]. As the pressure in the electron microscope is low, the gas might be so rarefied that the continuum assumption might not be applied. The Knudsen number, which describes the gas dynamics regime was calculated as in [6], in which the characteristic length was calculated locally as a ratio between density and the density gradient. The value of Knudsen number in this case was in range of $0 - 0.1$ (maximum behind the apertures). These values represent continuum flow as well as the slip flow regime, which is still defined by Navier-Stokes equations. The adjustment considering the slip lies in the application of the Maxwell slip boundary for the velocity and the Smoluchowski temperature jump condition [7]. These boundary conditions were applied using a FLUENT's in-built option, which is available for laminar flows. In this case, the Reynolds number was sufficiently low, so the laminar flow assumption was valid. In the electron optics area, the vacuum is so high that the continuum assumption cannot be applied. To avoid this complication, the flow in this area was ignored and only constant pressure was considered. The gas density was described by the ideal gas law, while the viscosity was defined by the kinetic theory. The pressure boundary conditions were set as: 500 Pa in the specimen chamber, 30 Pa in the differentially pumped chamber and 0 Pa in the electron optics area. This simplification was used as this study describes only a general microscope and the exact vacuum pump parameters were not known.

2.2 ELECTROMAGNETIC FIELD

ANSYS Fluent offers a possibility of implementing custom user defined functions and user defined scalars (UDS), which are variables defined by a general transport equation. The electromagnetic field was described using an electric potential φ [V] and a magnetic vector potential A [V·s·m^{-1}]. These equations were implemented in a steady state manner using UDSs as [8]:

$$-\nabla^2 \varphi = \frac{\rho_e}{\varepsilon} \tag{1}$$

$$-\nabla^2 \boldsymbol{A} = \mu \boldsymbol{J} \tag{2}$$

Where is the charge density ρ_e [C·m^{-3}], ε is the permittivity [F·m^{-1}], μ is the permeability [H·m^{-1}] and $\boldsymbol{J}$ is the current density [A·m^{-2}].

The accelerating voltage was set to 10 kV (-10 kV at the tungsten filament and 0 V at the anode). The simplified scintillator was defined by a boundary condition of 300 V. The magnetic field of the coils was generated by total current density of $5 \cdot 10^7$ A·m^{-2} for the Coil 1 and $4.3 \cdot 10^7$ A·m^{-2} for the Coil 2.

2.3 ELECTRON TRANSPORT

Even though ANSYS Fluent includes a discrete phase model, which describes the transport of particles, it cannot be directly used for the simulation of the movement of charged particles as it always assumes that particles are in a flowing fluid. The original equation, which considered a constant mass of the particles and the drag force was adjusted to consider relativistic effect and the influence of the electromagnetic field [9]. The final equation of the relativistic electron transport was described as:

$$\frac{\partial \boldsymbol{v}}{\partial t} = \frac{q(\boldsymbol{E} + \boldsymbol{v} \times \boldsymbol{B})}{m_0 \left(\frac{v^2}{c^2} \gamma_L{}^3 + \gamma_L \right)} \tag{3}$$

Where t is time [s], q is the charge [C], $\boldsymbol{E}$ is the electric field intensity [V·m^{-1}], $\boldsymbol{B}$ is the magnetic flux density [T], m_0 is the resting mass [kg], γ_L is the Lorentz factor [-].

The current density generated by the thermionic electron source was described by the Richardson-Dushmann equation, in which the material was defined as tungsten with a temperature of 2800 K [1]. Additionally, it was considered that primary electrons will release secondary electrons from the specimen. It was estimated that these secondary electrons had a random energy from 1 eV to 10 eV with a random direction pointing away from the specimen [10].

3 RESULTS

The simulation of particle trajectories in the studied electron microscope is shown in Figure 2. Without the focusing magnetic field, the particles naturally diverge, they do not reach the specimen chamber and are absorbed on the microscope walls. In the second case, electromagnetic coils are able to focus the electron beam on the specimen.
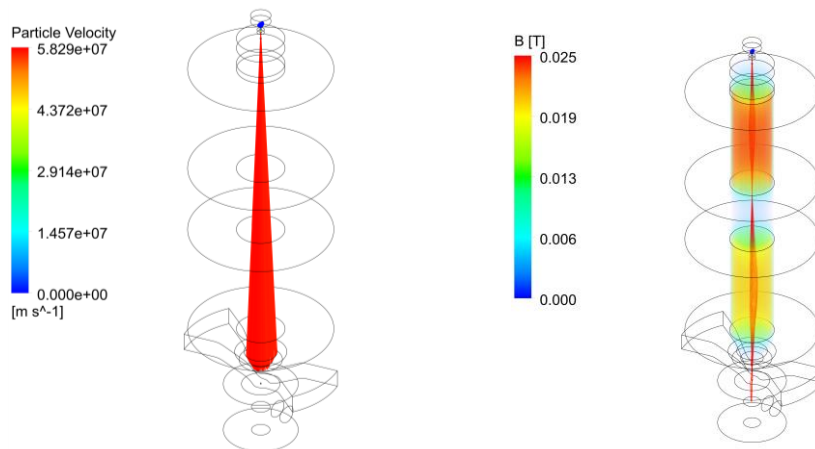

Figure 2: Electron trajectories without focusing coils (left) and with the magnetic field (right).

Figure 3 (left) shows the detail of electron source. It is possible to see the acceleration of electrons between the filament tip and the anode. As the particles are released from the filament, they are converging towards the crossover point and afterwards they start to diverge. Figure 3 (right) shows the detail of electron trajectories in the specimen chamber. While most of the primary electrons are scattered back at low angles, secondary electrons are released with random direction. Due to their low energy, these electrons are attracted by a simplified scintillation detector. Still, the low potential (300 V) at the detector cannot attract all secondary electrons.
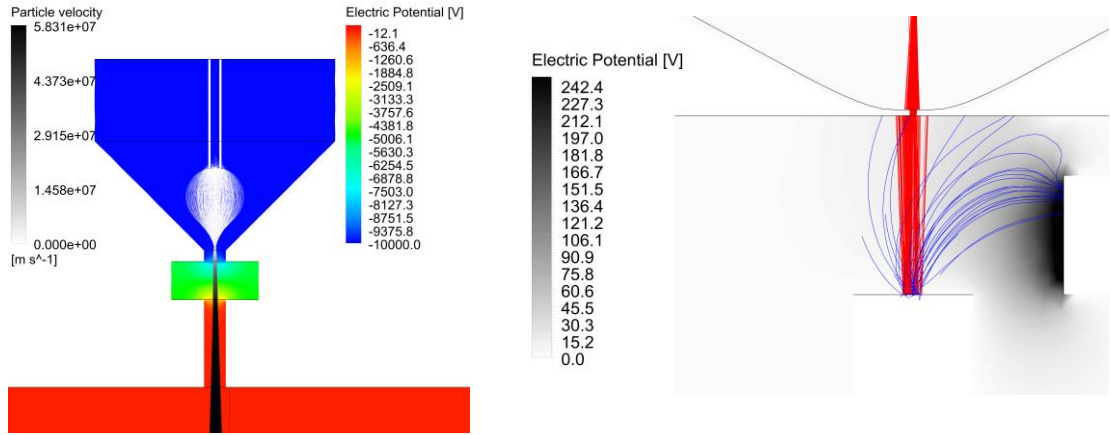
Figure 3: Detail of the electron source (left) and the detail of electron trajectories in the specimen chamber (right). In the specimen chamber, primary and back-scattered electrons are coloured red, while the secondary electrons are coloured blue.

Figure 4 show the gas velocity in the differentially pumped chamber. For better clarity, the displayed velocity is scaled to 100 m·s^{-1} (the maximum velocity was 460 m·s^{-1}). Due to the low pressure as well as the significant pressure difference between these chambers a supersonic choked flow can be observed [3]. As the gas moves through the aperture to the area with lower pressure, the velocity increases and the gas expands in all directions, which would not be visible for a subsonic flow in which only a narrow stream would be created. The supersonic flow forms at the end of the aperture and accelerates as it moves further away and is then decelerated to a subsonic flow by a shock wave.
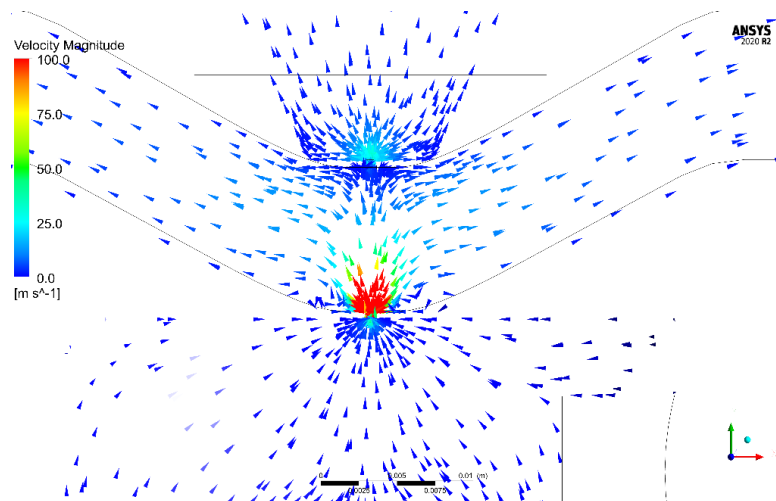

Figure 4: Rarefied gas flow in the differentially pumped chamber.

4   CONCLUSIONS

Numerical simulations of processes in electron microscopy can significantly help with designing and optimising new microscope parts as the experimental measurement might be often difficult or even impossible. The presented results show that the implemented custom model can describe the characteristic processes in an electron microscope. With this model, it might be possible to perform complex multiphysics simulations, which could increase the accuracy of the process description. From the gas flow simulation, it is possible to conclude that the flow regime significantly depends on the pressure and geometry conditions. While for some cases (high pressure), the continuum approximation with Navier-Stokes equations might be applicable, it is important to estimate the flow regime beforehand as there is a possibility that the Navier-Stokes equations will not be applicable. The presented electromagnetic model is able to describe the influence of electromagnetic field of trajectories of charged particles while considering relativistic effects. The theoretical velocity ($5.84 \cdot 10^7$ m·s^{-1}) was in an agreement with velocity obtained from numerical simulations ($5.83 \cdot 10^7$ m·s^{-1}). This model can be additionally extended with the description of stochastic particle-gas interactions to fully capture the whole process.

REFERENCES

[1]   N. Erdman, D. C. Bell, and R. Reichelt, „Scanning Electron Microscopy", in *Springer Handbook of Microscopy*., Cham: Springer International Publishing, 2019, pp. 231-305

[2]   G. D. Danilatos, "Theory of the Gaseous Detector Device in the Environmental Scanning Electron Microscope", *Advances in Electronics and Electron Physics*, vol. 78, no. 1, pp. 1-102, 1990.

[3]   Maxa, Hlavata, and Vyroubal, "Analysis of Impact of Conic Aperture in Differentially Pumped Chamber", *Advances in Military Technology*, vol. 14, no. 1, 2019.

[4]   F. R. Menter, "Two-equation eddy-viscosity turbulence models for engineering applications", *AIAA Journal*, vol. 32, no. 8, pp. 1598-1605, 1994.

[5]   F. G. Schmitt, "About Boussinesq's turbulent viscosity hypothesis: historical remarks and a direct evaluation of its validity", *Comptes Rendus Mécanique*, vol. 335, no. 9-10, pp. 617-627, 2007.

[6]   H. J. N. van Eck, W. R. Koppers, G. J. van Rooij, W. J. Goedheer, R. Engeln, D. C. Schram, N. J. L. Cardozo, and A. W. Kleyn, "Modeling and experiments on differential pumping in linear plasma generators operating at high gas flows", *Journal of Applied Physics*, vol. 105, no. 6, pp. 1-12, Mar. 2009.

[7]   Z. Guo, J. Qin, and C. Zheng, "Generalized second-order slip boundary condition for nonequilibrium gas flows", *Physical Review E*, vol. 89, no. 1, pp. 1-11, 2014.

[8]   P. Fiala, R. Kadlec, and J. Zukal, "Measuring Fluid Flow Velocities in the Context of Industry 4.0", in *2019 12th International Conference on Measurement*, 2019, pp. 295-298.

[9]   E. Munro, "Numerical simulation methods for electron and ion optics", *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 645, no. 1, pp. 266-272, 2011.

[10]   Y. -K. Kim, "Energy Distribution of Secondary Electrons", *Radiation Research*, vol. 64, no. 1, 1975.

COMPARISON OF MEASUREMENT METHODS FOR CHARACTERIZATION OF 3D PRINTED LOW-LOSS ARTIFICIAL DIELECTRIC SUBSTRATE BASED ON CROSS UNIT CELL

Petr Kaděra

Doctoral Degree Programme (3), FEEC BUT

E-mail: kadera@feec.vutbr.cz

Supervised by: Jaroslav Láčík

E-mail: lacik@feec.vutbr.cz

Abstract: This paper deals with the effective complex permittivity determination of low-loss 3D printed artificial dielectric substrate based on a cross unit cell. The two resonators working with TE_{011} and TM_{010} modes suitable for a uniaxial anisotropy determination are compared with the transmission / reflection waveguide method. To evaluate the methods' performance, the eigenmode analysis is carried out to provide precise reference values.

Keywords: Complex permittivity, artificial dielectric substrate, 3D printing, dielectric characterization, uniaxial anisotropy, resonators, waveguide

1 INTRODUCTION

A precise determination of an effective complex permittivity of low-loss 3D printed dielectric substrates plays an important role in the antenna and microwave circuit design [1]-[2]. For its determination, two measurement methods [3], [4] can be used. The first one is based on the exploitation two resonators [4] with known electric and magnetic field distributions allowing to calculate the material properties from change of the resonant frequency and cavity quality factor due to inserted sample of the material in the cavity. The second measurement method is based on the measurement of the transmission and reflection coefficient of a sample located in a waveguide and exploiting the non-iterative Nicolson-Ross-Weir (NRW) algorithm [5]. Both these measurement methods are able to measure anisotropy of a measured sample which is a typical feature of 3D printed structures.

In this paper, both the measurement methods are compared for the characterization of a 3D printed artificial low loss dielectric substrate based on a cross unit cell which is one of the basic elements exploited in the field of 3D printing [3]. For the evaluation of the suitability of the methods, the data of the eigenmode analysis of the cross unit cell is used [6]-[7]. Note that in this phase of our study, the data for the comparison was not obtained by the measurement of samples in the laboratory, but it was obtained by the modeling of the measurement methods in CST Studio Suite.

2 EFFECTIVE COMPLEX PERMITTIVITY DETERMINATION

2.1 EIGENMODE ANALYSIS METHOD

The method of eigenmode analysis is based on the modal analysis of an enclosed structure (defined by boundary conditions) when the electric and magnetic field distribution of each mode (representing a standing wave) are calculated. As a result, the eigenfrequencies and eigenvectors are determined and can be further used for calculation of the effective relative permittivity and tangent loss according to (1)-(3) [6]-[7].

$$\varepsilon_{\text{r,eff}} = \frac{\beta^2 - \left(\dfrac{c}{f}\right)^2}{4\pi^2}, \tag{1}$$

$$\tan\delta = \frac{P_{\text{D}}}{\pi \cdot f + 2 \cdot W}, \tag{2}$$

$$W = \frac{1}{2}\varepsilon_0\varepsilon_{\text{r,eff}}\int\left|\vec{E}\right|^2 \partial V + \frac{1}{2}\mu_0\mu_{\text{r}}\int\left|\vec{H}\right|^2 \partial V, \tag{3}$$

where $\varepsilon_{\text{r,eff}}$ is the effective relative permittivity, ε_0 is the permittivity of vacuum, μ_{r} is the relative permeability, μ_0 is the permeability of vacuum, β is the propagation constant, c is the speed of light, f represents the eigenfrequencies, $\tan\delta$ is the tangent loss, P_{D} denotes the dielectric power loss, W represents the total energy stored in the calculated structure, $\vec{E}, \vec{H}$ are eigenvectors corresponding to the electric and magnetic field, respectively and V is the volume of calculation domain.

The eigenmode analysis is applied in CST Studio Suite with the eigenmode JDM (Jacobi-Davidson Method) solver to determine the exact values of the effective relative permittivity and tangent loss of the investigated loss-less dielectric substrate. The dielectric substrate composition and boundary conditions settings are shown in Fig. 1. To evaluate this method, we assume the host permittivity ε_{h} of 1 (air) with the $\tan\delta_{\text{h}}$ of 0 and the inclusion permittivity ε_{i} of 2.5 with the $\tan\delta_{\text{i}}$ of $7.5 \cdot 10^{-3}$ at a frequency of 7 GHz, corresponding to the 3D printing filament (Prusament PLA Jet Black).
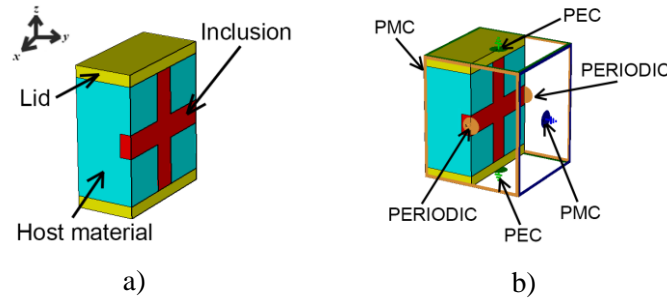

Fig. 1: The cross unit cell based dielectric substrate (a) and boundary conditions settings for the eigenmode analysis (b).

2.2 TE$_{011}$ AND TM$_{010}$ RESONATOR METHOD

Two resonant cavities with diameters of 56 mm and 32 mm and heights of 55 mm and 12 mm are exploited to find the effective complex permittivity of a cross unit cell dielectric substrate in the longitudinal (TE$_{011}$ mode) and the transversal (TM$_{010}$ mode) polarization direction, respectively [4]. The material of the cavity is gold with the electric conductivity σ of $5.7 \cdot 10^7$ S/m. The calculation is processed for a dielectric substrate of thickness 7.2 mm and the inclusion volume fraction (defined as the ratio of the volume of the inclusion material to the volume of the whole cross unit cell) between 0.2 and 1. The sample is located in the middle and at the bottom of TE$_{011}$ and TM$_{010}$ resonant cavities, respectively (Fig. 2). The feeding coupling loops are not depicted in this figure. The resonant frequency of the TE$_{011}$ mode in the empty cavity is 7.06637 GHz with the quality factor of 35263. Similarly, the resonant frequency of the TM$_{010}$ mode in the empty cavity is 7.17177 GHz with the quality factor of 8780.
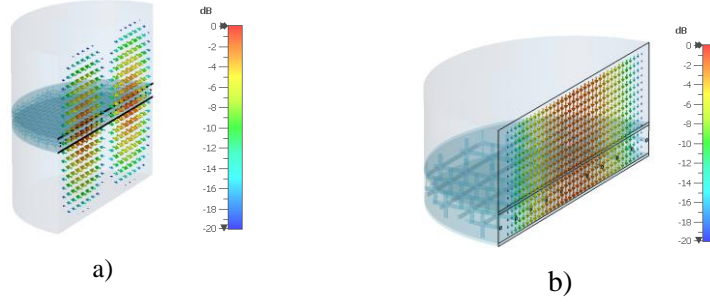
Fig. 2: The electric field distribution of TE_{011} resonator mode (a), electric field distribution of TM_{010} resonator mode (b). The cross unit cell based dielectric substrate with inclusion volume fraction 0.2 is located inside the resonators.

2.3 TRANSMISSION / REFLECTION NRW WAVEGUIDE METHOD

The WR137 waveguide with inner dimensions of 34.8488 mm x 15.7988 mm and propagation mode TE_{10} is assumed. The NRW algorithm with input transmission and reflection coefficients is utilized for the 3.6 mm thick sample. In the default sample arrangement (Fig. 3a), the extracted values of the effective complex permittivity correspond to the longitudinal direction. To obtain the values in the transversal direction, the elementary dielectric substrate unit cells have to be rotated by 90-degree with respect to the *x* axis. The waveguide model with inserted sample and electric field distribution at a frequency of 7 GHz is shown in Fig. 3.
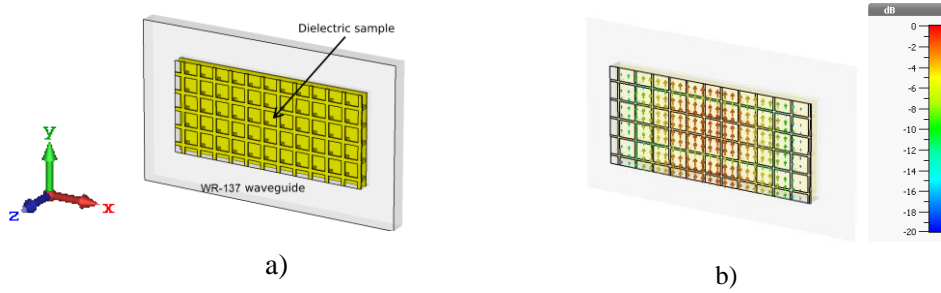

Fig. 3: The WR137 waveguide with cross unit cell based dielectric substrate with inclusion volume fraction 0.2 (cut in the middle of structure) (a), electric field distribution of TE_{10} mode at a frequency of 7 GHz (b).

3 RESULTS AND DISCUSSION

The achieved effective complex permittivity of the artificial dielectric substrate based on the cross unit cell in the longitudinal and transversal direction are presented in Fig. 4. The relative error (4-5) of the measurement methods (waveguide and resonators) related to the eigenmode analysis in Fig. 5 is calculated as follows:

$$\Delta_{\varepsilon_{r,eff}} = \left| \frac{\varepsilon_{r,eff\text{-}waveguide/resonator} - \varepsilon_{r,eff\text{-}eigenmode}}{\varepsilon_{r,eff\text{-}eigenmode}} \right| \cdot 100, \tag{4}$$

$$\Delta_{\tan\delta} = \left| \frac{\tan\delta_{waveguide/resonator} - \tan\delta_{eigenmode}}{\tan\delta_{eigenmode}} \right| \cdot 100, \tag{5}$$

where $\Delta_{\varepsilon r,eff}$ denotes the relative error of the effective relative permittivity and $\Delta_{\tan\delta}$ represent the relative error of tangent loss.

By comparison of the achieved results, the suitability of the measurement methods to accurately determine the effective complex permittivity of the artificial dielectric substrate based on the cross

unit cell can be confirmed. The relative error of the effective relative permittivity is under 2.2 % for both NRW waveguide and resonator methods, respectively, in both longitudinal and transversal directions for a PLA 3D printing material. The relative error of the tangent loss lies below 3.1 % and 7.1 % for NRW waveguide and resonator methods, respectively, in the longitudinal direction, and below 16 % and 5.2 % for NRW waveguide and resonator methods, respectively in the transversal direction.

Our study shows that both the NRW waveguide and resonator measurement methods are precise enough to determine the effective complex permittivity of the selected structure. However, in practice for low-loss materials, the resonator method could be more convenient because of its lower sensitivity to parasitic air gaps between sample and resonator's walls or sample defects compared to the transmission / reflection waveguide method. Such advantage of resonator method is pointed out in the recent study [8], where a multimode resonator cavity based on TM_{0n0} modes was employed with the tangent loss sensitivity in order of 10^{-4}-10^{-5}.

a)   b)

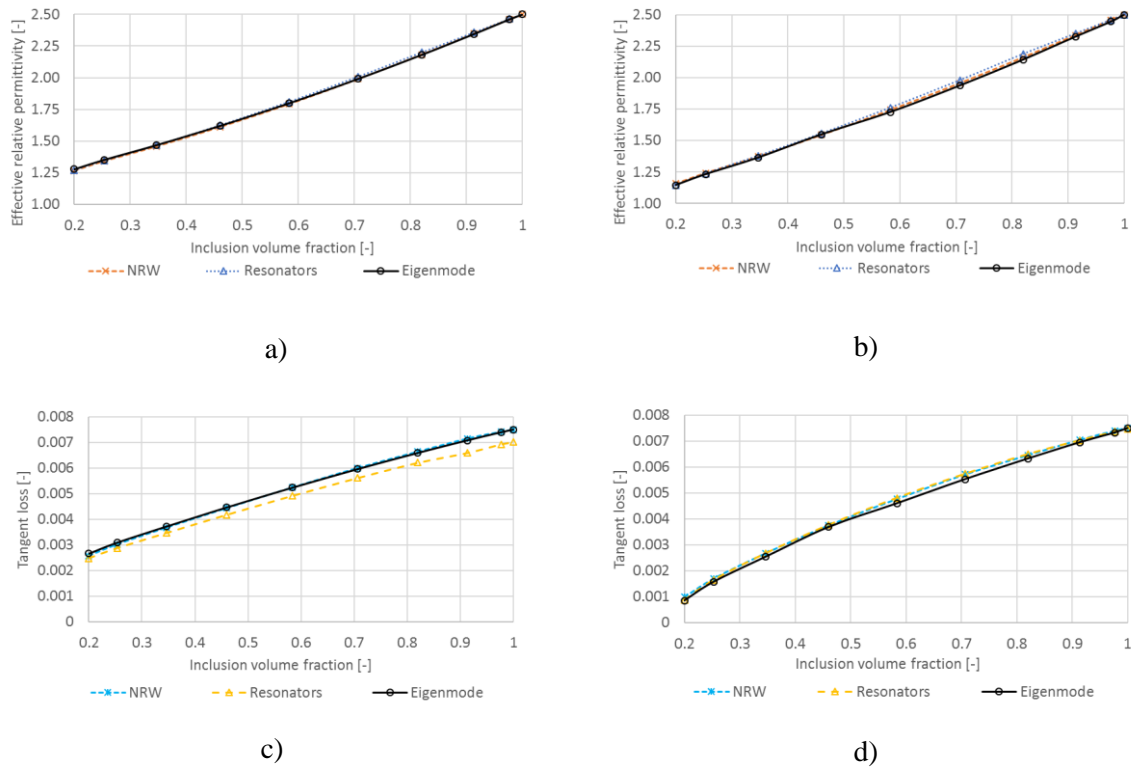c)   d)

Fig. 4:   The effective relative permittivity (a, b) and tangent loss (c, d) of the artificial dielectric substrate in longitudinal (a, c) and transversal (b, d) direction determined by the eigenmode analysis, NRW waveguide and resonator methods.

4   CONCLUSION

In this paper, we presented a comparison of the transmission / reflection waveguide method based on Nicolson-Ross-Weis algorithm and two resonator method based on TE_{011} and TM_{010} modes enabling to effectively describe effective complex permittivity of low-loss 3D printed artificial dielectric substrate based on the cross unit cell in the longitudinal and transversal directions. Both methods showed good ability in the anisotropy characterization. To evaluate the accuracy of the methods, the data of the eigenmode analysis of the cross unit cell was used. Future work will focus on resonators fabrication and methods' verification on various artificial dielectric substrate samples.
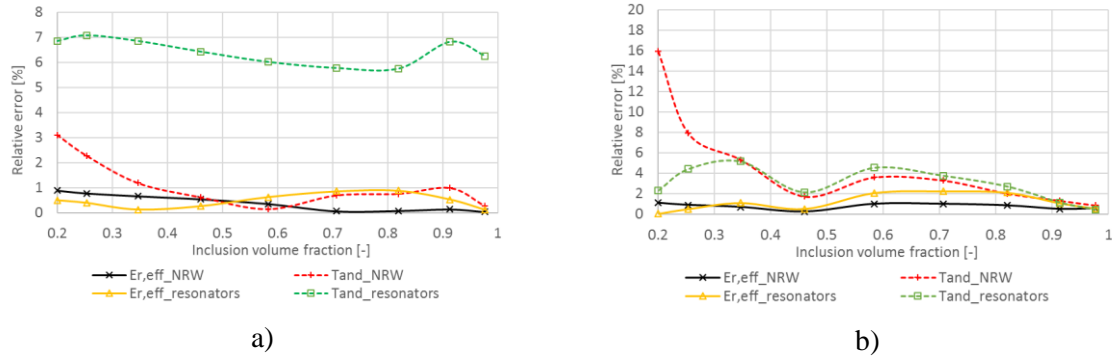
a)   b)

Fig. 5:   The relative error of effective relative permittivity and tangent loss of the artificial dielectric substrate in longitudinal (a) and transversal (b) direction.

REFERENCES

[1]   J. Castro, E. Rojas, A. Ross, T. Weller and J. Wang, "High-k and low-loss thermoplastic composites for Fused Deposition Modeling and their application to 3D-printed Ku-band antennas," in *2016 IEEE MTT-S International Microwave Symposium (IMS)*, San Francisco, CA, 2016, pp. 1-4.

[2]   E. Massoni *et al.*, "Characterization of 3D-printed dielectric substrates with different infill for microwave applications," in *2016 IEEE MTT-S International Microwave Workshop Series on Advanced Materials and Processes for RF and THz Applications (IMWS-AMP)*, Chengdu, 2016, pp. 1-4.

[3]   P. I. Dankov, "Characterization of Dielectric Properties, Resultant Isotropy and Anisotropy of 3D Printed Dielectrics," *in 2018 48th European Microwave Conference (EuMC)*, Madrid, 2018, pp. 823-826.

[4]   V. N. Levcheva, B. N. Hadjistamov and P. I. Dankov, "Two-resonator Method for Characterization of Dielectric Substrate Anisotropy," in *Bulgarian Journal of Physics*, vol. 35, pp. 33-52, 2008.

[5]   A. H. Boughriet, C. Legrand and A. Chapoton, "Noniterative stable transmission/reflection method for low-loss material complex permittivity determination," in *IEEE Transactions on Microwave Theory and Techniques*, vol. 45, no. 1, pp. 52-57, Jan. 1997.

[6]   M. Mrnka and Z. Raida, "An Effective Permittivity Tensor of Cylindrically Perforated Dielectrics," in IEEE Antennas and Wireless Propagation Letters, vol. 17, no. 1, pp. 66-69, Jan. 2018.

[7]   Li Mingyang, "Loss and Q Calculation Overview," [Online], [21-03-2021]. Available at: http://www.mweda.com/cst/cst2013/mergedprojects/cst_microwave_studio/special_overview/special_overview_loss_and_q_calculation_overview.htm.

[8]   P. Kopyt, J. Krupka and B. Salski, "Measurements of the complex anisotropic permittivity of microwave laminates," in *2020 23rd International Microwave and Radar Conference (MIKON)*, Warsaw, Poland, 2020, pp. 114-117.

TRANSIENT STABILITY ASSESSMENT FOR UNBALANCED FAULTS

Jan Koudelka

Doctoral Degree Programme (3), FEEC BUT

E-mail: xkoude20@stud.feec.vutbr.cz

Supervised by: Petr Toman

E-mail: toman@feec.vutbr.cz

Abstract: The paper describes the model applicable for transient stability assessment for unbalanced faults. The model itself is derived using symmetrical component representation of the grid. Sensitivity analysis of zero sequence line reactance, fault location and line length was done using the built model, the influence of unbalanced faults was discussed. The model is applicable for simplified stability assessment, but it could be used also for simulation purposes.

Keywords: transient stability, unbalanced operation, symmetrical components

1 INTRODUCTION

Electrical power grids are undergoing huge development these days, mainly due to changes in source base (increase of renewables and distributed generation), which enhances the requirements for reliable and resilient power system operation. This leads to the demand for transient stability assessment (TSA) enhancement. Currently used and researched methods are combining following aspects:

a. The simplest grid model used for TSA is called "Single Machine – Infinite Bus", abbreviated as SMIB (single line diagram is shown in Figure 1). Assessing stability using this model is straightforward and relatively easy [1]. Even large power systems are convertible to the SMIB model [2], but the precision of results need not be accurate.

b. It is difficult to define a measure for transient stability (how to quantify it). There is a general agreement on usage of Critical Clearing Time, abbrev. CCT [3]. CCT is the maximum allowable time of fault duration for which the assessed power system remains stable. Question remains, which fault is the most severe (leading to the lowest CCT). Usually, the three-phase short circuit occurring near to generator(s) in transmission system is used.

c. Commonly used methods for power system dynamics analyses are time domain simulations. Nevertheless, these simulations can be time-consuming and require computationally intensive, so the search for simplified methods is evident [4].

In scope of this paper is the TSA for power systems under unbalanced fault operation. Investigation of the influence of unbalanced faults on transient stability can be equally interesting, because this leads to the need of an advanced power grid modelling and can be practically used for investigation of real faults that occurred in the grid. In this paper, influence of shunt faults – Line-to-line (LL), Single line-to-ground (1LG) and Double line-to-ground (2LG) fault – and series faults – One line open (1LO) and Two lines open (2LO) – is compared using the SMIB model shown in Figure 1.
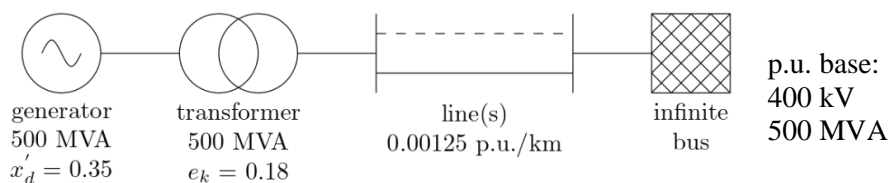

Figure 1:   Single line diagram of the SMIB model.

2   GRID MODEL FOR UNBALANCED FAULTS

The common method used for analysis of power systems under unbalanced operation is the method of symmetrical components. It is described e.g. in [5] and widely used for various applications. The main idea is to decompose the unsymmetrical system into three symmetrical components – positive (1), negative (2) and zero (0) sequence. The equivalent schemes of the SMIB model from Figure 1 for all three components are shown in Figure 2 for both shunt and series faults. These schemes respect the variable fault location on the line respected by coefficient α expressing the fault distance from the beginning of the line. The common simplification that grid elements are modelled only using the reactances [6] is kept.

| a) shunt faults | b) series faults |
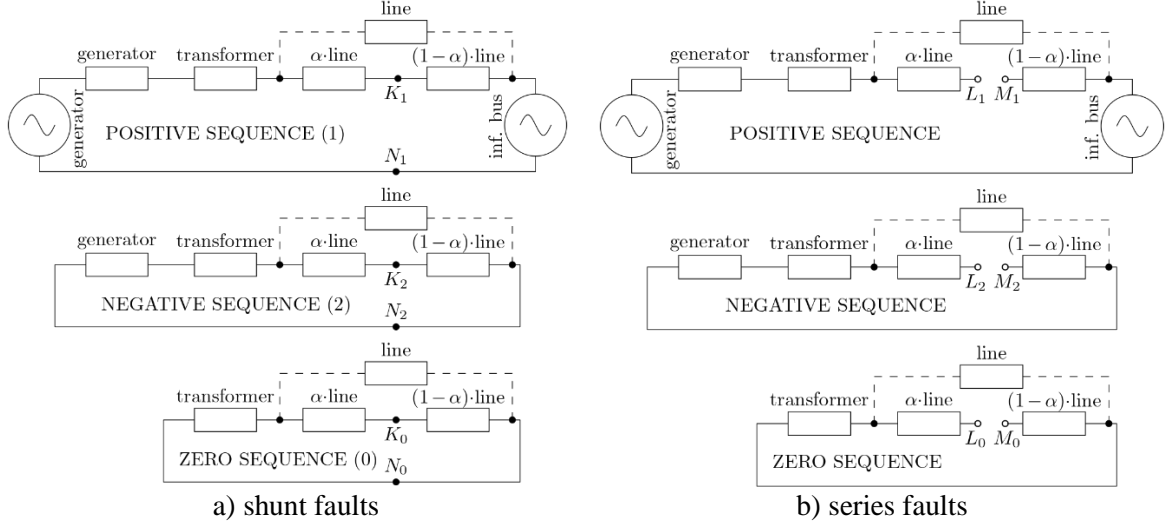
Figure 2:   Equivalent sequence representation of the grid (single-line diagram).

The key issue for further analysis is the interconnection of these schemes – interconnection of ports, which is derived for each fault separately e.g. in [5] and summarized in Table 1. To clarify the problem, it is possible to use the equivalent reactance X_Δ between the ports $K_1 - N_1$ for shunt faults and $L_1 - M_1$ for series faults. For stability assessment, it is important to calculate the transfer reactance between the machine and infinite bus X_{11} for 1 line operation, respectively X_{21} for 2 lines operation. Formulas derived from the system representation (Figure 3) are summarized in Table 1. Used indices are: g – generator, t – transformer, l – line.

Fault	Shunt faults			Series faults	
	LL	1LG	2LG	1LO	2LO
Connected ports	$K_1 - K_2,$ $N_1 - N_2$	$K_1 - N_0, N_2 - K_0,$ $K_2 - N_1$	$K_1 - K_2 - K_0,$ $N_1 - N_2 - N_0$	$L_1 - L_2 - L_0$ $M_1 - M_2 - M_0$	$L_1 - M_0, L_0 - M_2,$ $L_2 - M_1$
Neg. seq. reactance $X_{(2)}$ 1 line	$\dfrac{(1-\alpha)X_{l(2)} \cdot (X_{g(2)} + X_{t(2)} + \alpha X_{l(2)})}{X_{g(2)} + X_{t(2)} + X_{l(2)}}$			$X_{g(2)} + X_{t(2)} + X_{l(2)}$	
Neq. seq. reactance $X_{(2)}$ 2 lines	$\dfrac{\alpha(1-\alpha)}{2}X_{l(2)} + \dfrac{(1-\alpha)X_{l(2)} \cdot \left(X_{g(2)} + X_{t(2)} + \frac{\alpha}{2}X_{l(2)}\right)}{2X_{g(2)} + 2X_{t(2)} + X_{l(2)}}$			$X_{l(2)} + \dfrac{\left(X_{g(2)} + X_{t(2)}\right) \cdot X_{l(2)}}{X_{g(2)} + X_{t(2)} + X_{l(2)}}$	
Zero seq. reactance $X_{(0)}$ 1 line	$\dfrac{(1-\alpha)X_{l(0)} \cdot (X_{t(0)} + \alpha X_{l(0)})}{X_{t(0)} + X_{l(0)}}$			$X_{t(0)} + X_{l(0)}$	
Zero seq. reactance $X_{(0)}$ 2 lines	$\dfrac{\alpha(1-\alpha)}{2}X_{l(0)} + \dfrac{(1-\alpha)X_{l(0)} \cdot \left(X_{t(0)} + \frac{\alpha}{2}X_{l(0)}\right)}{2X_{t(0)} + X_{l(0)}}$			$X_{l(0)} + \dfrac{X_{t(0)}X_{l(0)}}{X_{t(0)} + X_{l(0)}}$	

Fault	Shunt faults			Series faults	
	LL	1LG	2LG	1LO	2LO
Equivalent reactance X_Δ	$X_{(2)}$	$X_{(2)} + X_{(0)}$	$\dfrac{X_{(2)}X_{(0)}}{X_{(2)} + X_{(0)}}$	$\dfrac{X_{(2)}X_{(0)}}{X_{(2)} + X_{(0)}}$	$X_{(2)} + X_{(0)}$
Transfer reactance 1 line X_{1l}	$X_{g(1)} + X_{t(1)} + X_{l(1)} + \dfrac{(1-\alpha)X_{l(1)}\left(X_{g(1)} + X_{t(1)} + \alpha X_{l(1)}\right)}{X_\Delta}$			$X_{g(1)} + X_{t(1)} + X_{l(1)} + X_\Delta$	
Transfer reactance 2 lines X_{2l}	$X_{g(1)} + X_{t(1)} + \dfrac{X_{l(1)}}{2} + \dfrac{(1-\alpha)X_{l(1)}\left(X_{g(1)} + X_{t(1)} + \frac{\alpha}{2}X_{l(1)}\right)}{\alpha(1-\alpha)X_{l(1)} + 2X_\Delta}$			$X_{g(1)} + X_{t(1)} + X_{l(1)} \cdot \dfrac{X_{l(1)} + X_\Delta}{2X_{l(1)} + X_\Delta}$	

Table 1: Transfer reactance

3 SENSITIVITY ANALYSIS

As was shown in Table 1, the value of transfer reactance is dependent on all three components impedances, but its calculation differs for each fault. Transfer reactance X_{1l} (X_{2l}) has significant influence on transient stability [6]. For TSA using simplified methods, the crucial value is the maximum transferable active power from the machine to the grid during the fault, which is, in a simplified way, inversely proportional to the transfer reactance.

To reveal the effect of selected factors on transfer reactance and therefore on transient stability, sensitivity analysis was done. The selected factors were zero sequence impedance, fault location and line length. Results accompanied by discussion are in the following subsections 3.1 – 3.3.

3.1 ZERO SEQUENCE IMPEDANCE

Negative sequence impedances are generally assumed to be the same as positive sequence impedances. The issue is the value of zero sequence impedance. Regarding power lines, zero sequence impedance is mainly influenced by the usage of ground wires. It can be calculated or measured, and its common range is 1÷5 times higher than the positive sequence reactance. This range was tested; results are shown in Figure 4. In studied case, line length was 100 km and fault location α = 0 %.

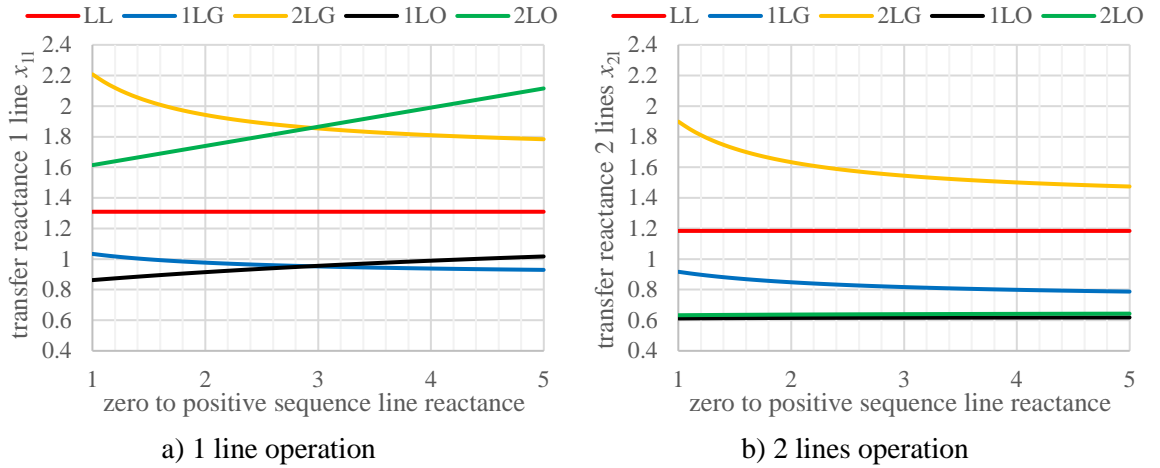

a) 1 line operation

b) 2 lines operation

Figure 3: Effect of zero sequence line reactance on transfer reactance.

From Figure 4, it can be observed that the most significant effect of increasing zero sequence line reactance is in case of 2LO (1 line operation; X_{1l} is increasing) and 2LG (X_{1l} and X_{2l} are decreasing). For 1LG and 1LO, the effect is small, for LL, there is no influence.

3.2 FAULT LOCATION

For analysis, fault location was tested in range 0÷99 % of the faulted line. Results for line length 100 km and zero sequence line reactance ratio 1 are shown in Figure 5. Results confirms the assumption the closer a shunt fault is located to the machine, the worse effect on stability it has. The significant influence is mainly in case of 2 lines operation. Referring to formulas derived in Table 1, there is no influence of fault location in case of series faults.
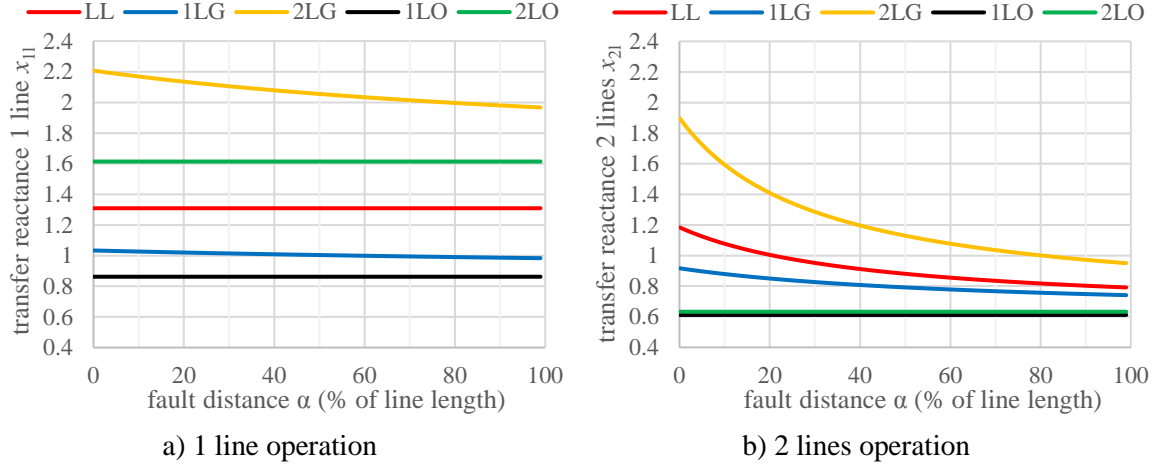

Figure 4: Effect of fault location on transfer reactance.

3.3 LINE LENGTH

The influence of line length on stability is not so straightforward comparing to aforementioned aspects. This is harder to explain, but in cases 3.1 and 3.2, the steady state before fault was the same, so it was reasonable to study only the transfer reactance. The method in this case was kept, but the interpretation must be different: the question in this case is, how much each fault will increase the transfer reactance, i.e. how much it will worsen the stability for the specific line length.

From Figure 6 can be seen that the worst case is the 2LG fault, especially for long line operation. It is interesting that the 2LO fault is the second most severe in 1 line operation, but the second lowest in 2 lines operation.

Grouping and ordering the faults regarding the transfer reactance, the order is 1LG – LL – 2LG for shunt faults and 1LO – 2LO for series faults.
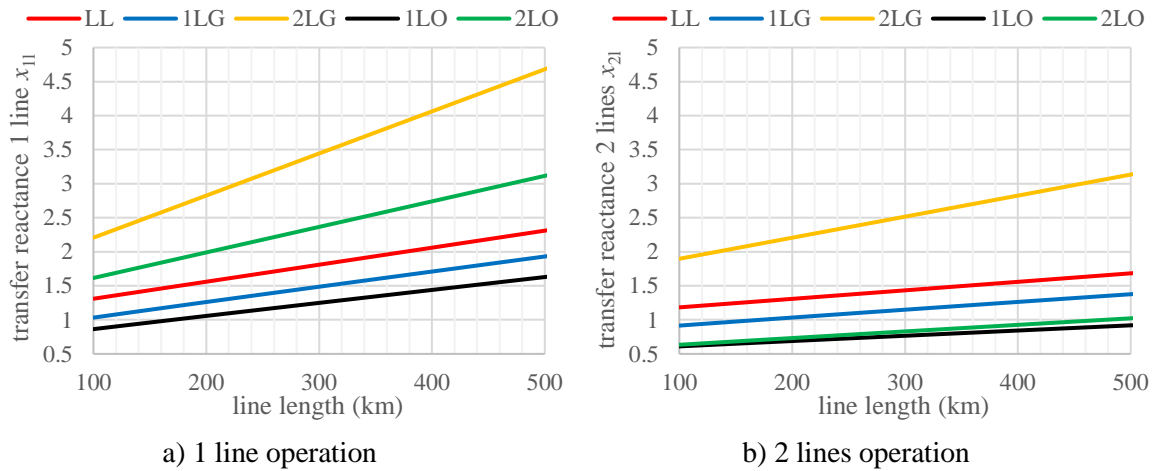

Figure 5: Effect of line length on transfer reactance.

4 CONCLUSION

The main scope of this paper was the power system representation for TSA for unbalanced faults. The simplified model SMIB was extended using symmetrical components method and the formulas for transfer reactance used for TSA were derived. The results are useful for simplified TSA methods as they allow to easily calculate the transfer reactance for unbalanced faults and do the simplified assessment using the SMIB representation of the grid. The symmetrical component representation could be also implemented to simulation tools (especially the user created ones e.g. in Matlab) allowing them to extend the simulation possibilities.

Sensitivity analysis of zero sequence line reactance, fault location and line length was carried out. Results show that the most severe unbalanced fault is the 2LG fault, which is also influenced by these factors the most. However, zero sequence line reactance has significant impact on 2LO fault. To get the accurate results, it is recommended to clearly assess the zero sequence reactance of the line. Findings also show that faults located closer to the synchronous generator have worse impact on transient stability.

ACKNOWLEDGEMENT

REFERENCES

[1]   SUN, Y., J. MA, J. KURTHS a M. ZHAN. Equal-area criterion in power systems revisited. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* [online]. 2018, **474**(2210) [cit. 2021-03-12]. ISSN 1364-5021. doi:10.1098/rspa.2017.0733.

[2]   DJALAL, M. R. et al. An Approach Transient Stability Analysis Using Equivalent Impedance Modified in 150 kV South of Sulawesi System. *Journal of Electrical and Electronic Engineering-UMSIDA* [online]. 2017, **1**(1), 1-7 [cit. 2021-03-12]. ISSN 2460-9250. doi:10.21070/jeee-u.v1i1.758.

[3]   SULISTIAWATI, I. B. et al. Critical Clearing Time prediction within various loads for transient stability assessment by means of the Extreme Learning Machine method. *International Journal of Electrical Power & Energy Systems* [online]. 2016, **77**, 345-352 [cit. 2021-03-12]. ISSN 01420615. Dostupné z: doi:10.1016/j.ijepes.2015.11.034.

[4]   K. MASLO et al. Methods for Dynamic Stability Assessment. In: *2019 20th International Scientific Conference on Electric Power Engineering (EPE)* [online]. IEEE, 2019, 2019, s. 1-5 [cit. 2021-03-12]. ISBN 978-1-7281-1334-0. doi:10.1109/EPE.2019.8778182.

[5]   ANDERSON, P. M.: *Analysis of faulted power systems*. New York: IEEE Press, 1995. ISBN 0-7803-1145-0.

[6]   KUNDUR, P., N. J. BALU a M. G. LAUBY. *Power system stability and control*. New York: McGraw-Hill, c1994. ISBN 0-07-035958-X.

[7]   Orság, F.: Vision für die Zukunft. Biometrie, Kreutztal, DE, b-Quadrat, 2004, s. 131-145, ISBN 3-933609-02-X

[8]   Drahanský, M., Orság, F.: Biometric Security Systems: Robustness of the Fingerprint and Speech Technologies. In: BT 2004 - International Workshop on Biometric Technologies, Calgary, CA, 2004, s. 99-103

VOLTAGE CONTROL IN MEDUIM VOLTAGE GRIDS EMPLOYING DISTRIBUTED POWER SOURCES

Jan Klusacek

Doctoral Degree Programme (1st year), FEEC BUT

E-mail: klusacekj@ feec.vutbr.cz

Supervised by: Jiri Drapela

E-mail: drapela@feec.vutbr.cz

Abstract: The main objective of the paper is to compare selected modes of voltage control in medium voltage (MV) distribution system (DS) by means of reactive power injection provided by distributed power generating plants (DPGPs). At first, rationale for voltage rise compensation and voltage control in power grids is made in order to select three reactive power control modes of DPGP. The control modes are described in terms of ability to autonomous operation and suitability of their use in MV DS is examined using parametric simulation.

Keywords: Distributed Power Generating Plants, Autonomous Control, Reactive Power Control

1 INTRODUCTION

In recent years, emphasis on integration of DPGP into low voltage (LV) and MV DSs is recognized and distributed generation is perceived as an instrument of transition to carbon neutral energy generation. On the other hand, in order to incorporate the DPGPs without destabilizing the DS operation [1] and in order to ensure the DPGPs operation without interrupting for all players [2], robust concept of integration with transparent requirements on the DPGPs in terms of compulsory functionalities must be established [3]. Selection of suitable strategy of voltage coordination in DSs, with regards to development level of technology being deployed to DPGPs [4], is crucial part of it. This paper is focused on comparison of selected voltage control modes implemented in DPGPs operating in MV DS with typical parameters. In the theoretical part of the paper, requirements on DPGPs control are identified and three approaches of voltage control in MV grid using DPGPs are proposed. Consequently, parametric model of MV grid with multiple DPGPs and realistic parameters in PSCAD is presented and individual control modes are compared by simulation. Finally, simulation results are discussed and topics for further development are identified.

2 PROBLEM DEFINITION

The MV network voltage control, unlike from high voltage (HV) networks monitored in a pilot node predominantly [5], employs an on-load tap changer (OLTC) with level voltage regulator (LVR) on HV/MV transformer, which controls the voltage of main busbar in substation discretely with time and voltage deviation hysteresis. Thus, voltage at the beginning of all branches is set centrally and actual voltage profile of individual branches depends on actual load and dimensioning of the components. However, serious issue with voltage coordination might occur [6], especially if power flow directions and consequently the voltage changes over two branches are in opposite. In order to mitigate voltage rise caused by generated active power, reactive power demand might be enforced and thus voltage rise might be compensated on a MV line with typical resistance/reactance ratio (R/X=1). The principle of the voltage mitigation is depicted in **Figure 1**, where effect of increasing reactive current from I_r^1 to I_r^2 is compared (current is inductive in consumption convention). Furthermore, using the geometry, calculation of voltage change scalar value might be simplified to formula

$\Delta V \cong R|\bar{I}|\cos\varphi + X|\bar{I}|\sin\varphi$ (1) for small angles resulting in $(0.9 < |\pm \cos\varphi| < 1)$, which shows the real-part-current-resistance and reactive-part-current-reactance linkages as well.
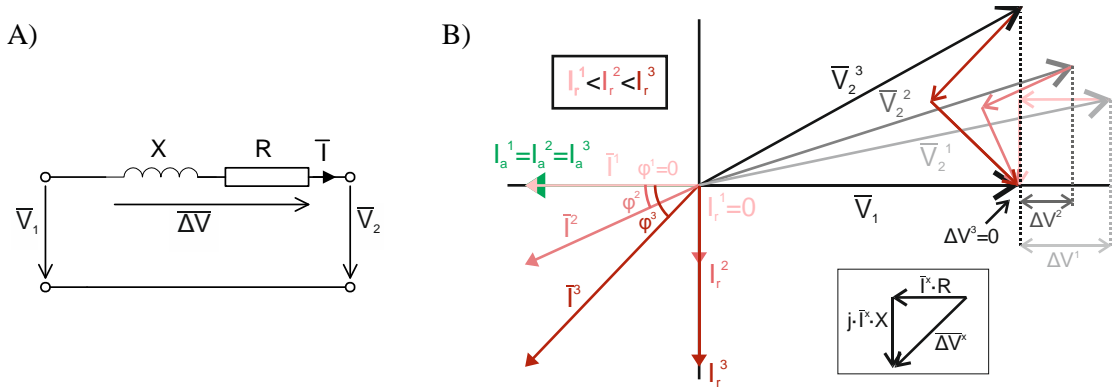
A)

B)

Figure 1:   1A: Single phase power line model, 1B: principle of voltage rise mitigation.

$$\Delta V \cong R|\bar{I}|\cos\varphi + X|\bar{I}|\sin\varphi \qquad (1)$$

Described principle of voltage magnitude compensation is implemented in [7], where real part of current is presented by means of active power *P*, reactive part by means of reactive power *Q* and limits are recommended by PQ diagram (**Figure 2A**). Furthermore, the document [7] proposes several strategies how to operate in the area, e.g. required $\cos\varphi$, implementation of Q(U) function, or required *Q*. Nevertheless, actual strategy for operation set point of power sources should be defined by DS operators individually, regarding local conditions/circumstances.

3   VOLTAGE CONTROL STRATEGIES

In this chapter, link between definition of setpoint in PQ diagram in order to mitigate the voltage change and grid voltage control is presented on control modes, which are described by signal schemes in **Figure 2**, where DPGP is considered as single power generating unit (PGU) with unique point of connection (PoC) to the DS. All the control modes are intended to control the voltage by reactive power (directly or indirectly) and the active power would be limited only if the local voltage reaches protection limits.
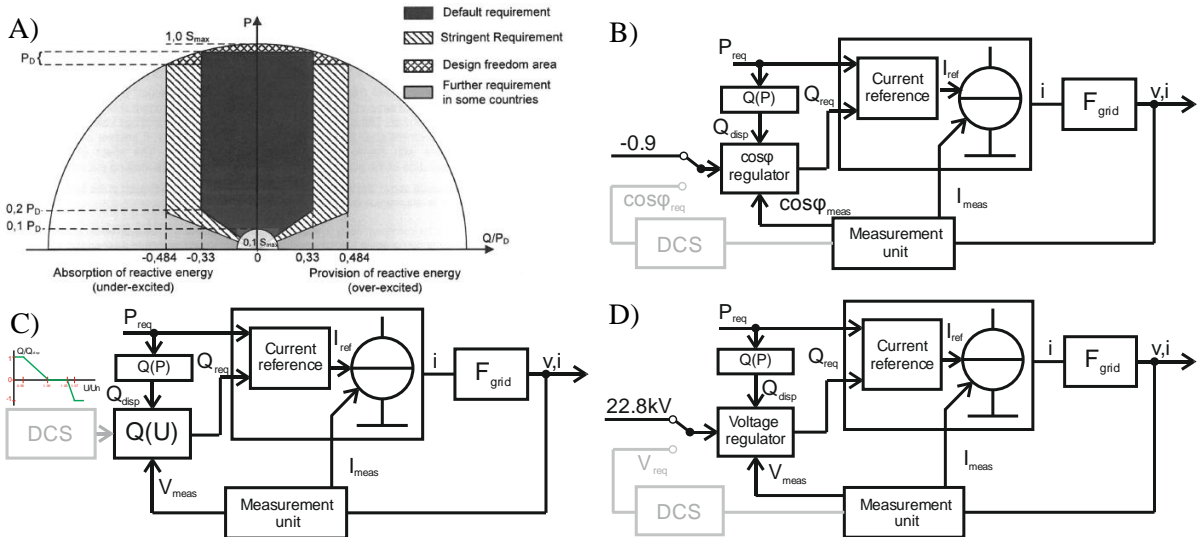

Figure 2:   Considered control modes and corresponding control and regulation schemes: A) PQ diagram according to [7], B) MOD1, C) MOD2; and D) MOD3.

The PGU in **Figure 2B** (MOD1) maintains requested cosφ using regulator and it is set either as a constant or it may be set based on grid steady state analysis in dispatching control system (DCS) provided via communication. Thus, information about local voltage and voltage of adjacent nodes may be incorporated. On contrary, PGU in **Figure 2C** (MOD2) implements Q(U) function in order to reflect voltage at PoC directly and communication with DCS is used for Q(U) parametrization. Both modes are in accordance with [7]. Control mode depicted in **Figure 2D** (MOD3) is inspired by control scheme of PGUs connected to HV level. It contains voltage regulator in order to maintain the voltage at PoC on requested value, which is set either as constant or it may be received from DCS on similar basis as in the case of MOD1. It is obvious, that emphasis on communication path is increased in the case of MOD1 and MOD3. Furthermore, penetration of measurement units in the grid should be relatively high in order to perform the steady state analysis correctly and keeping the voltage in desired limits in every node.

4 SIMULATION MODEL

Behavior of the PGU implementing proposed control modes has been evaluated in a test model of MV grid (**Figure 3**) in PSCAD software, where parametric simulation with fully symmetrical components was carried out. The model consists of a HV feeder, HV/MV transformer equipped with OLTC controlled by LVR, and of MV grid consisting of one radial branch in view. Rest of the branches are represented by means of an equivalent load at the substation busbar. Overall charging capacitive current of those branches is modelled using a single capacitor bank with nominal current of 200 A (~3 Mvar) and located at the substation busbar. The capacitive current rating is typical for rural MV networks with overhead lines. The single radial branch is modelled in details in order to examine its voltage profile when three PGUs: PV1, PV2 and PV3, with the same control mode (S1/S2/S3/S4 in **Table 1**) at the time, are connected in nodes N2, N3 and N4, respectively. The PGUs nominal power (2 MW) has been tuned together with impedance of the line sections between nodes (R=5.25 Ω, X_L=5.25 Ω) in order to induce typical performance of the tested control modes.
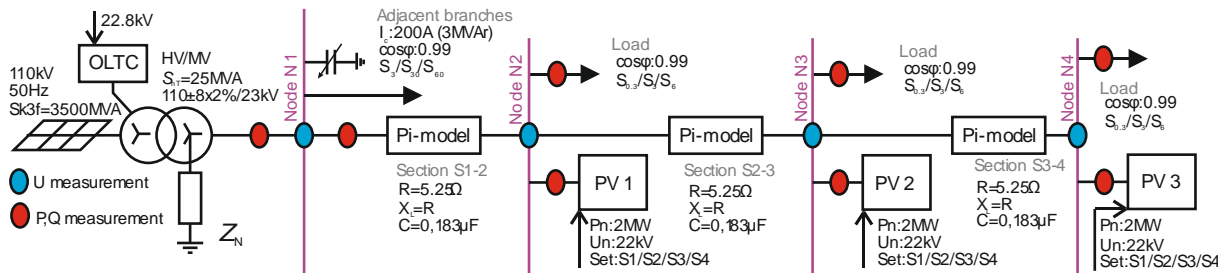

Figure 3: Scheme of test MV grid model with radial branch depicted in details.

Loads in nodes N2, N3 and N4, together with load in N1, simulate DS consumption and it varied in three steps emulating: 1) LOAD1 – moderate consumption, 2) LOAD2 – medium consumption and 3) LOAD3 – high level of consumption. During LOAD1, load power in N2, N3, N4 (branch loads) was set to 0.3 % of the transformer nominal power (S_{nT}) and load in N1 (grid load) was set to 3 % of S_{nT}. During LOAD2 branch loads and grid load power is 3 % and 30 % of S_{nT} respectively and during LOAD3 the branch loads and grid load power is 6 % and 60 % of S_{nT} respectively. Power factor of all loads was set to 0.99.

Setting	Control mode	Characteristic parameters				
S1	MOD1	PGU	PV 1	PV 2	PV 3	(- refers to in-ductive charac-ter in consump-tion convention)
		cosφ (-)	1.00	1.00	1.00	
S2	MOD1	PGU	PV 1	PV 2	PV 3	
		cosφ (-)	-0.9	-0.9	-0.9	
S3	MOD2					
S4	MOD3	PGU	PV 1	PV 2	PV 3	
		U_z (kV)	22.8	22.8	22.8	

	PV 1	PV 2	PV 3
$[X_1,Y_1]$	[0.93,1]	[0.93,1]	[0.93,1]
$[X_2,Y_2]$	[0.99,0]	[1.00,0]	[1.00,0]
$[X_3,Y_3]$	[1.04,0]	[1.05,0]	[1.05,0]
$[X_4,Y_4]$	[1.06,-1]	[1.07,-1]	[1.07,-1]

Table 1: Parameters of control modes of generators.

The varying parameters of the simulation were: 1) available generators power (in % of P_n, from 0 % to 100 % with step of 10 %), 2) control mode of generators (S1/S2/S3/S4) and 3) load level (LOAD1/LOAD2/LOAD3). Voltage, active power and reactive power were measured when assuming symmetrical loads/generators and they were recorded after reaching quasi steady-state.

5 SIMULATION RESULTS

The simulation results are presented in **Figure 4** to **Figure 6**. In the **Figure 4**, voltage profile of the branch can be observed when different generators settings are applied. The setting S0 (red set of parameters) indicates the state when no power is generated, the grid purely consumes and the state is considered as a reference characterizing the grid performance. According to the voltage profile, the MV branch is rather overloaded and LOAD3 forces the LVR to change the tap in all control modes.
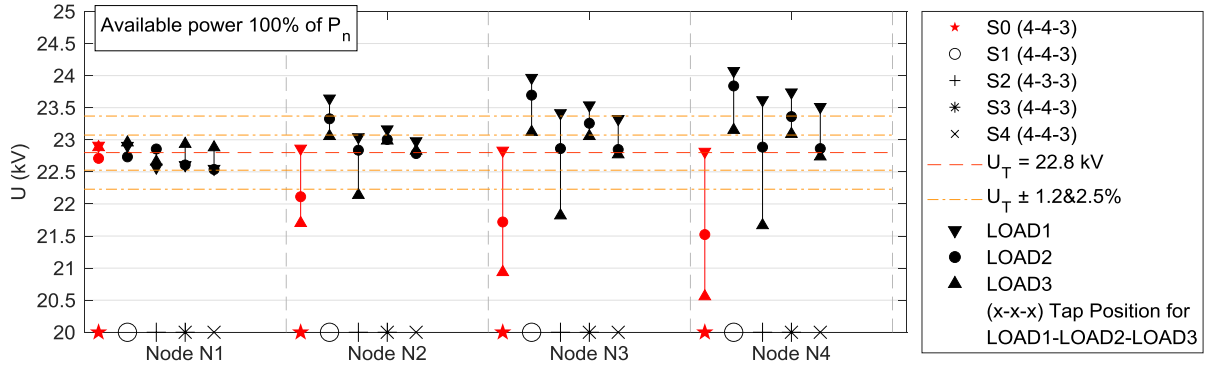

Figure 4: Voltage in nodes N1, N2, N3 and N4 under various generator settings and grid load, available power 100 % of P_n.

Based on **Figure 4**, S1 and S2 shows that voltage fluctuation, when different consumption levels are set, is higher than compared to S3 and S4. Simultaneously, **Figure 6** shows that if suitable requested power factor is set, the generator PV2 and PV3 active power is not limited due to overvoltage (see difference between S1 and S2 in the **Figure 6**). Therefore, it can be concluded that although the voltage rise due to generated active power might be eliminated by reactive power in MOD1, the voltage drop caused by change of load is not compensated as in case of S3 and S4.
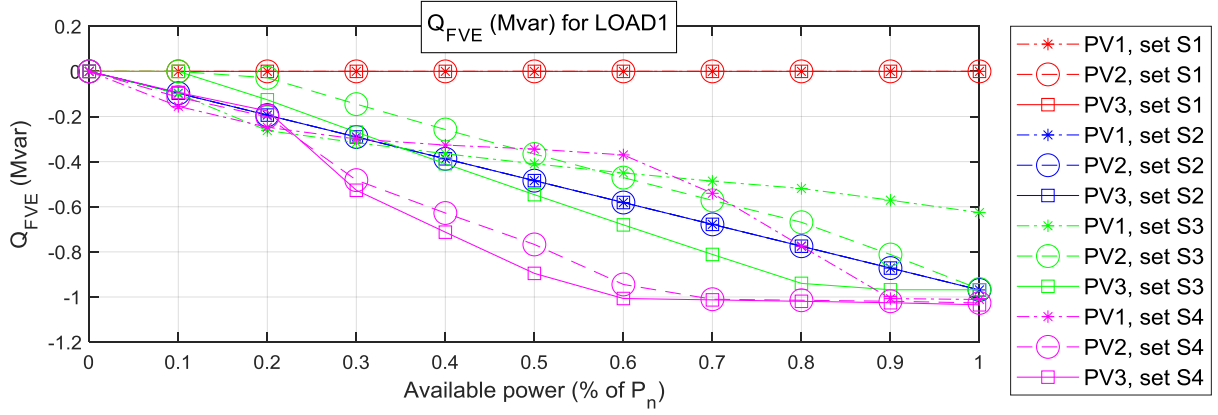
Figure 5: Compensating reactive power among available power of the generators at LOAD1.

Figure 5 shows development of compensating reactive power of each generator over increasing available power for LOAD1 (the least favorable case in terms of voltage rise compensation). It is obvious that total injection of reactive power is the highest for S4 (MOD3) and it correlates with voltage profile in **Figure 4** for LOAD1. Furthermore, contributions from each generator cannot be coordinated deterministically in MOD3 and the reactive power is limited only by PQ area. On contrary, determinism of reactive power increases from MOD2 (S3) to MOD1 (S1/S2), so deployment of the generators to potential superior Q coordination algorithm is technically enabled.
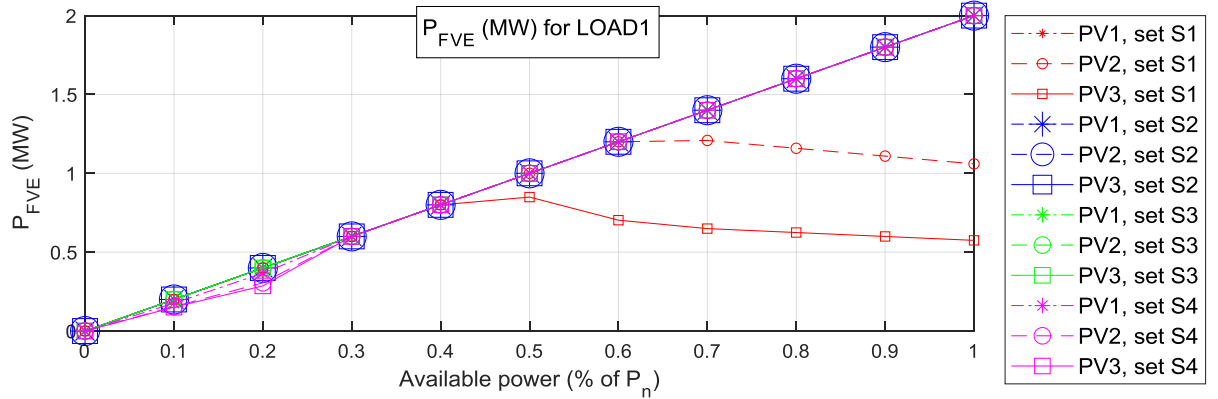
Figure 6: Active power among available power of the generators at LOAD1.

6 CONCLUSION

Three selected approaches of generators voltage control are proposed in the paper and their implementation is examined in model of MV grid using parametric simulation. The performance of the approaches is evaluated from several aspects: ability to compensate the voltage rise, ability to adapt on current local conditions (varying consumption), ability to coordinate reactive power directly and ability to adequate and stable autonomous operation when communication with DCS is interrupted. It may be concluded, that performance of control with Q(U) function is compromise among all the listed aspects and it is worth for future development in terms of, e.g. adapting the Q(U) curve to position in the grid and desired voltage profile or parametrization in order to achieve desired coordination of reactive power injection by each generator. Lastly, parametric model is introduced as a tool for analysis of distributed generators incorporation and for evaluation their control modes and settings.

ACKNOWLEDGMENT

REFERENCES

[1]   LV, Yazhou et al. Emergency Control Strategy for Transient Stability in Large Power Grid Considering Distributed Generation IEEE, 2020, doi:10.1109/ICPSAsia48933.2020.9208559

[2]   BAYINDI, Ramazan et al. Effects of renewable energy sources on the power system. 2016 IEEE International Power Electronics and Motion Control Conference (PEMC. IEEE, 2016, doi:10.1109/EPEPEMC.2016.7752029

[3]   WU, Yuan-Kang et al. Standards and Guidelines for Grid-Connected Photovoltaic Generation Systems: A Review and Comparison. IEEE Transactions on Industry Applications 2017, doi:10.1109/TIA.2017.2680409

[4]   HANNAN, M.A et al. Optimized controller for renewable energy sources integration into microgrid: Functions, constraints and suggestions. Journal of Cleaner Production 2020, doi:10.1016/j.jclepro.2020.120419

[5]   ESPEJO, Rafael et al. Analysis of transmission-power-grid topology and scalability, the European case study. Physica A: Statistical Mechanics and its Applications. 2018. doi:10.1016/j.physa.2018.06.019

[6]   NIGAR, Yasmin, et al. Visualising the effect of DG on voltage profile in medium voltage distribution networks. 2015 IEEE PES Asia-Pacific Power and Energy Engineering Conference (APPEEC). IEEE, 2015,doi:10.1109/APPEEC.2015.7381075

[7]   EN50549-2. Requirements for generating plants to be connected in parallel with distribution networks - Part 2: Connection to a MV distribution network - Generating plants up to and including Type B

OPTIMIZATION OF MESH NETWORK CONFIGURATION FOR REDUCTION OF SHORT CIRCUIT CURRENTS

Vit Krcal

Doctoral Degree Programme (2), FEEC BUT

E-mail: vit.krcal@vutbr.cz

Supervised by: David Topolanek

E-mail: topolanek@vutbr.cz

Abstract: This paper deals with LV mesh network reconfiguration in order to minimize short circuit currents. The reconfiguration rests in switching off selected network elements leading to increase of short circuit impedances in LV nodes. For optimal configuration, an algorithm of sequential disconnecting of elements is proposed, based on load flow and short circuit calculations. The reconfiguration procedure is applied to a model of a mesh MV/LV network. Reduction of short circuit currents in this network is presented, capturing impact of each step of the sequence. Based on the presented results, deploying other measures and further development of the proposed algorithm are discussed.

Keywords: Mesh network, optimization, short-circuit, network reconfiguration

1   INTRODUCTION

In urban areas, the MV/LV underground cable distribution networks are often structured in mesh topology [1]. The advantages of this kind of operation are mainly in high distribution reliability, low outage probability, low energy losses and stable voltage in the network [2]. These benefits are compensated by several shortcomings, such as high capital expenditures, poor clarity of network status and high level of short circuit (SC) currents. The problem of high short circuit currents is tackled in this paper. The objective is to design an algorithm to find possible network configuration with sufficiently low SC currents and, at the same time, fulfil requirements of distribution system operation. In this paper, a model of a dense-mesh distribution network is used. The network model data are fictious but the topology is based real network data provided by distribution system operator (DSO). Nevertheless, the resultant algorithm may be further developed and can be used for real mesh networks as well, providing valuable information for the DSO.

2   MESH NETWORK MODEL

The network model used for this study represents an urban underground MV/LV network arranged in mesh topology. This network is very dense due to a large number of interconnecting lines between LV nodes. The network consists of more than 1,000 nodes, more than 1,000 lines and c. 100 MV/LV transformers. There are usually two transformers in each transformer station. In the case of two transformers in parallel operation in a station, there is a bus coupler connecting LV buses of both transformers. Another bus couplers are situated in switchgears in some LV nodes, enabling sectionalizing the network. The current state of operation is considered with all the bus couplers (c. 50) closed (switched on). This configuration results in very high values of SC currents, which is demonstrated in section 4. Meshed topology of the network using graph visualization is depicted in Figure 3.

3   SHORT CIRCUIT CURRENT CALCULATION

In the case of the LV network, which consists exclusively of three phase cables, single line faults mostly escalate to three phase faults. Therefore, only three-phase balanced faults are considered

for purposes of this study. The evaluation of three phase SC currents is based on impedance matrix principle, which is described e.g., in [3]. Using this method, the bus impedance matrix is created as an inverse of the bus admittance matrix, which is straightforwardly constructed for any network configuration according to a well-known principle [3]. The diagonal bus impedances represent Thevenin impedances [4] and can be used in calculation of SC currents using equivalent voltage source at the SC location. According to IEC standard 60909-0 [5], the initial symmetrical short-circuit current can then be calculated as

$$I_k'' = \frac{c \cdot U_n}{Z_k \cdot \sqrt{3}},$$ (1)

where c is the voltage factor, U_n is nominal line to line system voltage and Z_k is the Thevenin impedance, derived from bus the impedance matrix $[\bar{Z}]$. This method can be modified for calculation of asymmetrical faults, using separate bus impedance matrices for each positive, negative and zero sequence [4]. In this paper, only initial symmetrical short-circuit currents (SC currents) are evaluated using formula (1).

4 CONFIGURATION OPTIMIZATION

Based on calculations carried out according to (1), it was revealed that, in considered current network configuration, the calculated values of SC currents are significantly higher than usual SC current rating (SCCR) of LV switchgears. This fact led to proposing of optimizing network configuration aiming to reduce the SC currents below the level of existing SCCRs. In reality, the procedure of network reconfiguration has to comply with actual possibilities of the DSO and should consider both technical and economic aspects. In this paper, two of simply executable operations are considered. The easiest way to reconfigure the network is to switch off the bus couplers placed in the LV network to weaken the network and hence increase the SC impedances. The other feasible, and therefore considered measure is to disconnect some MV/LV transformers, which have significant impact on SC impedances. Another measures of network reconfiguration leading to lowering SC currents, such as replacing the transformers or sectionizing the network into two or more individual networks, are not considered in this study. These possibilities could be taken into account in case that abovementioned considered measures do not lead to satisfying results.

4.1 CONDITIONS FOR NETWORK OPERATION

Each network steady state in the process of reconfiguration has to be checked for limiting operating conditions, meaning that selected parameters have to be in compliance with distribution system operation requirements and power quality standards. Several conditions derived from load flow calculation can be applied, from which some were used in simulation in this paper with following definition:

- *Voltage condition:* Voltages in all nodes have to be within the specified limits (±10% of V_n).
- *Loading condition:* Loading of branch elements (lines and transformers) have to be lower than specified limit (80%).
- *Reverse active power condition:* Active power in all transformers has to flow from MV to LV side.
- *Supply interruption condition:* All customers have to be connected to the network at all times.

Other criteria can be applied, such as contingency analysis (N-1 criterion), but were not implemented in this study. It is important to define a termination condition, which reflects the optimization objective:

termination condition (SC condition): SC currents in all nodes have to be lower than short circuit current rating of LV switchboards, $I_k'' \leq SCCR$.

4.2 PROPOSED ALGORITHM

According to principles described in the beginning of section 4, a simple algorithm for network reconfiguration has been designed. This algorithm is based on sequential disconnecting of elements in the network. The order, in which the elements are being disconnected, is based on the impact of those elements on the resultant SC currents. Due to the paper page restriction, this principle will be described in future publications. As mentioned earlier, the elements in this study represent LV bus couplers and MV/LV transformers.

This sequential approach, when the elements are disconnected one by one, reflects the real operations procedure and has following advantages:

- Every stage of the procedure is calculated, and operating conditions are checked for specified limits. This ensures non-problematic network operation during the procedure of elements disconnection (if performed in the same order). Impact of each of successive steps can be evaluated.
- The calculation time is much lower than in the case of calculating all possible network configurations. Considering 150 switchable elements, using a brute-force approach would result in $2^{150} \approx 10^{45}$ permutations. Contrast to that, in the case of sequential disconnecting, the maximum number of calculated combinations, given by sum of an arithmetic series, is $151 \cdot 150/2 \approx 10^4$.

Algorithm procedure is simplistically depicted in Figure 1. The initial network configuration is considered with all bus couplers and transformers switched on. The loads are derived from usual rated currents of fuses in LV network. The algorithm can be further enhanced and supplied with more inputs and conditions. Based on conditions mentioned in 4.1 the algorithm is terminated if:
- any of the limiting operating conditions is not fulfilled (optimization cannot be reached), or
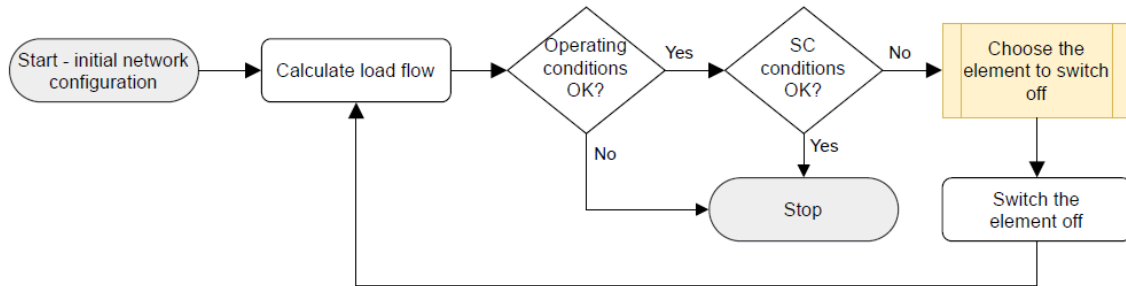- the termination condition and operating conditions are fulfilled (optimization reached).

Figure 1: Flowchart of proposed algorithm

4.3 RESULTS

The proposed algorithm was applied to mesh network model described in chapter 2. The initial network configuration respects described meshed configuration, meaning that all transformers are in operation and all the bus couplers are switched on. Limiting operation conditions specified in chapter 4.1 were applied. As the LV switchboard short-circuit current rating, one universal value of 40 kA was used for all LV nodes.

For this specific setting, the algorithm was terminated in the 82nd iteration, when the limiting operation conditions were not fulfilled due to two overloaded lines. The last feasible configuration was found in 81st iteration after 81 elements have been disconnected. In this stage, all the bus couplers have been switched off and considerable number of transformers have been disconnected. The optimization was not reached – SC currents in some nodes are still higher than the set SC withstand current rating. The number of nodes, where the SC condition is not fulfilled, is depicted for each iteration in Figure 2. The number was reduced by 75% between the initial configuration and the last feasible configuration, following c. linear trend. Actual numbers are normalized to the initial state. The difference between the initial and the final configuration is also depicted in Figure 3, where the

network topology is visualized. Red squares indicate nodes, where the SC condition is not fulfilled, green circles indicate nodes, where SC condition is fulfilled.
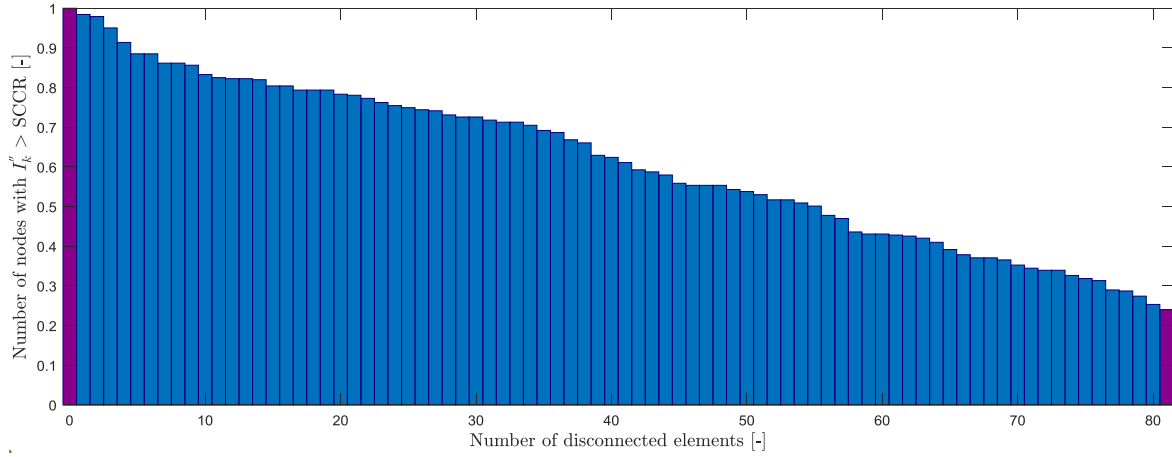

Figure 2: Number of nodes unfulfilling SC conditions throughout the procedure simulation
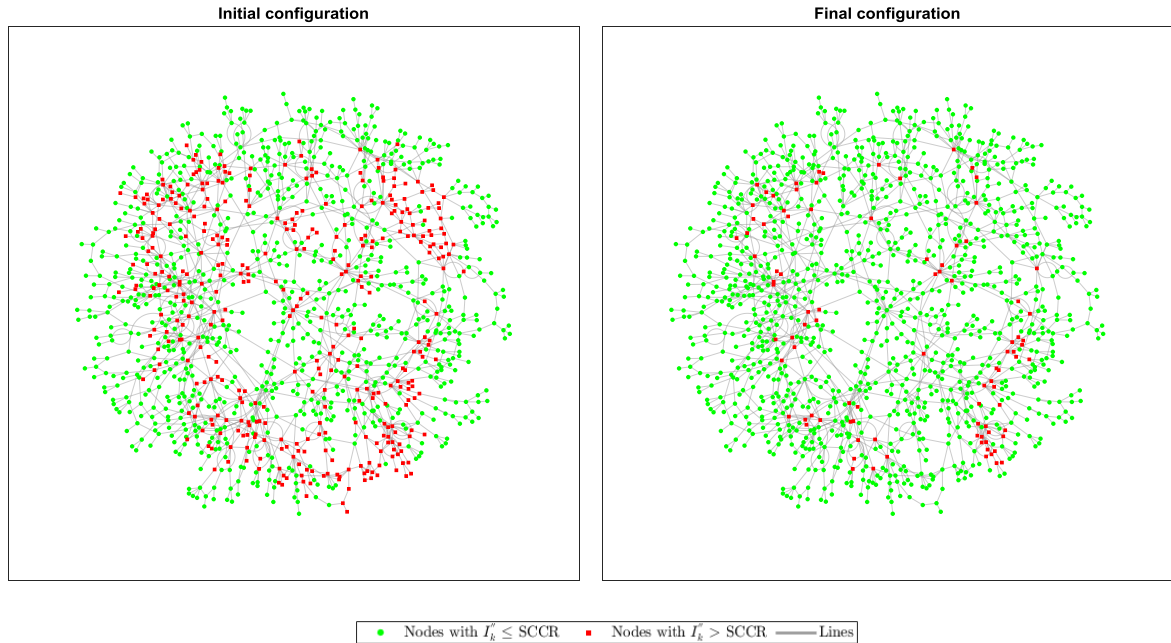

Figure 3: Comparison of fulfilling SC conditions in nodes for initial and final network configuration

The magnitudes of calculated SC currents throughout the reconfiguration procedure simulation are hinted via boxplot in Figure 4. The blue box covers values between 1st and 3rd quartiles, the red tick inside is the median, and the red points are outliers representing c. maximum 1% of values. The green line represents set SCCR value. An interesting phenomenon is to be seen between 11th and 49th iteration, where the maximum SC current in the network has not dropped despite 38 elements have been disconnected. Otherwise, the distribution of SC currents is quite consistent throughout the simulation with both the maximum and median values dropping roughly linearly. The values of I_k'' are normalized its maximum value in the initial configuration.
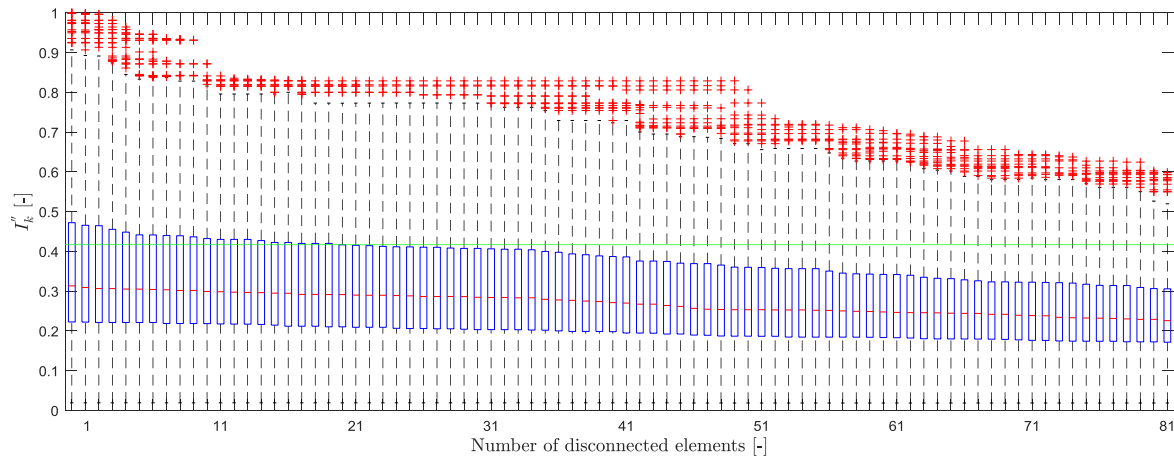
Figure 4: Statistical distribution of SC currents magnitudes in LV nodes throughout the procedure simulation

5   CONCLUSION

A simple and time-saving algorithm of mesh network reconfiguration leading to minimizing SC currents was designed and applied to a mesh MV/LV network model. The results show that sequential disconnection of selected elements in the network can lead to significant SC currents reduction while the operating conditions remain within defined limits. This procedure should be applied to real mesh networks providing valuable information for DSOs. Applying the specific procedure setting to the model has shown that reached SC current reduction is not sufficient since the SC currents in several nodes remain higher than usual SCCR of LV switchboards. That implies consideration of deploying further measures and including them into the algorithm. The proposed algorithm will be further enhanced, described and tested in future publications.

ACKNOWLEDGEMENT

REFERENCES

[1]   Y. Liao, Y. Weng, G. Liu and R. Rajagopal, "Urban MV and LV Distribution Grid Topology Estimation via Group Lasso," in IEEE Transactions on Power Systems, vol. 34, no. 1, pp. 12-27, Jan. 2019, doi: 10.1109/TPWRS.2018.2868877.

[2]   Toman, P., Drápela, J., MIŠÁK, S. et al.: Provoz distribučních soustav. 1. 1. Praha: České vysoké učení technické v Praze, 2011. 264 s. ISBN: 978-80-01-04935-8. (CS)

[3]   Anderson, P., Analysis of faulted power systems. IEEE Press power system engineering series, IEEE Press, 1995. ISBN: 9780780311459.

[4]   Anyaka, Boniface & Ozioko, Innocent. (2020). Transmission line short circuit analysis by impedance matrix method. International Journal of Electrical and Computer Engineering (IJECE). 10. 1712. 10.11591/ijece.v10i2.pp1712-1721.

[5]   IEC 60909-0:2016, Short-circuit current calculation in three-phase a.c. systems – Calculation of currents

PREFERENCE BASED AND IDEAL MULTI-OBJECTIVE OPTIMIZATION APPLIED ON HIGH-TORQUE FERRITE ASSISTED SYNCHRONOUS RELUCTANCE MACHINE

Ladislav Knebl

Doctoral Degree Programme (4), FEEC BUT

E-mail: ladislav.knebl@vutbr.cz

Supervised by: Cestmir Ondrusek

E-mail: ondrusek@feec.vutbr.cz

Abstract: This paper introduces comparison of optimization algorithms applied on high-torque ferrite-assisted synchronous reluctance machine. The comparison is focused not solely on two algorithms within the same multi-objective optimization approach - preference based or ideal, but also on comparison of these two approaches. The genetic algorithm and self-organizing migrating algorithm in both approaches are used to find optimal solution. The optimization goal is an optimal parameter combination to achieve the highest torque and power factor, while developing the lowest torque ripple. The optimized design will be evaluated by the 2D finite element analysis in steady-state analysis.

Keywords: steady-state, synchronous reluctance motor, finite element analysis, optimization

1   INTRODUCTION

The synchronous motor (SM) branch development has taken a big leap during the last 30 years [1], mainly due to discovery of rare-earth magnets - either neodymium-iron-boron or samarium-cobalt compounds, but also due to the pressure put on efficiency in area of electric motors (EM) (new more strict European standards - IEx). Nowadays, the situation in the SM development has reached a point, where the SMs are working with superior efficiency over other drive types, but also due to the increased permanent magnet (PM) prices the PM-assisted SMs have become expensive solutions.

The low-cost SMs have thus become attractive research topic aiming to develop cheaper SM competitor to low-cost induction machines. The low-cost machines suitable for traction applications, the topic of this paper, taking cost effective point of view into account are synchronous reluctance machines, while only PM-assisted synchronous reluctance machines are usually considered due to the higher torque, thus higher power density ratio. The price of the PMs is affected by multiple factors, the main being the geopolitical factor, majority of rare-earth ores is located in a few countries. Therefore, the solution of this price reduction approach lays in a use of cheaper PM material - ferrite magnet. The price of ferrites seems to be not affected by the price fluctuations of other PM ores and its cost remains just a portion of rare-earth competitors.

The topology choice, regardless of its undisputed importance is just a starting point of the whole design process. The next step in the machine design process is to provide reasonably chosen set of geometry parameters and their optimal values. The simplest way of optimization could be performed by trial and error approach - parametric analyses or manual testing. The optimal design could be found also by optimization algorithms (OA), which is commonly used technique resulting usually in better and efficiently found solutions. The commonly used algorithms in EM optimization design are the evolutionary algorithms. The best known algorithm in this group is the genetic algorithm (GA), that is found in many variations and also in both preference based and ideal multi-objective optimization (MOO) versions. The other OA that is considered in this paper, that is also found in

both MOO variations is self-organizing migrating algorithm (SOMA).

This paper considers high-torque ferrite-assisted synchronous reluctance (FASR) machine used in ship propulsion application for an MOO algorithm comparison using three objectives as an aim for optimization. These objectives, that are evaluated by 2D finite element analysis, are as follows - electromagnetic torque, torque ripple and power factor.

2 OPTIMIZED MACHINE

The investigated machine is the FASR machine has ten poles and is equipped with the integral slot winding distributed in 60 slots. The machine dimensions are listed in Table 1 and the parametric geometry is depicted in Figure 1. The optimized machine has three flux barriers, that was found optimal for this machine type within the limited machine dimensions.

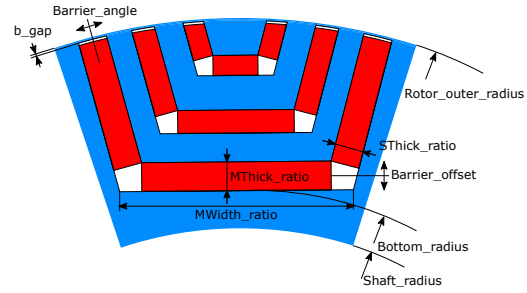| Parameter	Symbol	Value
Rated output torque	T_N	3100 Nm
Rated speed	n	150 min^{-1}
Stator out. diam.	D_{out}	584 mm
Stator inner diameter	D_{in}	470 mm
Stack length	L_{stk}	306 mm

Table 1: Key machine parameters

Figure 1: Parametric optimized geometry

The parameters MThick_ratio and SThick_ratio are related to the rotor space volume. A barrier width parameterized by the MWidth_ratio is related to the pole-width. Other parameters are used to distort the equidistant barrier placement Barrier_offset or barrier end angle Barrier_angle. Both outer and inner rotor radius defined Rotor_outer_radius, Shaft_radius respectively are fixed during the optimization. The bridge between the barrier end and air gap b_gap is fixed as well to 0.5 mm. The number of optimized parameters is therefore 15 (5 parameters for each barrier) plus parameter Bottom_radius defining the bottom barrier placement.

3 OPTIMIZATION ALGORITHMS

The OA is according to [2] is a programmable process or function, that takes some sort of input parameter or value that modify characteristics of the device or function to find the minimum or maximum value of a result or a output. The evolutionary algorithms works on repeating certain algorithm procedures, when each repeated procedures combination, i.e. generation in GA or migration in SOMA, starts where the previous ended. By this repetition is supposed to find better design each iteration and therefore ideally result in an globally (best result in a investigated space) optimal solution. The preference based MOO (PB-MOO) works on a principle of combining the normalized (dimensionless) objective values with their weight factors, increasing or decreasing aim of the optimization into single value, which is then minimized. To this combination of weight coefficients and normalized objectives is usually referred to as a "Cost function" or a "Fitness function". Algorithms working on this principle are usually less sophisticated and less efficient in the global optimum search. The ideal MOO (I-MOO) is on the other hand programmed to optimize the objectives equally, without influencing the algorithm objectives aims. This kind of algorithm is easier to initialize, since no weight coefficients nor desired values are required. I-MOO algorithms are more sophisticated and the result of optimization is easier to interpret, since each individual carries direct objective information not only cost function value. Also instead of single optimal solution, whole set of optimal solutions is delivered, this set is called Pareto optimal set of solutions.

3.1 GENETIC ALGORITHM

This algorithm is based on the Darwinian theory of evolution. The principle is based on combination of following principles - survival of the fittest, reproduction and the occasional positive mutation [2]. The repetition of these procedures ideally results in a optimal solutions. The whole process of GA is depicted in Figure 2(a). The I-MOO is following the same principle, but instead of sorting the population based solely on the cost function are the individuals sorted by non-dominant sorting process. This procedure is sorting individuals in Pareto fronts. In each front are individuals, where improvement of one objective leads to a decay of a second or other objective and also according to the distance between the solutions within the same Pareto set. Many I-MOO versions of GA were applied in EM design, but the most popular version, that is also used in this paper, is non dominated sorting genetic algorithm (NSGA-II).
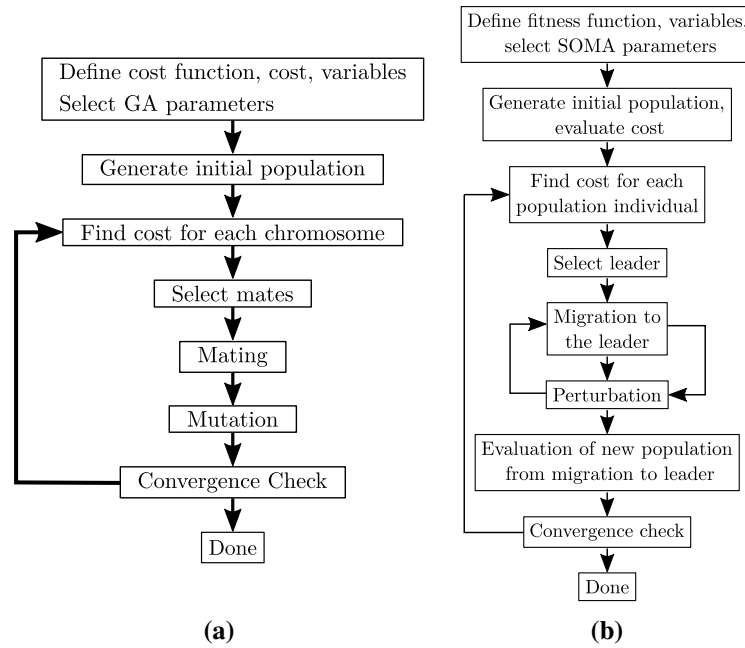

Figure 2: (a) Principle of GA (b) Principle of SOMA

3.2 SELF-ORGANIZING MIGRATING ALGORITHM

The principle of SOMA is based on the predator-prey situation, where is a herd of predators searching for a prey in a specific area. The original author is in [3] presenting the behavior on a pack of wolves searching for a food. An assumption is, that one member of predator herd is always closest to the prey, the member becomes a herd leader, and other members are traveling toward him. Seldom occur a situation, where one predator become even closer to the prey comparing to the original leader and other members changes traveling trajectories toward the new leader. This procedure, called migration, is repeated multiple times until the prey is found. The prey is in optimization terminology representing the global optimum and the herd of predators is the optimized population. In PB-MOO the global solution has the lowest cost function value. In I-MOO is the "prey" represented by the Pareto set of optimal solutions toward the population is migrating to. The SOMA principle is depicted in 2(b).

3.3 OPTIMIZED PARAMETER RANGES AND ALGORITHMS SETTING

The geom. parameters defined are varied during the optimization within the ranges listed in Table 3. The optimization algorithms have specific set of parameters that allow the user to influence the optimization processes. All the parameter values in Table 4 are chosen based on OA founders recom-

Parameter	Flux-barrier 1	Flux-barrier 2	Flux-barrier 3
MWidth_ratio, -	$0.2 \div 0.85$	$0.2 \div 0.85$	$0.2 \div 0.85$
MThick_ratio, -	$0.2 \div 0.85$	$0.2 \div 0.85$	$0.2 \div 0.85$
SThick_ratio, -	$0.2 \div 0.85$	$0.2 \div 0.85$	$0.2 \div 0.85$
Barrier_offset, (mm)	$-2 \div 2$	$-2 \div 2$	$-2 \div 2$
Barrier_angle, (deg)	$10 \div 22$	$-3 \div 3$	$-3 \div 3$
Bottom_radius, (mm)	$165 \div 180$		

Table 3: Boundaries of the optimized parameters.

mended values or the authors good experience in EM optimization. The common parameter across the algorithm is the fitness function evaluation (FFE), that set the same boundaries in terms of algorithm time demands to all OA. Therefore 20,000 FFE evaluations were supposed to be done by each OA. Weight coefficients in both PB-MOO algorithms were chosen to equally optimize all objectives.

OA/Parameter	SOMA	MOSOMA
Initial population	50	50
Mig. agents (SOMA-all)	-	15
Number of mig. steps	20	10
Mig. path length	2.1	1.7
Perturbation	0.3	0.1

OA/Parameter	GA	NSGA-II
Initial population	50	50
Crossover	3	0.5
Mutation rate	0.2	0.2

Table 4: OA parameters initialization

4 OPTIMIZATION RESULTS

The optimized results from PB-MOO are depicted as a development of the cost function value over the individual count, whereas the I-MOO algorithms figure depict distributed solutions in the objective space. Optimal solutions are highligled in all figures with the corresponding objectives value.

All algorithms except the SOMA algorithm evaluated 20,000 FFE, the SOMA algorithm met the convergence criteria sooner, therefore the algorithm finished. The results clearly shows the advantage of the I-MOO optimization procedures over PB-MOO approach. Both I-MOO algorithms sufficiently found solutions, that meet the desired objectives listed in Table 1, while the NSGA-II algorithm delivers evenly distributed solutions on and between the Pareto fronts. The PB-MOO algorithms failed to find a optimal solution in this application, the reason could be the high number of optimized parameters, that favors the I-MOO solvers. Chosen NSGA-II optimized result (highlighted by a star in Figure 3(d)) flux-density surface plot is depicted in Figure 4.

5 CONCLUSION

This article presents the comparison of four optimization algorithms in both I-MOO and PB-MOO approaches applied to the ferrite-assisted synchronous reluctance machine. The selected OAs that exist in both versions are GA and SOMA. The I-MOO approach of both algorithms resulted to be superior over their PB-MOO counterparts, while GA in both MOO approaches dominated over SOMA. This could be caused by relatively high amount of optimized parameters within the decision space. The optimization goal was reached by both I-MOO algorithms with higher torque ripple in case MOSOMA.
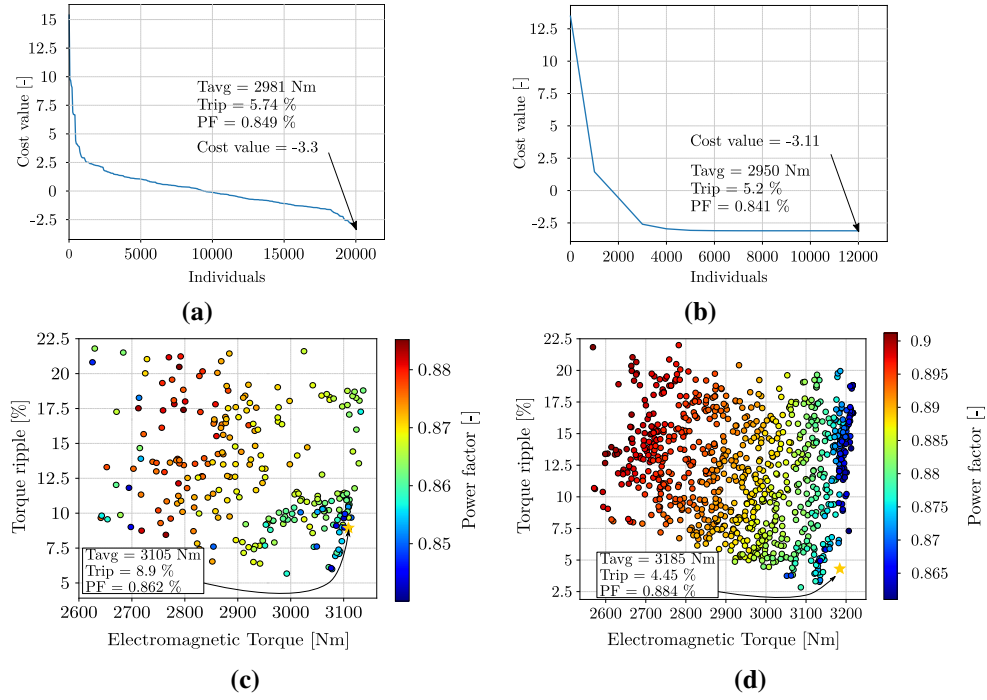
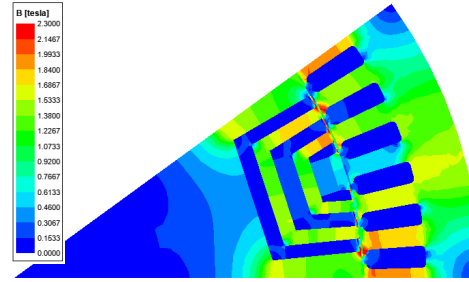**Figure 3:** a) Principle of GA b) Population sorting in I-MOO version of GA

Figure 4: Flux-density surface plot in optimized machine

REFERENCES

[1] Gieras, Jacek and Wing, Mitchell. *Design of synchronous motors with rare-earth surface permanent magnets*, 08 1994.

[2] Randy Haupt. *Practical genetic algorithms*. John Wiley, Hoboken, N.J, 2004.

[3] Donald Davendra and Ivan Zelinka. *Self-Organizing Migrating Algorithm: Methodology and Implementation*, Volume 626. 01 2016.

239

CONSTRUCTION OF THE ACTIVE MAGNETIC BEARING CONTROL SYSTEM

Jiri Ctibor

Doctoral Degree Programme (5), FEEC BUT

E-mail: jiri.ctibor@vut.cz

Supervised by: Ivo Pazdera

E-mail: surname@feec.vutbr.cz

Abstract: This paper talks about the design of active magnetic bearing control system. The main design issues are proposed and discussed. The discussion covers the control system requirements, selecting the main components, and proposes the own solution. Finally, the designed prototype is shown and its concept is verified by an experiment utilizing current control loops at their maximal capacity and electromagnetic interference is observed.

Keywords: AMB, Active Magnetic Bearing, Control System, ARM, Digital Signal Processing

1   INTRODUCTION

High speed electrical machines are the current state-of-the-art in the electric motor and drives department. With the growing demand on drives with higher rotational speeds, it is also going with growing demands on the mechanical design of such machines additionally on their bearings. The conventional mechanical bearings are insufficient at high rotational speeds. There are basically two possible substitutions of a mechanical bearing. First of them are hydrodynamic bearings having disadvantages such as low dynamics and damping capability. The second type is magnetic bearings. In magnetic bearing systems, the rotor desired position is kept by a type of electromagnetic force. In this article is proposed a concept of active magnetic bearing system. The system is still being developed, its current control capability is only one radial bearing but thanks to its modular concept it will be extended to control the whole high-speed rotational machine with two radial and one axial bearing.
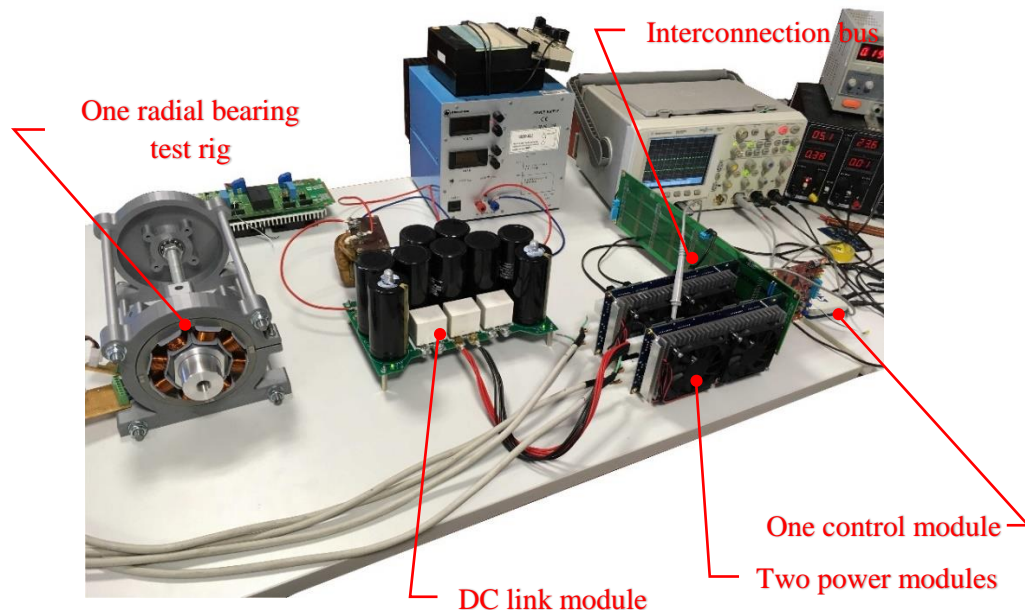

Figure 1: test rig overview

2   DEFINITION OF A PROBLEM

Active magnetic bearing is a complex electromechanical system performing a task with five degrees of freedom position control. As apparent from Figure 2 the rotor position could be defined in Cartesian system by five axes: x_1, y_1 controlled by radial bearing 1; x_2, y_2 controlled radial bearing 2, and finally z controlled by the axial so-called thrust magnetic bearing. The system complexity becomes higher when realized that control of the position in one axis is done by so-called differential actuators. The differential actuators are two opposite electromagnets, see Figure 3, biased together producing the desired forces used to adjust the position of the rotor in the respective axis. The principle and advantages of a differential actuator are explained in [1] and [2]. Substantial fact for design of the control system is that the differential actuator contains two coils energized by two different currents. That means to control the rotor position in five axis with five differential actuators, literally it is required to control ten different electrical currents. By recommendation from experienced colleagues from other universities, the bandwidth of the position control should be at least 10 kHz. To guarantee the proper performance and dynamics of the position control loops, the subsequent current control loops have to be at least by one order faster than the position loop. Those requirements define the performance of the system, which has to be able to control ten currents in the 10 μs window and five positions in the 100 μs window.
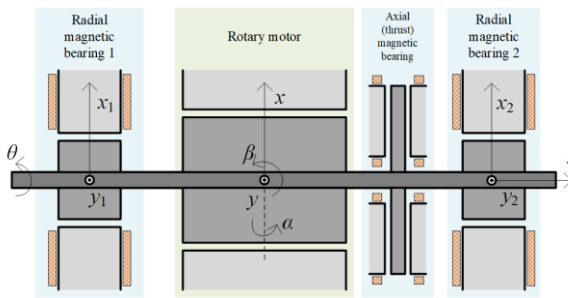

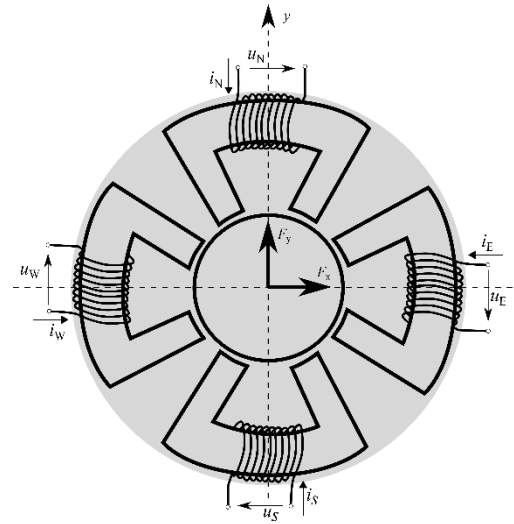**Figure 2:** AMB system radial section view **Figure 3:** radial bearing, axial section view

3   SYSTEM CONCEPT

Regarding to the system complexity mentioned in Chapter 2, it was a big issue how to design such a large system. Respecting the low experience with AMB systems, the most familiar concept is exposed to be a modular system. In that way, there are two main types of modules in the system, see Figure 4. First of them is the power module. Each power module is capable to drive one differential actuator, so the final count of these modules is five. The second module type is a control board. At the first iteration, one control board is designed to be capable to drive two differential actuators of one radial bearing. All control system modules are connected using an interconnection bus, which is a bare PCB attached by PCIe connectors, so the control and power modules are tightly interconnected between each other to ensure the robustness of the whole system.

The interconnection bus arranges three auxiliary power supply buses. The 24 V power supplying bus is designed to supply the SKF eddy current sensors, whose detailed description is introduced in the next chapter. The 5 V power supply is for MCUs power supply and current sensing. The separate power supply 12 V is for the power MOSFETs driving circuit, because it is expected to be burdened by electromagnetic interference. There is also no common ground between the modules to avoid electromagnetic interference. The switching and analogue signals are guided differentially, which is

a cheaper way than, for example, optically as proposed here [3]. The main power supply was calculated to be approximately 50 VDC and is connected separately from the DC link module.
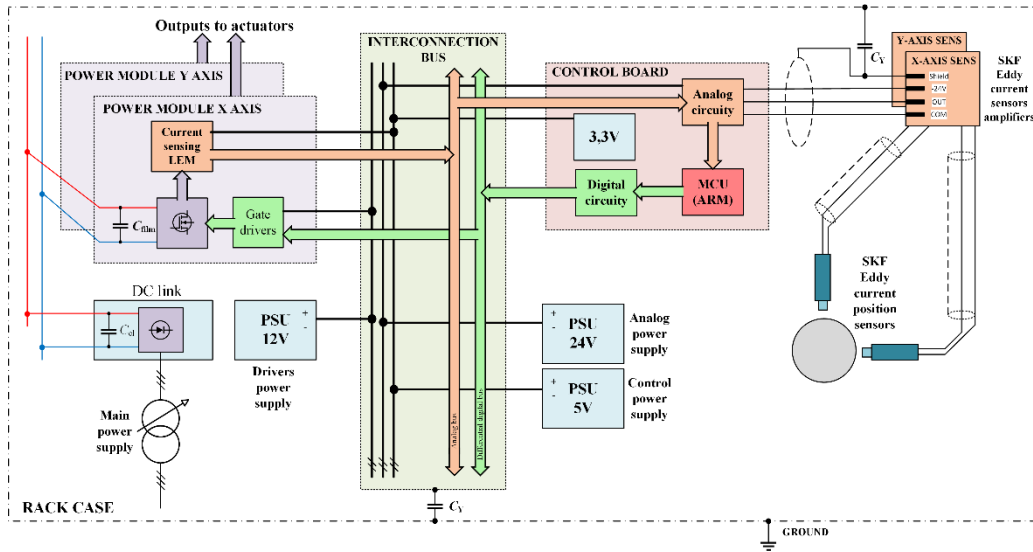

Figure 4: Overall system conception block schematic

3.1   CONTROL UNIT

There are several options which control unit to choose for such a complex system. The goal of this project is to make a stand-alone embedded system, which is compact and ready for practical use. Therefore, there are probably three options – PLC, FPGA, or Microcontroller. PLC control system for such application is not applicable in general because the majority of PLCs on the market haven't got such fast sampling periods as is desired for this application and moreover the programming languages of PLCs are not suitable for performing complex sequential algorithms. The FPGAs are more powerful, but the implementation of a sequential algorithm is complicated and finally it is not a cheap way by means of HW costs. Therefore, the most possible option for choosing a control unit is a microcontroller. As proposed in Chapter 2, AMB system has the character of fast digital signal processing tasks. The market offers for these applications so-called digital signal processors, such as Texas Instruments DSPs with C2000 cores used for example here [3]. However, it is commonly known that DSPs with Harvard architecture are going to be outdated in the near future which is proven, for example by NXP's (Freescale) termination these microcontrollers production. The DSP product lines are being substituted by a new architecture ARM Cortex-M core designed also for digital signal processing. Therefore, this kind of microcontroller was chosen, specifically STM32G4 product family. This microcontroller offers similar power of the floating-point core with a set of DSP instructions, peripherals with similar parameters, e.g. High-Resolution Timer (HRTIM) for generating up to 12 PWM signals, with maximal resolution 184 ps, 5×12-bit AD sigma-delta convertors or specific DSP peripherals - CORDIC for trigonometric functions acceleration and FMAC filter mathematical accelerator for implementation of digital HW filters.

3.2   POWER STAGE

There are plenty of articles describing advanced topologies of AMB power stages, such as shared-bridge convertors published at [4], [5] or [6]. This topology saves the count of the switching components but increases the drive algorithm complexity and brings challenges on the current caring capacity of the shared-phase leg which can be easily broken due to high current stress. Furthermore, this topology doesn't respect the selected modular conception, because all switching components need to be close together. Therefore, the traditional topology of power convertor has been chosen to feed every single coil by a half-bridge.
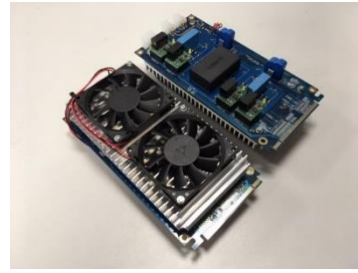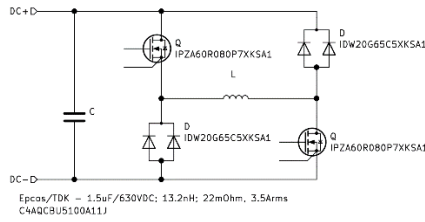
Figure 5: topology of a single converter **Figure 6:** two power modules (2 half bridges in each)

By means of the high frequency bandwidth of the current loop, also high PWM frequency needs to be introduced. To decrease the switching power losses and electromagnetic interference, novel semiconductor technology *SiC* power switches are employed, see Figure 5.

3.3 SENSING OF PHYSICAL VARIABLES IN THE SYSTEM

To enable high stiffness and damping of the bearings the axis position has to be sampled very precisely. The traditional inductive position sensors are insufficient for this job, because of its low bandwidth around 1 kHz. As proposed in Chapter 2, the position sensor optimal bandwidth is 10 kHz. A little research has been done to select proper position sensors. The following sensor covers our technical requirements: Micro-Epsilon offering capacitive sensors capaNCDT series and SKF offering CMSS eddy current position sensors. However, only SKF sensor fulfil also our financial budget requirements. The force of the actuators is strongly dependant on the coil currents, so the current also has to be sampled at comparatively high resolution and high bandwidth 100 kHz at least. Finally the following sensors are employed in the system (Table 1).
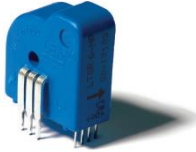
Current sensor	Position sensor
LEM Current Transducer LTSR 15-NP	SKF CMSS 75-LM1
Hall effect	Eddy currents
Nominal RMS current 15 A Frequency bandwidth 200 kHz Output voltage $2.5 \pm (0.625 \times I_p/I_{pn})$	Measurement range 0.25 to 2.25 mm Scale factor 7.87 V/mm Frequency response DC to 10 kHz Linear characteristic ±25μm for 5 m cable length

Table 1: presentation of selected physical sensors

4   EXPERIMENTAL RESULTS

For verification of the designed system functionality and robustness following experiment was performed. Two different actuators were energized in different instants approximately to 12 A, which is the full current capacity and mutual interference was observed. Experiment result very well, as apparent from Figure 7, there is no recognizable interference in the instant energizing second coil by full current step (time 10 ms). The data was acquired from the microcontroller by specialized routine via serial port and visualized in MATLAB.
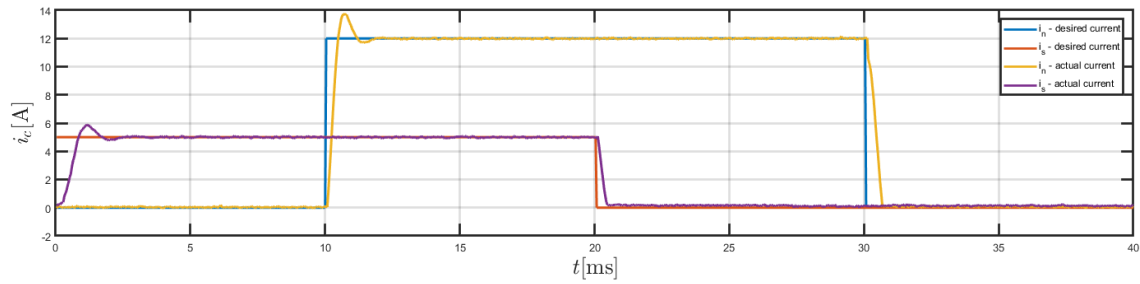
Figure 7: experimental result, current loop interreference robustness

5   CONCLUSION

In this article was presented the concept and design of the active magnetic bearings control system and several issues were discussed and explained. As a verification of system functionality an experiment was performed and shown its results. The project is in stage that hardware layer can be considered as finalized and tested. Actual work is software development, which can be divided into two main layers. First of them is low level layer covering peripheral configuration and using of hardware abilities of the MCU together with PWM generation and PID current control loop design. This part could be also considered as finalized. Second high-level software layer will be about the position sensing and control, which will be definitely covered in future work and publications

REFERENCES

[1]   Akira Chiba ... [et al.]., Magnetic bearings and bearingless drives [Online-Ausg.]. Amsterdam: Elsevier/Newnes, 2005.

[2]   J. Ctibor, I. Pazdera and J. Knobloch, "One-axis Radial Active Magnetic Bearing Simulink Model with Respect to Nonlinear Ferromagnetics", in 2018 18th International Conference on Mechatronics - Mechatronika (ME), 2018, pp. 1-6.

[3]   A. Schulz, M. Schneeberger and J. Wassermann, "A Reliable Switching Amplifier driving an Active Magnetic Bearing - Experimental Results", in 2005 IEEE International Conference on Industrial Technology, 2005, pp. 389-394.

[4]   Z. Hu, D. Jiang, H. Sun and R. Qu, "A Shared-bridge Converter with Reversed Current Direction for Active Magnetic Bearing Drive", in 978-89-5708-313-0, 2019, pp. 326-332.

[5]   D. Jiang, T. Li, Z. Hu and H. Sun, "Novel Topologies of Power Electronics Converter as Active Magnetic Bearing Drive", IEEE Transactions on Industrial Electronics, vol. 67, no. 2, pp. 950-959, 2020.

[6]   J. Yang, D. Jiang, H. Sun, A. Li and Z. Liu, "Series-winding Topology Converter for Active Magnetic Bearing Drive", IEEE Transactions on Industrial Electronics, pp. 1-1.

ELECTROMAGNETIC ANALYSIS OF LINE-START SYNCHRONOUS RELUCTANCE MACHINES OPTIMIZED BY MEANS OF TOPOLOGY OPTIMIZATION

Iveta Lolová

Doctoral Degree Programme (1), FEEC BUT

E-mail: Iveta.Lolova@vutbr.cz

Supervised by: Jan Bárta

E-mail: bartaj@feec.vutbr.cz

Abstract: This paper deals with the electromagnetic analysis of a line-start synchronous reluctance motor optimized by means of topology optimization based on the normalized Gaussian network. Topology optimization using the normalized Gaussian network represents the novel approach for electrical machines improvements. This article evaluates this approach as promising due to the electromagnetic analysis results of selected individual of the Pareto front. The main contribution of this article is a comprehensive electromagnetic analysis of line-start synchronous reluctance machine optimized by the above mention method.

Keywords: Ansys Maxwell, electromagnetic analysis, FEM, line-start synchronous reluctance motor, the normalized Gaussian network, topology optimization, transient analysis

1 INRODUCTION

There are worldwide efforts for a reduction in energy consumption within the framework of sustainable development. The electrical rotating machines are one of the biggest consumers of electricity. Thus, the efficiency requirements for electrical rotating machines is raised considerably by the efficiency classification standard IEC 60034-30. Presented research deals with the line-start synchronous reluctance machines (LSSynRM) due to the promising performances of this type of machine [3]. Although LSSynRM can reach the high-efficiency requirements of the efficiency classification standard, there is a demand for further improvements. The improvements of LSSynRM are done by the means of topology optimization (TO) based on the normalized Gaussian network (NGnet). Comprehensive electromagnetic analysis of LSSynRM optimized through the described methodology is the main goal of this paper and it can be found in the following sections.

The paper is organized as follows: Section 2 presents the TO based on NGnet in the general and developed method for its application to LSSynRM. Moreover, it contains the results of TO performed in [2] and it presents one individual geometry selected from the Pareto front for the more extensive electromagnetic analysis because the electromagnetic analysis was not the main goal of [2]. So the electromagnetic analysis is presented in Section 3. The analysis is divided into three parts: 3.1 Starting performances, 3.2 Steady-state analysis and 3.3 Synchronization capability. Section 4 concludes the paper.

2 TOPOLOGY OPTIMIZATION BASED ON THE NGNET OF LINE-START SYNCHRONOUS MACHINES

TO is an approach when the geometry of the optimized part of an electrical machine is changed directly without any previous expectations or limitations for topology. This is the main advantage of TO compared to parametric optimization, which works with predefined geometry such as number of flux barriers, bars etc. The authors of [1] explains this matter in more detail. They also discuss the disadvantages of TO such as a high number of optimized parameters and the possibility of achieving

unfeasible geometries. These disadvantages can be partly solved by TO based on NGnet, which is described in [2] and [3] in more detail. The principle of the TO is shown in Fig. 1. The geometry of the rotor is fully determined by the assignment of pixels in the rotor to the predefined material such as iron or aluminum depending on the values of NGnet.
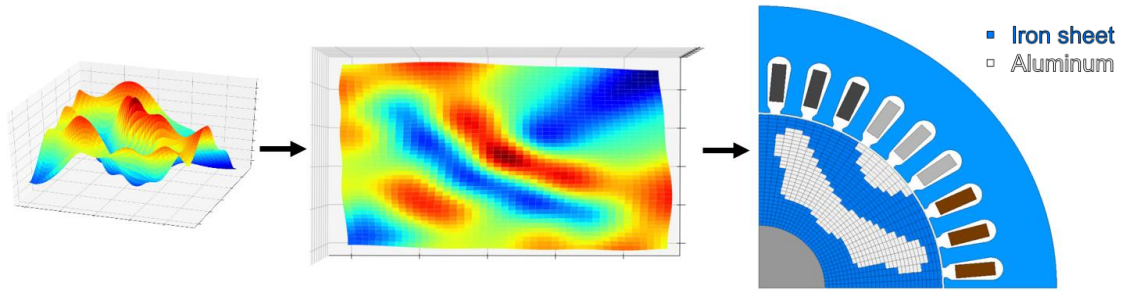

Figure 1: The NGnet defines geometry of LSSynRM rotor.

The TO considering NGnet was firstly presented by authors in [4]-[6] and it was used for permanent magnet and reluctance synchronous machines optimization. Although there are notable complications of application TO to line-start synchronous machines, the author of this paper developed the methodology for TO. The methodology includes a developed algorithm for the evaluation of a single individual, which leads to a significant reduction of the computational part. The comprehensive description of this algorithm can be found in [2] and [3] as it is out of the scope of this paper. The software tools used for TO are SyMSpace [7], Ansys Maxwell Electronics Desktop and Python.

The LSSynRM was selected as the type of machine for the optimization. LSSynRM is based on a conventional induction motor with a rated power of 2.2 kW and a rated speed of 1500 rpm. The main objectives of the optimization were: efficiency, power factor and torque ripple in the steady-state. The running time of TO was approximately one and half month and nearly 15 000 individuals were evaluated. The obtained Pareto front with roughly forty individuals is shown in Fig. 2. Although the results of the Pareto front were already presented in [2], this paper deals with the more extensive electromagnetic analysis of the selected individual. The individual selected for comprehensive electromagnetic analysis is highlighted in Fig. 2 and its topology is shown in Fig. 3, where a grey part is an iron sheet and orange parts are aluminum. The electromagnetic analysis is performed in the software Ansys Maxwell Electronics Desktop and it is presented in the following section.
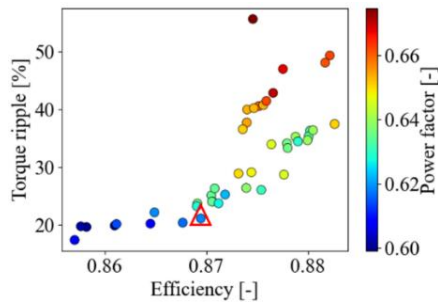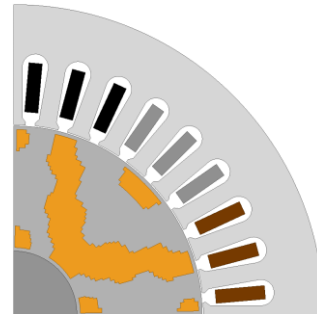

Figure 2: Pareto front [2].

Figure 3: The selected geometry for comprehensive analysis.

3 COMPREHENSIVE ELECTROMAGNETIS ANALYSIS OF SELECTED INDIVIUAL

The geometry of the selected individual strongly depends on the shape of NGnet, which is changed by the genetic algorithm DECMO2 [8] in SyMSpace during TO. It can be seen in Fig. 3, that the main flux barrier and main flux guide are rather thick. The thickness is caused by the low number of Gaussian functions used in NGnet. The low number of Gaussian functions, and thus a small number of optimized parameters, was necessary to ensure the reasonable time of optimization. This produces

not ideal geometry with only one main flux barrier. Three bars are located near to surface of the rotor. Their functionality is the same as in an induction motor. Besides, two more bars are located in the inner area of the rotor, which do not serve as squirrel cage due to their location too far from the surface of the rotor. Also, the whole geometry seems insufficient from the mechanical point of view because of the thin iron bridge between the shaft and the main flux barrier. Even though the mentioned deficiencies in geometry, an electromagnetic analysis was performed to evaluate the performance of obtained LSSynRM as well as the performed topological optimization.

3.1 STARTING PERFORMANCES

The analyzed machine is a line-start synchronous reluctance machine. Thus, the supply source is set as a voltage source with a nominal voltage of 230 V. The stator winding connection is a star. The step time of transient analysis is set to 0.1 ms and the final time is 300 ms. The fan load with nominal load torque 14 Nm at 1500 rpm is connected to the shaft and the load moment of inertia is equal to the moment of inertia of the rotor. The starting process of the motor is quite dynamic and it is shown in Fig. 4. The speed gradually rises and it stabilizes at 1500 rpm in 180 ms. The torque during the starting process is shown in Fig. 5. The torque dependency on the speed is shown in Fig. 6, where the point of synchronization can be seen. The starting current is ten-times bigger compared to the nominal rated current at steady-state and it is shown in Fig. 7.
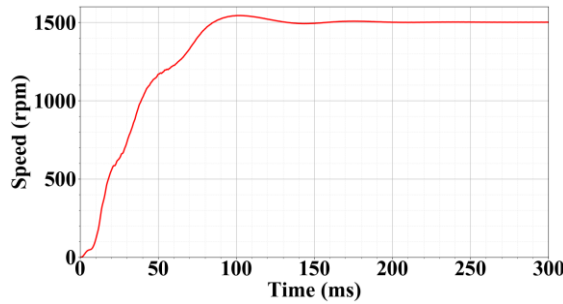
Figure 4: The speed during the starting process.
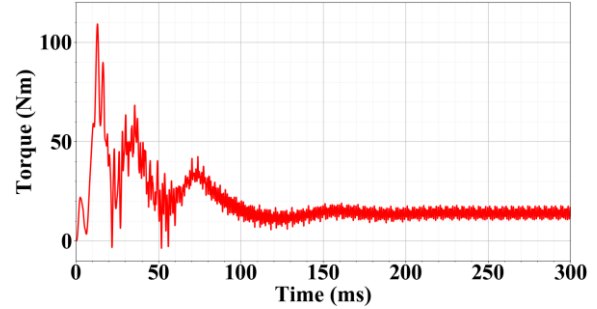
Figure 5: The torque during the starting process.
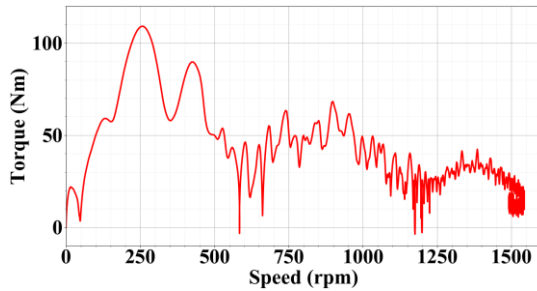
Figure 6: The torque dependency on the speed.
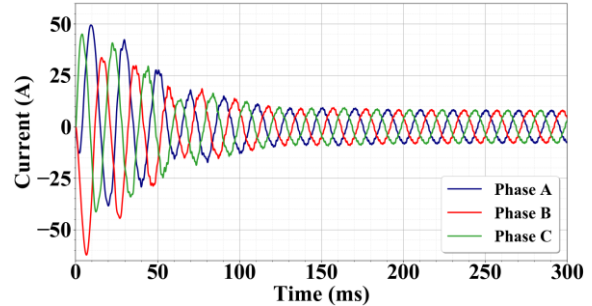
Figure 7: The currents during the starting process.

Rated power	kW	2.2
Rated speed	rpm	1500
Rated torque	Nm	14.1
Rated line-to-line voltage	V	400
Rated phase current	A	5.52
Efficiency	-	0.862
Power factor	-	0.629
Fundamental harmonic of flux density	T	0.761
Torque ripple	%	25.45

Table 1: Main parameters of selected individual at steady-state.

3.2 STEADY-STATE ANALYSIS

The last two periods of the starting process analysis described before are used for steady-state analysis. The main parameters of steady-state analysis are shown in Tab. 1. The motor has a power factor 0.629 and an efficiency of 86.2 %, thus this machine, unfortunately, does not reach even IE3. The magnetic flux density distribution is shown in Fig. 8 and the magnetic flux density in the middle of the air-gap is shown in Fig. 9. It can be seen that flux is forced to go also through the inner area of the rotor and the shaft. The flux density in this area is almost 2 T. On the one hand, the saturation in this area can be lowered if there were not two inner bars that decrease permeance in this area. On the other hand, the saliency ratio could also decrease due to the absence of these inner bars.
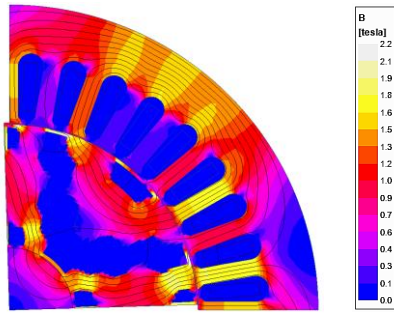

Figure 8: Flux density distribution in steady-state.
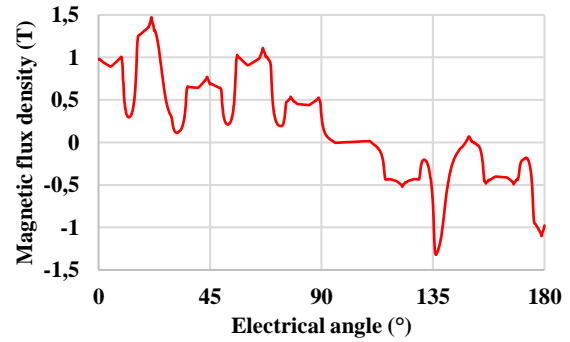
Figure 9: Magnetic flux density in the middle of the air-gap.

The torque ripple is shown in Fig. 10 and its value is 25. 6 % of the nominal torque. Fig. 11 shows the distribution of the losses in the motor at the steady-state. The resistive losses in stator represent almost 60 % of all losses. It is caused by a quite low power factor. The core losses are multiplied by the correction coefficient which considers the manufacturing effect on the material. The waveform of the current is shown in Fig. 12. The torque dependency on the load angle is shown in Fig. 13, but as this machine is supplied from the grid it stabilizes at a load angle of 21° corresponding to 14.1 Nm. The nominal load torque on the grid voltage supply is highlighted in Fig. 13.

Figure 10: The torque behaviour in steady-state.

Figure 11: The distribution of the losses in steady-state.

Figure 12: The currents behaviour in steady-state.

Figure 13: The torque dependency on the load angle.

3.3 SYNCHRONIZATION CAPABILITY

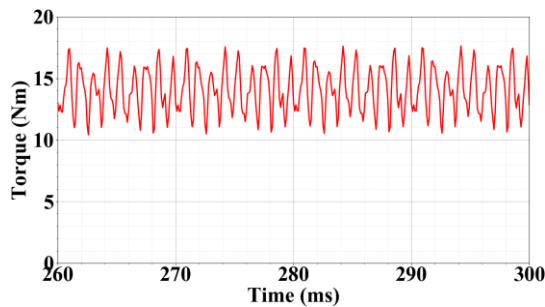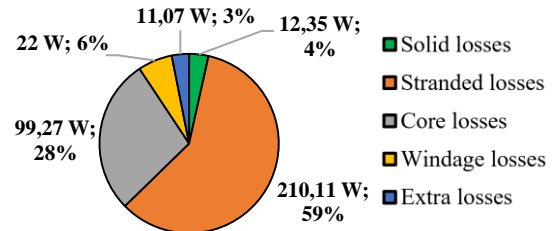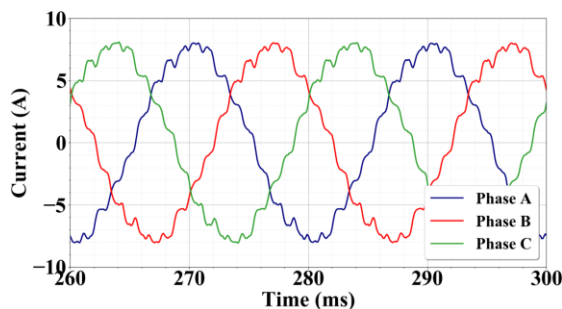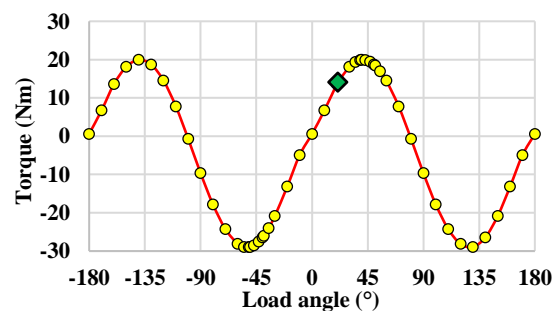The synchronization capability is crucial for the application of LSSynRM. It is determined by critical inertia, which can be synchronized corresponding to the given statical load torque. The critical inertia decreases with bigger load torque. The synchronization capability of selected individual is shown in Fig. 14.
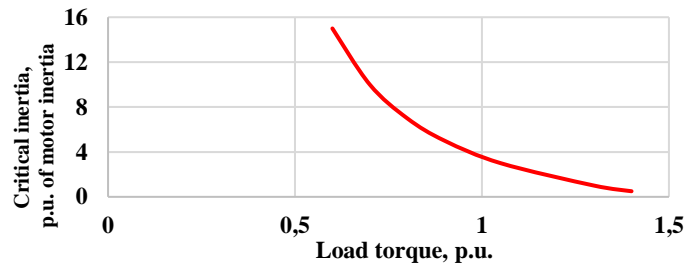

Figure 14: The synchronization capability.

4 CONCLUSION

The presented paper deals with the electromagnetic analysis of LSSynRM optimized by means of TO based on NGnet. The result of the analysis shows that the LSSynRM optimized by such an approach has the potential to reach considerably good performance. Thus, the TO based on NGnet is suitable for the improvement of LSSynRM, but there have to be done further development of this methodology for achieving the best results.

ACKNOWLEDGEMENT

REFERENCES

[1] A.-C. Zăvoianu, G. Bramerdorfer, E. Lughofer, and S. Saminger-Platz, "Multi-objective topology optimization of electrical machine designs using evolutionary algorithms with discrete and real encodings," in *Computer Aided Systems Theory – EUROCAST 2017*. Cham: Springer, 2018, pp. 331–338.

[2] I. Lolová, J. Bárta, G. Bramerdorfer, and S. Silber, "Topology optimization of line-start synchronous reluctance machine," *In 2020 19th International Conference on Mechatronics - Mechatronika (ME),* 2020, pp. 55–61.

[3] I. Lolova, "Topology optimization of the line-start synchronous machines," Brno, 2020. [Online]. Available: https://www.vutbr.cz/studenti/zav-prace/detail/125799.

[4] S. Sato, T. Sato, and H. Igarashi, "Topology optimization of synchronous reluctance motor using normalized gaussian network," in *IEEE Transactions on Magnetics*, vol. 51(3), IEEE, 2015.

[5] T. Sato, K. Watanabe, and H. Igarashi, "Multimaterial topology optimization of electric machines based on normalized gaussian network," in *IEEE Transactions on Magnetics*, vol. 51(3), IEEE, 2015.

[6] F. Guo and I. P. Brown, "Simultaneous magnetic and structural topology optimization of synchronous reluctance machine rotors," in *IEEE Transactions on Magnetic*s, vol. 56, no. 10, pp. 1–12, 2020.

[7] S. Silber, W. Koppelstatter, G. Weidenholzer, G. Segon, and G. Bramerdorfer, "Reducing Development Time of Electric Machines with SyMSpace," in *2018 8th International Electric Drives Production Conference (EDPC) [online]*, IEEE, 2018, pp. 1–5.

[8] A.-C. Zăvoianu, E. Lughofer, G. Bramerdorfer, W. Amrhein, and E. P. Klement, "Decmo2: A robust hybrid and adaptive multi-objective evolutionary algorithm," in *Soft Comput*, vol. 19, Springer, 2015, pp. 3551–3569.

ASYMPTOTIC CHARACTERIZATION OF SOLUTIONS OF EMDEN-FOWLER TYPE DIFFERENCE EQUATION

Evgeniya Korobko

Doctoral Degree Program (2), FEEC BUT

E-mail: xkorob01@stud.feec.vutbr.cz

Supervised by: Josef Diblík

E-mail: diblik@feec.vutbr.cz

Abstract: The paper derives an asymptotic formula describing the long-time behaviour of a solution of a nonlinear Emden-Fowler type difference equation.

Keywords: Difference equation, Emden-Fowler type equation, asymptotic characterization, auxiliary system.

1 INTRODUCTION

We discuss the asymptotic behaviour of a solution of Emden-Fowler type difference equation

$$\Delta^2 u(k) \pm k^\alpha u^m(k) = 0, \tag{1}$$

where $k \in \mathbb{N}(k_0) := \{k_0, k_0 + 1, ...\}$, $k_0 \in \mathbb{N}$, $k_0 > 0$, $u \colon \mathbb{N}(k_0) \to \mathbb{R}$, $\Delta u(k) = u(k+1) - u(k)$ is the first and $\Delta^2 u(k)$ the second difference with $\alpha, m \in \mathbb{R}$, $m > 0$, $m \neq 1$. When the $+$ sign is assumed in (1), the below result holds only if $m = p/q$ where p and q are integer numbers such that $p - q$ is odd. Equation (1) is a difference analogue of the Emden-Fowler differential second-order equation, known from the theory of ordinary differential equations (we refer to [4]), being significant in astrophysics, cosmology, atomic physics, and other areas.

2 MAIN RESULT

We will prove that there exists a solution $u = u(k)$ of (1) such that

$$u(k) = ak^{-s} + bk^{-(s+1)} + O(k^{-(s+\gamma+1)}) \tag{2}$$

when $k \to \infty$ where

$$a = [\mp s(s+1)]^{1/(m-1)}, \quad b = as(s+2)/(s+2-ms), \quad s = (\alpha+2)/(m-1), \tag{3}$$

$\gamma \in (0,1)$ is a fixed number and O is the Landau order symbol big "O". The equation (1) was investigated in [3] where the existence of a solution with asymptotic behaviour determined by formula (2) was proved under the assumption that $s \in (-2, -1)$ and $m < 0$. Our main result below establishes the validity of formula (2) under a different set of conditions.

Theorem 1. *Let $s > 0$. If*

$$m < \frac{(s+2)(s+3)}{s(s+1)} \tag{4}$$

then there exists a solution $u = u(k)$ of equation (1) defined on $\mathbb{N}(k_0)$, where k_0 is sufficiently large, with asymptotic behaviour determined by the formula (2).

Proof. In the proof we will refer to some parts of the paper [3]. Let

$$u(k) = a/k^s + b/(k^{s+1})(1 + Y_0(k)), \tag{5}$$

$$\Delta u(k) = \Delta(a/k^s) + \Delta\left(b/k^{s+1}\right)(1 + Y_1(k)),$$

$$\Delta^2 u(k) = \Delta^2\left(a/k^s\right) + \Delta^2\left(b/k^{s+1}\right)(1 + Y_2(k))$$

where $Y_i \colon \mathbb{N}(k_0) \to \mathbb{R}$, $i = 0, 1, 2$ are new unknown functions. Then, equation (1) can be converted (for details we refer to [3, Part 3]) into a system

$$\Delta Y_0(k) = F_1(k, Y_0, Y_1) := (-(s+1)k^{-1} + O(k^{-2}))(-Y_0(k) + Y_1(k)), \tag{6}$$

$$\Delta Y_1(k) = F_2(k, Y_0, Y_1) := (-(s+2)k^{-1} + O(k^{-2}))(ms(s+2)^{-1}Y_0(k) - Y_1(k) + O(k^{-1})). \tag{7}$$

This will allow us to use [3, Lemma 2] (being a modification of [1, Theorem 8]) to prove Theorem 1. As the scheme of the proof is the same as that of Theorem 1 in [3], we will only emphasize in detail the parts that are different, referring the remaining parts to this source. Let $\varepsilon_i > 0$, $i = 1, \ldots, 4$, $\gamma > 0$ and $\beta > 0$ be fixed. Define, as in [3, Part 4], auxiliary functions $b_1(k) := -\varepsilon_1/k^\gamma$, $c_1(k) := \varepsilon_2/k^\gamma$, $b_2(k) := -\varepsilon_3/k^\beta$ and $c_2(k) := \varepsilon_4/k^\beta$. To apply [3, Lemma 2], the following inequalities

$$F_1(k, b_1(k), Y_1) < b_1(k+1) - b_1(k), \tag{8}$$

$$F_1(k, c_1(k), Y_1) > c_1(k+1) - c_1(k), \tag{9}$$

$$F_2(k, Y_0, b_2(k)) < b_2(k+1) - b_2(k), \tag{10}$$

$$F_2(k, Y_0, c_2(k)) > c_2(k+1) - c_2(k) \tag{11}$$

must hold whenever $-\varepsilon_3 k^{-\beta} \leq Y_1 \leq \varepsilon_4 k^{-\beta}$ in (8), (9) and $-\varepsilon_1 k^{-\gamma} \leq Y_0 \leq \varepsilon_2 k^{-\gamma}$ in (10), (11). Let us find the conditions for validity of inequalities (8) – (11). As we assume $m > 0$ and $s > 0$, we have $ms > 0$, $s + 1 > 0$ obtaining:

$$F_1(k, b_1(k), Y_1) = \left(-\frac{s+1}{k} + O\left(\frac{1}{k^2}\right)\right) \cdot \left(\frac{\varepsilon_1}{k^\gamma} + Y_1(k)\right) < \left(-\frac{s+1}{k} + O\left(\frac{1}{k^2}\right)\right) \cdot \left(\frac{\varepsilon_1}{k^\gamma} - \frac{\varepsilon_3}{k^\beta}\right)$$

$$< b_1(k+1) - b_1(k) = -\frac{\varepsilon_1}{(k+1)^\gamma} + \frac{\varepsilon_1}{k^\gamma} = \frac{\varepsilon_1 \gamma}{k^{\gamma+1}}\left(1 + O\left(\frac{1}{k}\right)\right).$$

Then we have the following

$$-\frac{s+1}{k^{\gamma+1}}\varepsilon_1 + \frac{s+1}{k^{\beta+1}}\varepsilon_3 + O\left(\frac{1}{k^{\beta+2}}\right) < \frac{\varepsilon_1 \gamma}{k^{\gamma+1}} + O\left(\frac{1}{k^{\gamma+2}}\right)$$

or, after simplifying,

$$\frac{s+1}{k^{\beta+1}}\varepsilon_3 + O\left(\frac{1}{k^{\beta+2}}\right) < \frac{\varepsilon_1(\gamma+s+1)}{k^{\gamma+1}} + O\left(\frac{1}{k^{\gamma+2}}\right).$$

If $k \to \infty$, then the last inequality hold either if $\gamma < \beta$, or if $\gamma = \beta$ and $\varepsilon_3 < \varepsilon_1(\gamma+s+1)/(s+1)$. Similarly, inequality

$$F_1(k, c_1(k), Y_1) = \left(-\frac{s+1}{k} + O\left(\frac{1}{k^2}\right)\right) \cdot \left(-\frac{\varepsilon_2}{k^\gamma} + Y_1(k)\right) > \left(-\frac{s+1}{k} + O\left(\frac{1}{k^2}\right)\right) \cdot \left(-\frac{\varepsilon_2}{k^\gamma} + \frac{\varepsilon_4}{k^\beta}\right)$$

$$> c_1(k+1) - c_1(k) = \frac{\varepsilon_2}{(k+1)^\gamma} - \frac{\varepsilon_2}{k^\gamma} = -\frac{\varepsilon_2 \gamma}{k^{\gamma+1}}\left(1 + O\left(\frac{1}{k}\right)\right)$$

implies either $\gamma < \beta$, or $\gamma = \beta$ and $\varepsilon_4 < \varepsilon_2(\gamma + s + 1)/(s+1)$. Hence, both the above inequalities hold if either $\gamma < \beta$, or if

$$\gamma = \beta, \quad \varepsilon_3 < \varepsilon_1 \frac{\gamma + s + 1}{s+1}, \quad \varepsilon_4 < \varepsilon_2 \frac{\gamma + s + 1}{s+1}.$$

In much the same way we estimate (recall that $ms > 0$):

$$F_2(k, Y_0, b_2(k)) = \left(-\frac{s+2}{k} + O\left(\frac{1}{k^2}\right)\right) \cdot \left(\frac{ms}{s+2} Y_0(k) + \frac{\varepsilon_3}{k^\beta} + O\left(\frac{1}{k}\right)\right) < \frac{ms}{k} \cdot \frac{\varepsilon_1}{k^\gamma} - \frac{s+2}{k} \cdot \frac{\varepsilon_3}{k^\beta}$$

$$+ O\left(\frac{1}{k^{\gamma+2}}\right) + O\left(\frac{1}{k^{\beta+2}}\right) + O\left(\frac{1}{k^2}\right) < b_2(k+1) - b_2(k) = \frac{\varepsilon_3 \beta}{k^{\beta+1}}\left(1 + O\left(\frac{1}{k}\right)\right).$$

Then, inequality

$$\frac{ms}{k^{\gamma+1}} \varepsilon_1 - \frac{s+2}{k^{\beta+1}} \varepsilon_3 + O\left(\frac{1}{k^{\gamma+2}}\right) + O\left(\frac{1}{k^2}\right) < \frac{\varepsilon_3}{k^{\beta+1}} \beta + O\left(\frac{1}{k^{\beta+2}}\right)$$

holds if either $\gamma > \beta$ (but this contradicts to above restrictions), or if $\gamma = \beta < 1$ and $\varepsilon_1 < \varepsilon_3(\gamma + s + 2)/ms$. Finally, we estimate

$$F_2(k, Y_0, c_2(k)) = \left(-\frac{s+2}{k} + O\left(\frac{1}{k^2}\right)\right) \cdot \left(\frac{ms}{s+2} Y_0(k) - \frac{\varepsilon_4}{k^\beta} + O\left(\frac{1}{k}\right)\right) > -\frac{ms}{k} \cdot \frac{\varepsilon_2}{k^\gamma} + \frac{s+2}{k} \cdot \frac{\varepsilon_4}{k^\beta}$$

$$+ O\left(\frac{1}{k^{\gamma+2}}\right) + O\left(\frac{1}{k^{\beta+2}}\right) + O\left(\frac{1}{k^2}\right) > c_2(k+1) - c_2(k) = -\frac{\varepsilon_4 \beta}{k^{\beta+1}}\left(1 + O\left(\frac{1}{k}\right)\right).$$

Then, inequality

$$-\frac{ms}{k^{\gamma+1}} \varepsilon_2 + \frac{s+2}{k^{\beta+1}} \varepsilon_4 + O\left(\frac{1}{k^{\gamma+2}}\right) + O\left(\frac{1}{k^2}\right) > -\frac{\varepsilon_4 \beta}{k^{\beta+1}} + O\left(\frac{1}{k^{\beta+2}}\right)$$

holds if either $\gamma > \beta$ (this case is excluded), or if $\gamma = \beta < 1$ and $\varepsilon_2 < \varepsilon_4(\gamma + s + 2)/ms$. Summing up all the conditions, we get (except for $m > 0$, $s > 0$) the system of inequalities

$$0 < \gamma = \beta < 1, \ \varepsilon_3 < \varepsilon_1 \frac{\gamma + s + 1}{s+1}, \ \varepsilon_4 < \varepsilon_2 \frac{\gamma + s + 1}{s+1}, \ \varepsilon_1 < \varepsilon_3 \frac{\gamma + s + 2}{ms}, \ \varepsilon_2 < \varepsilon_4 \frac{\gamma + s + 2}{ms}. \quad (12)$$

Then, the necessary conditions for the solvability of this system are expressed by the below inequalities

$$0 < \gamma = \beta < 1, \ 1 < \frac{(\gamma + s + 1)(\gamma + s + 2)}{ms(s+1)}. \quad (13)$$

As $ms(s+1) > 0$, it is easy to see that the last inequality is equivalent to the following one

$$\gamma^2 + \gamma(2s+3) + (s+1)(s+2 - ms) > 0. \quad (14)$$

Consider the equation

$$\gamma^2 + \gamma(2s+3) + (s+1)(s+2 - ms) = 0. \quad (15)$$

Its discriminant D is positive since $D = (2s+3)^2 - 4(s+1)(s+2-ms) = 4ms(s+1) + 1 > 0$ and the roots $\gamma_\pm$ of (15) are

$$\gamma_\pm = \frac{1}{2}\left[-(2s+3) \pm \sqrt{4ms(s+1)+1}\right].$$

Obviously $\gamma_- < 0$ and

$$\gamma_+ < 1 \implies \sqrt{4ms(s+1)+1} < 2s+5 \implies m < \frac{(s+2)(s+3)}{s(s+1)}.$$

The last inequality guarantees the existence of a $\gamma \in (0,1)$ such that inequalities (13) hold. Moreover, if a $\gamma \in (0,1)$ is fixed, then it is easy to show that there exist an $\varepsilon_i > 0$, $i = 1, \ldots, 4$ such that the system of inequalities (12) is satisfied. Then, by the above-mentioned results, system (6), (7) has a solution $(k, Y_1(k), Y_2(k))$, $k \in \mathbb{N}(k_0)$, where k_0 is sufficiently large, such that $-\varepsilon_1 k^{-\gamma} \leq Y_0(k) \leq \varepsilon_2 k^{-\gamma}$, $-\varepsilon_3 k^{-\gamma} \leq Y_1(k) \leq \varepsilon_4 k^{-\gamma}$. Formula (2) follows from (5). $\qquad \square$

3 COROLLARY TO MAIN RESULT

In the following corollary, values of α and m are specified such that Theorem 1 holds.

Corollary 1. *If $s > 0$ and either*
i) $0 < m < 1$ and $\alpha < -2$
or
ii) $m > 1$ and

$$-2 < \alpha < \frac{1}{2}\left[-(m-1) + \sqrt{(m-1)^2 + 16m}\right],$$

then the conclusion of Theorem 1 holds.

Proof. Assumption (4) of Theorem 1 holds if

$$ms(s+1) < s^2 + 5s + 6.$$

According to formula (3), this inequality will be valid if

$$m(\alpha+2)(\alpha+m+1) < (\alpha+2)^2 + 5(\alpha+2)(m-1) + 6(m-1)^2.$$

This inequality, after some computations, turns into

$$(m-1)\left[\alpha^2 + \alpha(m-1) - 4m\right] < 0. \tag{16}$$

As $m > 0$, inequality (16) will hold if either

$$0 < m < 1, \quad \alpha^2 + \alpha(m-1) - 4m > 0 \tag{17}$$

or

$$m > 1, \quad \alpha^2 + \alpha(m-1) - 4m < 0. \tag{18}$$

First, analyze the inequality (17). Since $s > 0$, (3) implies $\alpha < -2$ and

$$\alpha^2 + \alpha(m-1) - 4m = (\alpha - m)(\alpha + 2m) + 2(m-1)^2 - (2+\alpha) > 0.$$

That is, in the case *i*), inequality (16) holds and, consequently, assumption (4) of Theorem 1 is valid.

Next, analyze the inequality (18). In this case, $m > 1$ and (3) implies $\alpha > -2$. Considering the equation

$$\alpha^2 + \alpha(m-1) - 4m = 0$$

we find its roots

$$\alpha_{\pm} = \frac{1}{2}\left[-(m-1) \pm \sqrt{(m-1)^2 + 16m}\right].$$

Let us show that $\alpha_- < -2$. Obviously, $(m-1)^2 + 16m > 16$ and

$$-\sqrt{(m-1)^2 + 16m} < -4.$$

Therefore,

$$\alpha_- = \frac{1}{2}\left[-(m-1) - \sqrt{(m-1)^2 + 16m}\right] < -2.$$

Now, show that $\alpha_+ > -2$. This inequality is equivalent with

$$\sqrt{(m-1)^2 + 16m} > (m-1) - 4,$$

which holds for $1 < m \leq 5$ and, taking the second power, for $m > 5$, is equivalent with

$$(m-1)^2 + 16m > (m-1)^2 - 8(m-1) + 16,$$

which obviously holds as well. We conclude that the second inequality in (18) holds if $-2 < \alpha < \alpha_+$. Thus, the case *ii*) holds as well. $\square$

4 EXAMPLE

Let $m = 2$, $\alpha = 1$. Then, equation (1) takes the form

$$\Delta^2 u(k) \pm k u^2(k) = 0. \tag{19}$$

By (3), we obtain $s = 3$, $a = \mp 12$, $b = \pm 80$. Condition (4) is satisfied since

$$m < (s+2)(s+3)/s(s+1) = 2.5$$

and Theorem 1 is applicable. Therefore, by formula (2), there exist two solutions to equation (1) with the asymptotic behaviour

$$u(k) = \mp 12 k^{-3} \pm 80 k^{-4} + O(k^{-(4+\gamma)})$$

when $k \to \infty$ where $\gamma \in (0, 1)$ is a fixed number.

5 CONCLUSION

The results obtained generalize those published in [2] where only one case of equation (19) with the term $-ku^2(k)$ was considered. The method used seems to be efficient for further investigation of the behaviour of solutions to Emden-Fowler types of discrete equations. For example, it can be expected that new results will be achieved if the constants α and m satisfy sets of assumptions different from those described by Theorem 1. Another challenge for further investigation is the following one. Using a discretization suggested by the forward difference formula

$$\Delta f(k) = \frac{f(k+h) - f(k)}{h},$$

where f is a well-defined function and h is a step of discretization, the Emden-Fowler differential second-order equation can be transformed into a difference one. Analyze the asymptotic behaviour of the solutions to a derived difference equation and show that, for $h \to 0$, formulas describing the asymptotic behaviour give formulas known for differential Emden-Fowler equation.

REFERENCES

[1] Diblík, J.: Discrete retract principle for systems of discrete equations. Comput. Math. Appl., 42, 2001, p. 515–528 .

[2] Korobko, E.: On solutions of a discrete equation of Emden-Fowler type. The Student conference EEICT 2020, Faculty of Electrical Engineering and Communication, Brno University of Technology, 2020, p. 441–445.

[3] Diblík, J., Korobko, E.: Solutions of perturbed second-order discrete Emden-Fowler type equation with power asymptotics of solutions. Mathematics, Information Technologies and Applied Sciences 2020, post-conference proceedings of extended versions of selected papers, Brno, 2020. p. 30–44, ISBN 978-80-7582-366-3.

[4] Bellman, R.: Stability theory in differential equations. Dover Publications, Inc., New York, 2008, 176 pp.

THE CONCENTRATION OF A NEAR NON-RADIATIVE MAGNETIC FIELD BY CHANGING THE ANGLE OF THE SPIRAL COILS ON THE SUBSTRATE

Josef Pokorný

Doctoral Degree Programme (3), FEEC BUT

E-mail: xpokor61@vutbr.cz

Supervised by: Petr Marcoň

E-mail: marcon@feec.vutbr.cz

Abstract: This paper deals with the concentration of a near non-radiative magnetic field using two coils excited with opposite polarity and offset to each other. This allows variability in the location of the receiving coil of the charging element off-center of the transmitting coil. The magnetic induction radiation was simulated in Maxwell using the eddy current simulation. Therefore, it is only an inductive power transfer (IPT).

Keywords: Wireless power transfer, Substrate spiral coil, inductive power transfer

1 INTRODUCTION

Thanks to wireless power transmission, today's small electronics and sensors are experiencing a great boom. A big advantage of the device is the possibility to seal hermetically, and thus the possibility to meet the requirements for dust tightness and resistance to fluid pressure and other adverse environmental conditions. The disadvantage of the complexity of the wireless transmission topology is compensated by the integration of the transmitter and receiver on the chip [1]. With the support of several discrete components, the wireless charging chips can act as an inverter on the transmitting side along with the communication protocol. Thus, even on the receiving side, it ensures stable resonance, rectification and setting of the output voltage by means of a converter. As a result, the device can also be embedded in various types of material, which must not absorb a magnetic non-radiation close field.

Serial topologies are used for power transfers in units of watts. These topologies achieve the efficiency of the whole system ($\eta \leq 0.7$). Hybrid topologies composed of a combination of capacitor and coil connections are used to transmit higher powers, especially in automobiles. These topologies achieve high efficiency ($\eta \leq 0.95$) even with a large misalignment of the receiving and transmitting coils [2, 3]. Frequency switching solutions or duty cycle control adjustments improve the magnetic coupling coefficient problem but are still limited to a small value [4]. Recently, multi-coil WPT systems have become increasingly popular, especially in applications with free placement features. Key benefits of a multi-coil system include improvements in working area and system efficiency compared to conventional dual-coil systems [5]. It is stated that by means of identical coils they extend the possibility of the receiving coil deflection. In [8] they propose a multicoil induction line field based on resonance. In contrast, in [9] they solve a similar principle of supply in the opposite direction in order to orient the radiated magnetic field.

In order to concentrate the magnetic field, where the receiving coil is located, I present this article with helical coils designed on a common FR-4 PCB substrate. The proposed method makes it possible to improve the concentration of the magnetic field and to improve the range and efficiency of wireless WPT transmission.

2   PROPOSED DESIGN AND CONFIGURATION

Due to the involvement of the opposite polarity of the coils, a higher concentration of the magnetic field is expected at the point where the threads are rotated by the angle. The proposed coils are designed with the same inductance. Therefore, the magnetic field of the coils is evenly affected by the angle of rotation. The inner coil L_{in} is inside the outer coil L_{out}. The inner coil is placed on the pivot, so it can change the angle α and the direction s with respect to the static coil L_{out}. Both coils are excited from the same AC voltage source with opposite winding polarity (Figure 1).
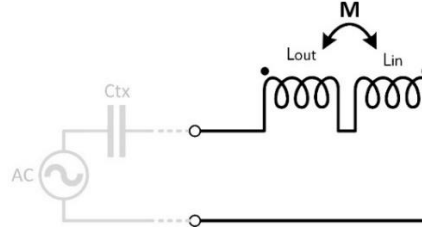

Figure 1: Connection of coils with opposite polarity of the winding.

$$L_V = L_{out} + L_{in} - 2M \text{ (H)} \tag{1}$$

In Equation 1, a minus sign applies if the windings are connected to each other (the end of the coil L_{out} with the end of the coil L_{in}, the terminals are the beginning of the coil winding L_{out} and the beginning of the coil L_{in}). where L_v is the resulting inductance, L_{out} is the outer coil, L_{in} is the inner coil and M is the mutual inductance.

$$L = K_1 \mu_0 \frac{N^2 D_{av}}{1 + K_2 \varphi} \text{ (nH)} \tag{2}$$

The calculation of the inductance of each of the coils is based on the empirical equation 2 according to [6]. Where the coefficient K_1 and K_2 depend on the type of the layout. For octagonal layout, $K_1 = 2.25$ and $K_2 = 3.35$. [7]. N is the number of turns, μ - magnetic constant, D_{av} - average diameter of coil calculed from equation 3 from parameter on figure 3:

$$D_{av} = \frac{D_o + D_i}{2} \text{ (µm)} \tag{3}$$

$$\varphi = \frac{D_o - D_i}{D_o + D_i} \text{ (-)} \tag{4}$$

Using eq. 4, we get the fill factor needed to calculate the total induction in eq. 2.
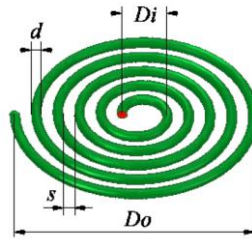
Figure 2: Parameters for calculating inductance of coil.

Table 1 compares the dimensions of the outer and inner coils for shaping the magnetic field.

Table 1: Dimensions of L_{in} and L_{out} on substrate FR-4.

	Inductance	Number of turns (N)	Inside diameter (Di)	Outside diameter (Do)	Space (s)
Inner coil (L_{in})	4 µH	21	5 mm	19 mm	0.15 mm
Outer coil (L_{out})	4 µH	10	23 mm	29.3 mm	0.15 mm
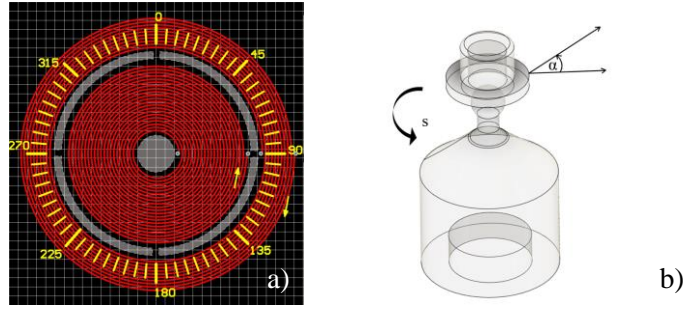
Figure 3: a) Layout of coils in Altium designer (grid 1 mm). b) Pivot with cilinder for inner coil L_{in}.

Figure 3a shows the design of a coil with 360° marking for possible measurements with multiple pairs of coils used in the matrix. The perforations are within the production limit of 0.2 mm and are sufficient for supply with enamelled wire. The design is adapted to the production requirements of the manufacturer. In Figure 3b, the support pivot for the L_{in} coil is modelled and printed using a 3D printer. Thus, the coil can be offset in all directions relative to the coil L_{out} and at the same time rotated by an angle α.

3   RESULTS

The simulation was performed in ANSYS solver, which uses the finite element method (FEM). Using harmonic analysis called magnetic: eddy current. Simulation was set with a current of 1 A and a frequency of 250 kHz with the opposite direction of excitation in L_{in} according to the connection in Figure 1. Solving was setup to maximum number of passes 5 and percent error 0.2. Convergence standard was set as: refinement per pass 10% and minimum number of passes on 2. Figure 4a shows, as expected, a uniform radiated magnetic field from the flat helical coils in the z-axis.

a) $\alpha= 0°$ b) $\alpha= 15°$

c) $\alpha= 45°$ d) $\alpha= 90°$

Figure 4: Dependence of magnetical inductance in z-axis in air space on angle α on inner coil $L_{in.}$

For Figures 4b and 4c, the maximum concentration of the magnetic field perpendicular to the turns of the coil L_{in} can be seen. Figure 4d shows that when $\alpha = 90\,°$ the magnetic field of the coils will have sharp maxima in x, y, z axes, which is undesirable and visible from the graph in Figure 5.
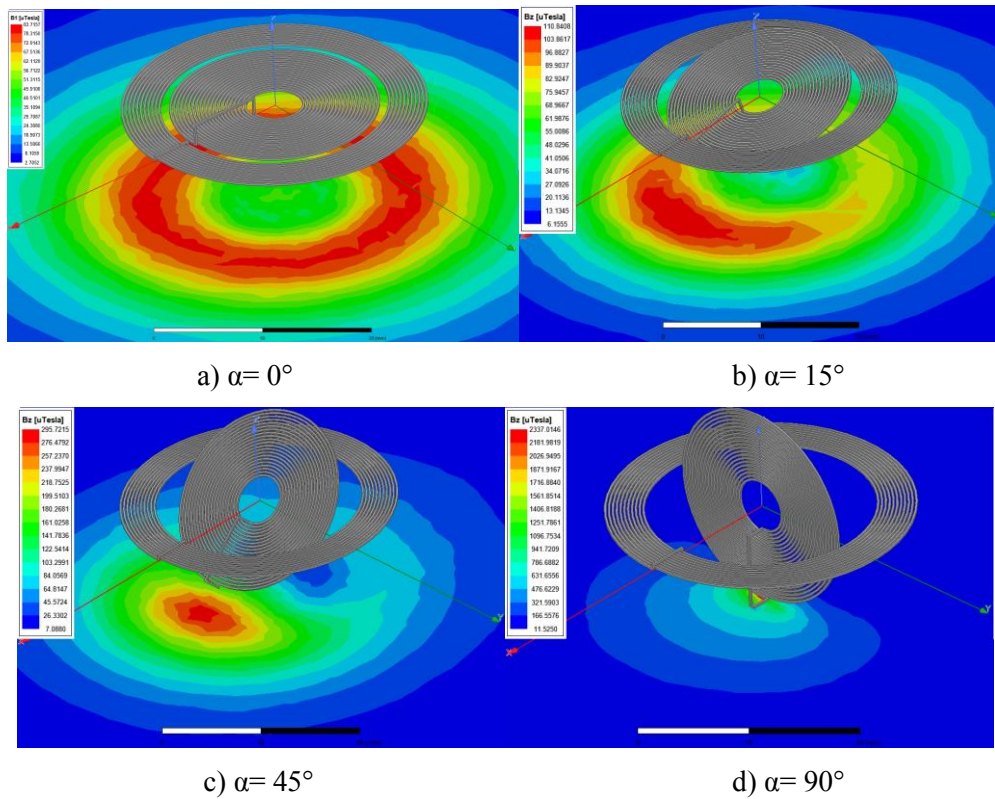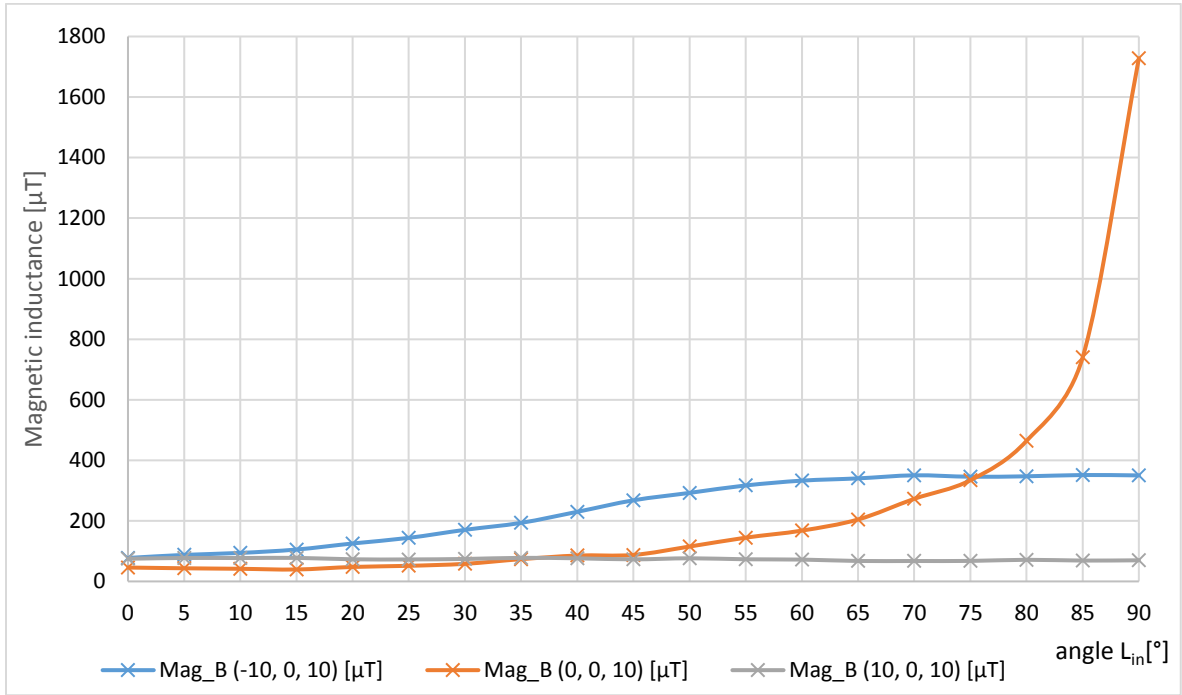

Figure 5: Depence magnetic induductance on angle of inner coil in 3 points in z-axis

Here, the magnetic induction is measured at three points in Euclidean space, which are named by using coordinates in millimeters. It can be seen from the graph, that at the point Mag_B (0, 0, 10), the magnetic induction increases exponentially and reaches a maximum, when the point touches almost the edge of the coil L_{in}. The graph shows, that the coil L_{in} should be rotated to an angle $\alpha = 60\,°$, when the magnetic induction reaches 332 µT at the point Mag_B (-10, 0, 10) while the other two points Mag_B (0, 0, 10) and Mag_B (10, 0, 10) of the magnetic field are kept at the half of the magnetic induction value of the Mag_B (-10, 0, 10).

4   DISCUSSION

In this article, the misalignment and rotation of one coil winding is used, which results in the concentration of the magnetic near non-radiation field. By constructing a matrix of this coil type, precise shaping of the magnetic field, can be achieved for different sizes and positions of the receiving coil.

5   CONCLUSION

This article describes a new arrangement of two coils on a FR-4 substrate, whereby the inner coil can be offset relative to the other in all directions thanks to the pivot. In addition, the inner coil is fed in the opposite direction. The simulation results show, that when the inner coil is rotated by 35 °, the magnetic induction concentration is doubled at the focus point. At the same time, it is effective not to exceed an angle of 60 °, when it begins to exceed the radiation in the x and y axes.

REFERENCES

[1]   Renesas Introduces World's First 60 W Wireless Power Receiver IC [online]. https://www.everythingrf.com/News/details/11513-Renesas-Introduces-World-s-First-60-W-Wireless-Power-Receiver-IC

[2]   Madzharov, N., & Hinov, N. (2019). Flexibility of Wireless Power Transfer Charging Station Using Dynamic Matching and Power Supply with Energy Dosing. *Applied Sciences*, *9*(22). https://doi.org/10.3390/app9224767

[3]   Mohammad, M., Choi, S., Islam, Z., Kwak, S., & Baek, J. (2017). Core Design and Optimization for Better Misalignment Tolerance and Higher Range of Wireless Charging of PHEV. *IEEE Transactions on Transportation Electrification*, *3*(2), 445-453. https://doi.org/10.1109/TTE.2017.2663662

[4]   Moghaddami, M., Sundararajan, A., & Sarwat, A. I. (2018). A Power-Frequency Controller With Resonance Frequency Tracking Capability for Inductive Power Transfer Systems. *IEEE Transactions on Industry Applications*, *54*(2), 1773-1783. https://doi.org/10.1109/TIA.2017.2779425

[5]   Sritongon, C., Wisestherrakul, P., Hansupho, N., Nutwong, S., Sangswang, A., Naetiladdanon, S., & Mujjalinvimut, E. (2018). Novel IPT Multi-Transmitter Coils with Increase Misalignment Tolerance and System Efficiency. In *2018 IEEE International Symposium on Circuits and Systems (ISCAS)* (pp. 1-5). IEEE. https://doi.org/10.1109/ISCAS.2018.8351305

[6]   Wheeler, H. A. (1928). Simple Inductance Formulas for Radio Coils. *Proceedings of the IRE*, *16*(10), 1398-1400. https://doi.org/10.1109/JRPROC.1928.221309

[7]   Mohan, S. S., del Mar Hershenson, M., Boyd, S. P., & Lee, T. H. Simple accurate expressions for planar spiral inductances. *IEEE Journal of Solid-State Circuits*, *34*(10), 1419-1424. https://doi.org/10.1109/4.792620

[8]   Mirbozorgi, S. A., Bahrami, H., Sawan, M., & Gosselin, B. (2014). A Smart Multicoil Inductively Coupled Array for Wireless Power Transmission. *IEEE Transactions on Industrial Electronics*, *61*(11), 6061-6070. https://doi.org/10.1109/TIE.2014.2308138

[9]   Kallel, B., Kanoun, O., & Trabelsi, H. (2016). Large air gap misalignment tolerable multi-coil inductive power transfer for wireless sensors. *IET Power Electronics*, *9*(8), 1768-1774. https://doi.org/10.1049/iet-pel.2015.0800

FROM MAGNETIC RESONANCE TO THE LOW EARTH ORBIT: SIMULATION AIDED DESIGN OF BIRDCAGE ANTENNA WITH APPLICATION IN THE ATMOSPHERE-BREATHING ELECTRIC PROPULSION ENGINE

Karel Juřík

Doctoral Degree Programme (1), FEEC BUT

E-mail: xjurik14@stud.feec.vutbr.cz

Supervised by: Dušan Nešpor

E-mail: nespord@feec.vutbr.cz

Abstract: This paper deals with the design and validation of the birdcage coil and with the proposal of its implementation into the Atmosphere-Breathing Electric Propulsion Engine (ABEP). The coil was designed, analysed, and tuned using ANSYS HFSS software and the results were evaluated against the measurement of the prepared prototype. The difference between the measured and required frequency was approximately 2.7 %. Although the coil needs to be further optimized, the ability to achieve the desired frequency was demonstrated.

Keywords: MRI, birdcage, RF coil, ABEP, electric propulsion, numerical analysis

1   INTRODUCTION

Utilization of Very Low Earth Orbits (VLEO) for satellite operations would bring a wide range of benefits. The observation of Earth through the reduced thickness of the atmosphere leads to improved resolution, signal-to-noise ratio (SNR), and is enabling novel technologies and approaches. Moreover, thanks to the constant drag of the atmosphere, the satellites cannot become space debris, they will quickly fall after the end of their lifetime. [1]

However, the presence of the atmosphere brings several challenges as well. The constant friction makes a demand for a source of a continuous thrust, which will compensate it. If the standard concept of propulsion engines was used, there would be a need for an enormous amount of propellant, or the lifetime of such a device would be significantly reduced. This problem is addressed by the utilization of atmospheric particles. Then, there is an infinite source of the propellant, which only must be properly ionized and accelerated to produce the thrust.

The mentioned conditions pose strict requirements on such an engine. At VLEO, there is a low density of atmospheric particles, moreover, they are composed mainly of oxygen and nitrogen. While the first condition requests high ionization efficiency, the later one requires the corrosion resistance of the whole system, even to the oxygen ions and radicals. In the ideal case, the electrodes should be placed outside the discharge channel to prolong the operation time. [1][2]

Among many attempts to address these challenges, the solution proposed by Romano et al. is one of the most promising. Their approach consists of employing the birdcage coil for ionization of the atmosphere particles and their acceleration by the variable electromagnetic field of the birdcage coil, in addition to an external static magnetic field. The frequency 40.68 MHz for driving the birdcage was used. [3]

The combination of the variable electromagnetic field, together with the static magnetic field leads to the formation of helicon waves and acceleration of both ions and electrons together. [3],[4]

2   BIRDCAGE COIL THEORY

Birdcage coils generally consist of two end rings, connected by equally spaced legs and capacitors. There are a few different ways of birdcage coil construction. The two most used are low-pass resonator, where the capacitors are placed in the middle of the coil legs, and high-pass design (shown in **Figure 1**), where the capacitors are placed in the end rings, between the legs.

The concept of Birdcage coils was originally developed for use in Magnetic Resonance Imaging (MRI). The main advantages are the great signal-to-noise ratio and excellent rf field homogeneity in the whole volume of the coil. Since the resonant frequency depends on several mutually independent variables (coil diameter, length and width of the legs, capacitor values, etc.), the frequency can be broadly tuned to different values for a single set of its dimensions. [5]

Every birdcage coil has N/2 resonance modes (where N is the number of its legs), for high-pass design, there is one extra, so-called antiresonance (AR) mode. Each mode has a different distribution of electromagnetic field inside the coil, only the first resonance mode provides the homogeneous magnetic field. For low-pass design, the first mode corresponds to the lowest frequency, while for the high-pass coils, the frequency of the first mode is the second highest (after the AR peak). [3],[5]
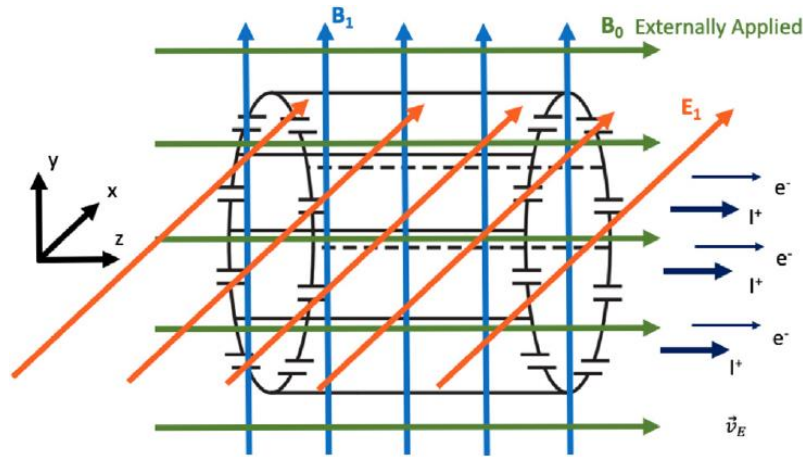

Figure 1: A high-pass birdcage coil with its electric (E_1) and magnetic (B_1) fields. The static magnetic field B_0 is externally applied to accelerate the charged particles. Adapted from [3].

3   EXPERIMENTAL

We decide to base our ABEP system on the widely used frequency 13.56 MHz, due to the availability of signal generators, amplifiers, and other high-frequency parts.

The Birdcage Builder software [6] was used for the first estimation of the coil parameters. The coil diameter, as well as the width of the endrings and the legs, were fixed to the desired values (which were derived from the diameter of the ABEP discharge channel and the available copper tape width respectively). Other parameters, the type (High-pass vs. Low-pass), coil length, and the value of the capacitors were roughly tuned to meet the desired frequency of 13.56 MHz. The first concept consisted of an eight-leg high-pass coil with a diameter of 5 cm and length 6.2 cm, with the legs and endrings 1.2 cm wide, and the capacity of the capacitors 6.8 nF.

These values were subsequently used for building the model in ANSYS HFSS software. The capacitors were modelled using Lumped RLC boundary condition. Since there are only a few discrete capacity values of the capacitors widely available on the market, the capacity was fixed to 6.8 nF and the only remaining degree of freedom consists of the length of the coil. Using finite element method (FEM) analysis, the accurate value of the length was found. The length was corresponding to the first resonance peak at the frequency 13.56 MHz, and it was determined to be 8.75 cm (see the difference compared to the results of the Birdcage Builder app).

To confirm the model's fidelity, the birdcage coil with calculated parameters was built. Polyvinylchloride (PVC) tube was used as a supporting structure. The tube was covered with the self-adhesive copper tape, following the layout of the model (see **Figure 2**). The endrings were thoroughly soldered to the individual legs to ensure good conductivity even at high frequencies. In the endrings gaps, the SMD capacitors in the 1206 housing of capacity 6.8 nF were used.

Finally, the coil was equipped with a BNC connector and its parameters were measured using the vector network analyser Rohde & Schwarz ZVL 9 kHz – 6 GHz.
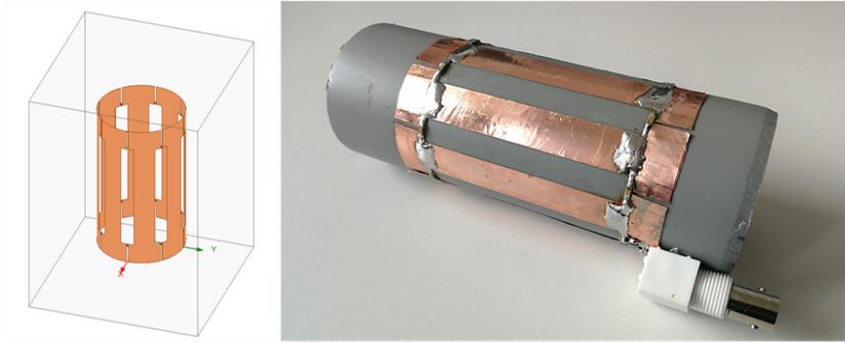
Figure 2: Left: the model of the coil created in ANSYS HFSS; right: prototype of the coil.

4   RESULTS AND DISCUSSION

Figure 3 shows the results of the numerical analysis of the coil and the values obtained from the real measurement of the prepared prototype. In both spectra, there are clearly visible the first resonance peak at a frequency of around 13.5 MHz, the second peak located around 11.5 MHz, and the anti-resonant peak at the frequency around 20.5 MHz. The other peaks are clearly visible in the simulation only. In the measured data, the two highest peaks are broadened and overlapped.

Using the FEM, the frequency of the first peak was determined at 13.56 MHz, while the measured data shows the peak located at 13.93 MHz. The difference is 2.73 %. The major deviation between the numerical analysis results and the measured values lies in the return loss at the resonance frequencies. The difference can be explained as the capacitors were considered ideal in the simulation.

Moreover, the impedance of the coil was not matched to 50 Ω. The impedance matching, as well as the fine-tuning of the frequency, is the next task in the development of the high-frequency part of the propulsion engine. The task will not be trivial, because of the need for an auto-matching network since the plasma ignition will strongly alter the impedance of the whole system.
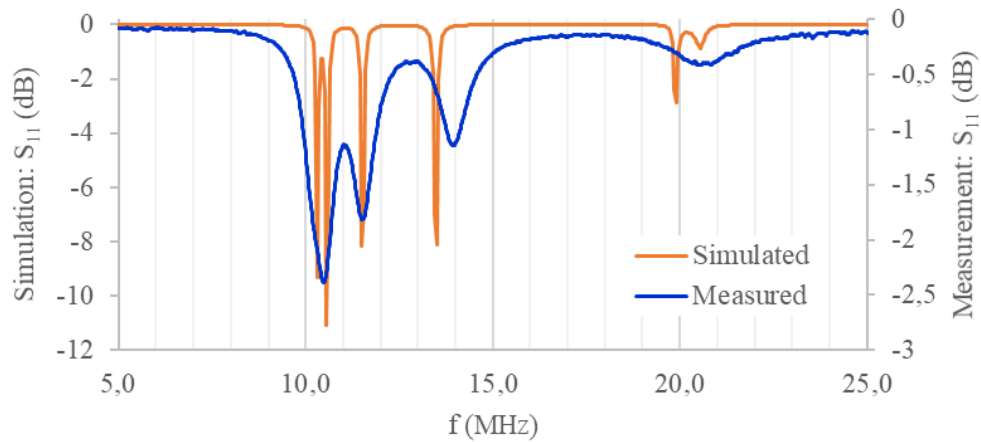
Figure 3: Simulated and Measured S_{11} parameters of the Birdcage coil.

To support our assumption, the peak at 13.56 MHz corresponds to the right resonance mode, the magnetic field magnitude and direction were plotted for this frequency. The results, shown in **Figure 4** were plotted perpendicularly to the coil axis, in the middle of the coil length. There is clearly visible uniformity of the magnetic field inside the volume of the coil.
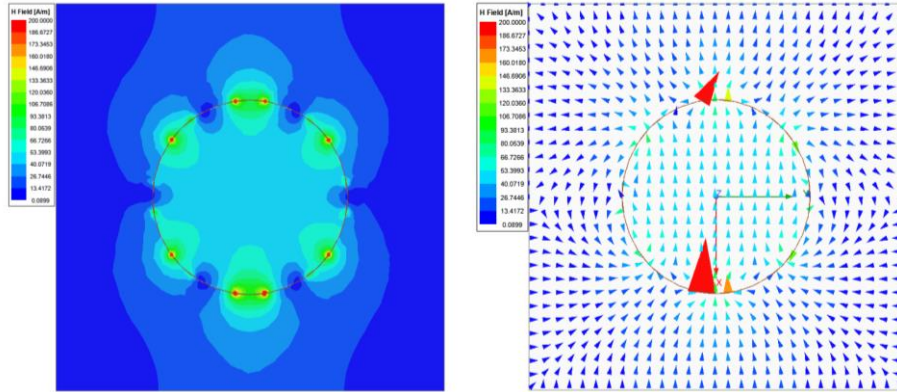
Figure 4: First resonance mode; left: the magnitude of the magnetic field; right: Magnetic field direction. Top view (XY plane). Plots were obtained using ANSYS HFSS.

To confirm the findings, the magnetic field amplitudes were also plotted for frequencies corresponding to other resonance modes. As it is clearly visible in **Figure 5**, the other resonance modes, located at a) 10.33 MHz, b) 10.57 MHz and c) 11.52 MHz for the higher resonance modes and at d) 20.02 MHz for the anti-resonant mode, do not provide homogeneous magnetic field across the coil.
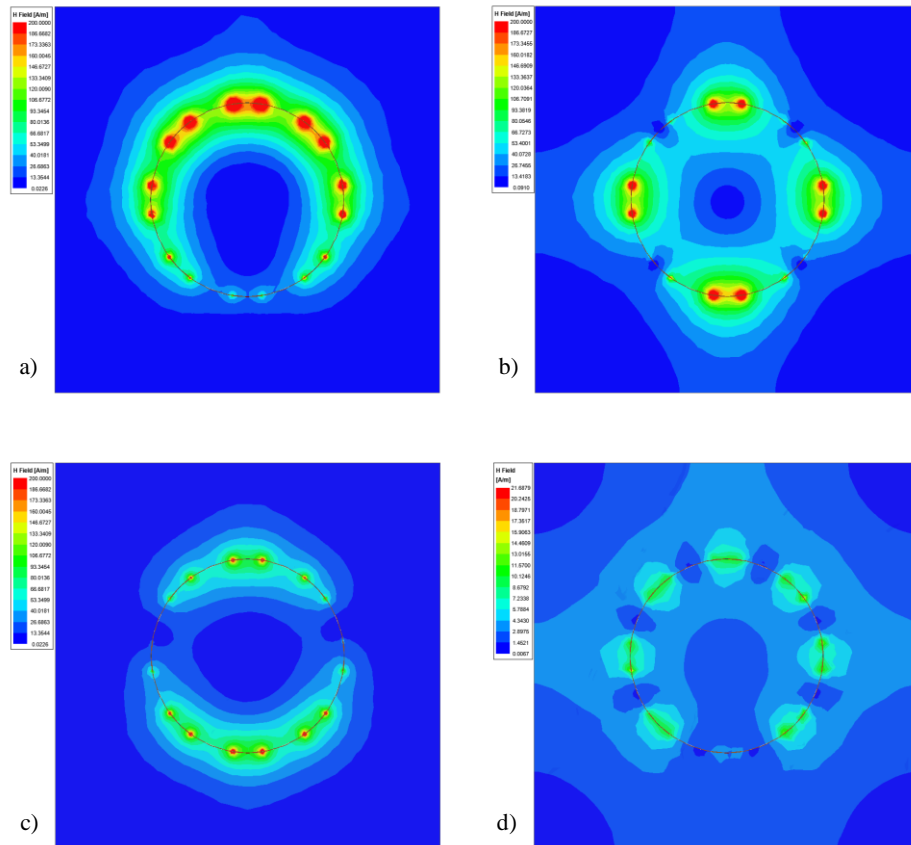
Figure 5: Distribution of magnetic field magnitude for a) 4th, b) 3rd, c) 2nd resonance and d) anti-resonance modes. Top view (XY plane). Plots were obtained using ANSYS HFSS.

5 CONCLUSION AND OUTLOOK

To sum up, within this work the Birdcage coil operating at 13.56 MHz was designed and created. Its parameters were measured and compared with the results of the numerical analysis results. The first resonance mode lies at the frequency of 13.93 MHz, which makes the difference 2.73 % compared to the results of the analysis and the desired frequency. The birdcage coil will be subsequently tuned, matched, and employed in the atmosphere-breathing electric propulsion engine. This may require an automatic matching network since the impedance of the low-pressure gas and plasma case might differ strongly.

REFERENCES

[1]   CRISP, N.H., P.C.E. ROBERTS, S. LIVADIOTTI, et al. The benefits of very low earth orbit for earth observation missions. Progress in Aerospace Sciences. 2020, 117, ISSN 03760421. Available at: doi:10.1016/j.paerosci.2020.100619

[2]   JACKSON, Stephen W. a Robert MARSHALL. Conceptual Design of an Air-Breathing Electric Thruster for CubeSat Applications. Journal of Spacecraft and Rockets. 2018, 55(3), 632-639. ISSN 0022-4650. Available at: doi:10.2514/1.A33993

[3]   ROMANO, F., Y.-A. CHAN, G. HERDRICH, et al. RF Helicon-based Inductive Plasma Thruster (IPT) Design for an Atmosphere-Breathing Electric Propulsion system (ABEP). Acta Astronautica. 2020, 176, 476-483. ISSN 00945765. Available at: doi:10.1016/j.actaastro.2020.07.008

[4]   GUITTIENNE, Ph., E. CHEVALIER a Ch. HOLLENSTEIN. Towards an optimal antenna for helicon waves excitation. Journal of Applied Physics. 2005, 98(8). ISSN 0021-8979. Available at: doi:10.1063/1.2081107

[5]   HAYES, Cecil E, William A EDELSTEIN, John F SCHENCK, Otward M MUELLER a Matthew EASH. An efficient, highly homogeneous radiofrequency coil for whole-body NMR imaging at 1.5 T. Journal of Magnetic Resonance (1969). 1985, 63(3), 622-628. ISSN 00222364. Available at: doi:10.1016/0022-2364(85)90257-4

[6]   CHIN, Chih-Liang, Christopher M. COLLINS, Shizhe LI, Bernard J. DARDZINSKI a Michael B. SMITH. BirdcageBuilder: Design of specified-geometry birdcage coils with desired current pattern and resonant frequency. Concepts in Magnetic Resonance. 2002, 15(2), 156-163. ISSN 1043-7347. Available at: doi:10.1002/cmr.10030

THE APPLICATION OF TEMPO CALCULATION FOR MUSICOLOGICAL PURPOSES

Matej Istvanek

Doctoral Degree Programme (2), FEEC BUT

E-mail: xistva02@stud.feec.vutbr.cz

Supervised by: Zdenek Smekal

E-mail: smekal@feec.vutbr.cz

Abstract: Beat tracking systems capture time positions of beats within digital recordings. They are used, for example, in streaming portals, but applications in the musicological analysis are often neglected. In this article, two different methods of beat tracking systems are tested—conventional and the state-of-the-art—on the specific motif of a string quartet music, which is one of the most complex tasks for beat detectors in general. The aim here is to determine which system is better for musicology purposes. This often involves determining not only the position of individual beats and estimating the tempo but also the accuracy of determining their number. Evaluation analysis may be suitable for comparing the accuracy of detectors, but may not necessarily reflect the requirements of musicological analysis. The results of selected detectors show that a system based on a recurrent neural network seems to be the most suitable.

Keywords: beat tracking, musicology, music analysis, music information retrieval, tempo estimation

1 INTRODUCTION

Beat tracking and rhythmic analysis are one of the key and most developed problems in the field of Music Information Retrieval (MIR). Algorithms determine the rhythmic or metric structure of a digital recording by specifically manipulating the audio signal and extracting valid information from it. Previously, the problem was grasped in various ways—the most successful algorithms were based mainly on the calculation of periodicity and the distribution of onsets over time. With the development and availability of artificial neural networks, virtually all MIR topics have been transformed and state-of-the-art algorithms have been replaced. Musicological analysis can deal, among other things, with the comparison of different interpretations or performances of the same composition. Here is the advantage of the MIR field, which provides the possibility of extracting the required information by machine, i.e. in much larger quantities, unified, faster and perhaps even more objectively. The well-known cases of cooperation between MIR researchers and musicologists include [1] and [2]. In addition to the UK, Austria and Germany, for example, also participate in the cooperation [3]. Although these are two theoretically very different fields, they have many challenges in common. Estimation of rhythmic structure and tempo by automatic methods provides the possibility of complex musical analysis. But are conventional detection systems really suitable for musicological analysis?

2 METHODS

The analysis consists of comparing and possibly modifying several beat tracking methods in an application for musicological analysis. This is specific for the type of music on which detection systems are usually not trained and tested and also for their focus on the same recordings, but different interpretations. As part of the development of new methods and the improvement of existing systems, for example at the MIREX competition [4], it is very rare to see testing exclusively on classical music

or the same composition using different interpretations. The most commonly used datasets contain a large number of recordings of various genres and musical instruments. The system may then be able to generalize better, but its specialized application may fall behind.

2.1 BEAT TRACKING

Earlier detection systems used periodicity in the onset strength envelope by modifying the spectral difference or spectral flux. An example of an advanced system is the beat tracking included in the librosa module [5] for the Python language. Figure 1 describes the signal flow in a conventional system and a neural network system. The left part of the image corresponds to librosa processing. The system is hereinafter referred to as lib. The problem with the detector is generally the inability to adapt to a rapidly changing tempo. Leap changes of tempo and meter are not expected. In the basic setting, an average tempo of 120 beats per minute (bpm) is assumed, from which the adjustable tolerance is determined. Therefore, it is a question of to what extent systems with a similar architecture can adapt to the agogic and complex rhythmic structure of string quartets.
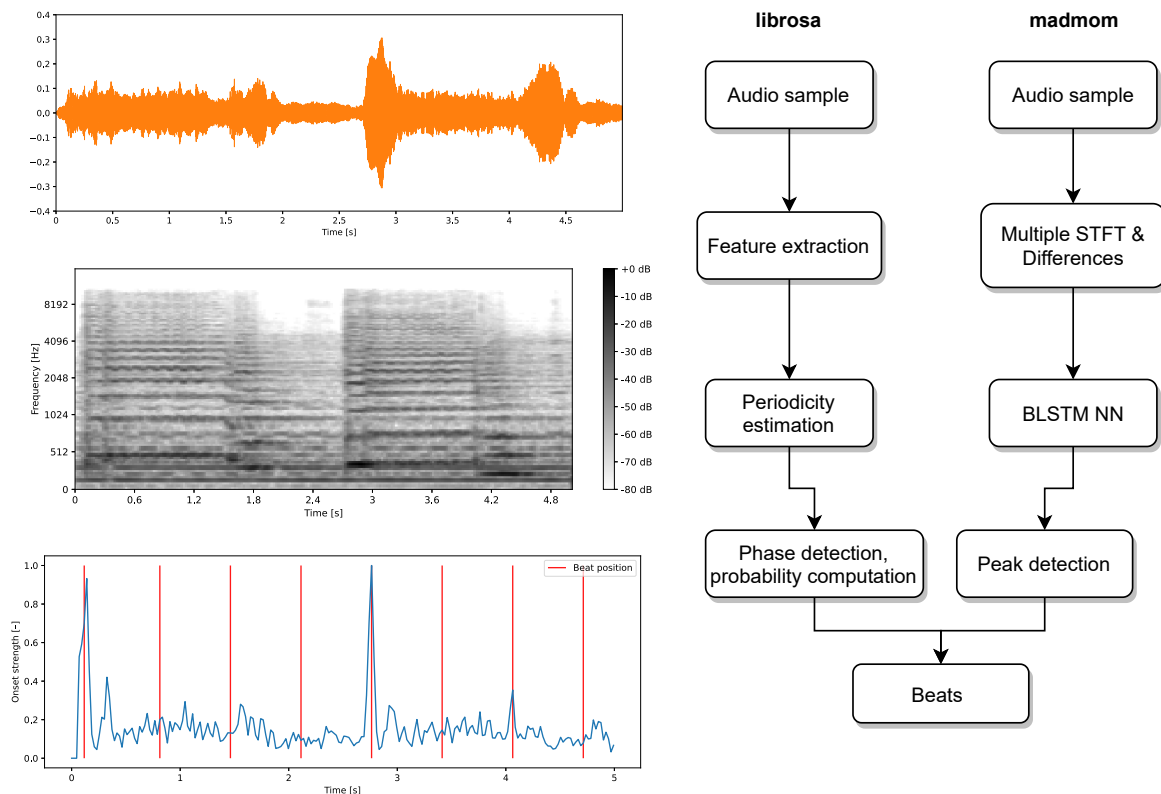
Figure 1: Signal flow of beat tracking systems [5, 6].

The second tested system is from the madmom module [6]. It uses a bidirectional recurrent neural network with Long Short-Term Memory (LSTM) cells, which are based on the determination of beat times according to a modified spectral envelope. The type of architecture is chosen deliberately here. By adding LSTM cells to the network, it is possible to store the time information that is theoretically necessary to determine the longer-term rhythmic structure, and thus the correct detection of musical beats. Finally, a dynamic Bayesian network (DBN) approximated by a Hidden Markov Model (HMM) to compute a probability of a beat within given frames (the neural network is trained on 100 frames per second resolution) is used. However, the question is how well the system can handle string quartets, as the network has not been trained for this style of music.

2.2 GLOBAL TEMPO

How can we compute the average tempo of recordings? Is it necessary to have a beat tracking system? For many applications of these algorithms (music recommendations, classification, genre recognition) within streaming portals, this is more or less the only available option. This procedure may not be necessary for musicological analysis. In a musicological analysis, we often work with one selected composition—then, the acquisition of recordings from different performers (i.e. different interpretations) takes place and the analysis is conducted. The advantage of comparing the same composition, but different musical performances, lies in the possibility of using music notation. Obtaining a notation of a given composition in the *.pdf* and *.xml* formats is typically not a problem (musicologists are usually familiar with notation software such as Sibelius or the open-source variant, MuseScore). Information about the number and absolute positions of the beat times (concerning the melodic and harmonic sequence) can be obtained from the notation. Thus, the global (average) tempo (GT) of a musical motif can simply be calculated as $GT = 60 \cdot N/d$, where N = the number of beats and d = time duration of the motif. In this way, however, we do not obtain information about the time position of individual beats, but only about the average tempo of the analyzed section.

2.3 DATASET

For the analysis, a first motif (bars 4–10) of the *String Quartet No. 1 e minor "From My Life"* by Bedrich Smetana was used. Together with prof. Spurny from Masaryk University, we have collected 33 different interpretations. This motif contains a total of 12 beats, which were manually annotated using Sonic Visualiser software to obtain ground truth data. In addition to GT, GTT (Ground Truth Tempo) is introduced. Manual ground truth annotations can be used to obtain GTT. First, the time difference d between successive beats was calculated, which was then averaged (*ad*) and the relationship $GTT = 60/ad$ was used. This value should theoretically be equal to GT, in fact, it will be slightly different, as manual annotation and also segmentation of digital recordings is not absolutely accurate. Accuracy can be determined by the difference between GT and GTT.

2.4 APPROACH

First, the motif was segmented from all available recordings of the database. GT, GTT, and their difference were calculated. Then, all recordings were analyzed by detection systems lib and madmom. The parameters were first left in the default settings (**lib**), then other parameterizations were tested:

- **lib_t** (*t* stands for tuned): the original sampling frequency $f_s = 44,100$ Hz was not changed and a different Mel spectrogram setting (138 mels), length of FFT (2048 samples), and hop factor (512 samples) was used.

- **lib_avgGT** (average Global Tempo): a specific parameter was selected that affects the expected start tempo (start_bpm). Thus, the system does not calculate the predominant tempo from the most frequently used value of 120 bpm but automatically calculates the position of the beats according to the most suitable candidate based on the selected tempo, including the adjustable tolerance (tightness).

The advantage of lib_avgGT is mainly the elimination of so-called octave tempos (120 bpm is the octave tempo of 60 bpm and so on). However, this parameter cannot be set separately in many cases, as we do not know the expected tempo. In a musicological analysis, however, we can know GT in advance, as we have recordings and sheet music available. Therefore, lib_avgGT uses the average GT of the entire tested database, 88 bpm. The last system is a detector using the previously mentioned bidirectional recurrent neural network.

3 RESULTS

The results are attached in Table 1. All the mentioned detectors were used for the musical motif, GT was calculated from the notation and GTT from the ground truth annotation. The mean, median and standard deviation were computed for all items. The absolute difference between GT and GTT is on average 0.318 bpm, which shows the high accuracy of manual segmentation and annotation of beats positions. The standard deviation, in this case, is 0.261. The lib and lib_t system show very inaccurate tempo detection, the average tempo value according to lib_avgGT is closest to GTT (the difference is 4.783 bpm, which can be considered a promising result). Madmom differs by 8.977 bpm.

Table 1: Tempo estimation of the string quartet motif.

Track	GT	GTT	Δ	lib	lib_t	lib_avgGT	madmom
CD01	100.000	100.660	0.660	117.454	123.047	95.703	99.548
CD02	93.506	93.576	0.070	78.303	143.555	86.133	72.398
CD03	88.615	88.877	0.261	234.908	90.666	89.103	88.829
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
CD33	98.630	100.660	2.030	56.174	139.675	99.384	100.457
Mean	**88.129**	**88.447**	**0.778**	**130.943**	**119.298**	**93.230**	**79.470**
Median	87.681	88.539	0.487	93.994	120.185	92.285	80.932
Std	7.592	7.853	0.751	79.510	13.101	7.717	15.721

Table 2 shows the mean and median number of beats determined for all tracks and systems. Here we can see the advantage of the madmom system. Even though the average tempo determined by lib_avgGT is closest to the reference, the median number of detected beats is 12 (11 for lib_avgGT). Besides, the number of tracks in which exactly 12 beats were detected is 25 for madmom and only 3 for lib_avgGT (out of a total of 33 recordings).

Table 2: Number of beats detected for each track from the dataset.

Track	lib	lib_t	lib_avgGT	madmom
CD01	13	14	11	12
CD02	7	18	10	9
CD03	28	12	11	12
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
CD33	5	18	11	12
Avg	15.788	16.212	**11.636**	11.152
Median	11	16	11	**12**
12 beats	1	2	3	**25**

4 DISCUSSION

Although lib_avgGT is the best system chosen according to the results of average tempo detection, the most suitable system for musicological analysis of the rhythmic structure and tempo estimation is madmom (RNN based). The accuracy of detectors is usually compared by f-score (f-measure)

metrics, however, this is not the most important factor for application in music analysis. Until the systems reach an f-score of about 0.95 and higher, a manual correction will always be needed for the analysis to be truly meaningful. The key factor here is the minimization of the researcher's time in correcting the actual time positions of the beats in the individual recordings. Of course, this article is limited to comparing only a few detectors, but librosa is considered a conventional system and madmom a state-of-the-art [4]. It is also a question of whether neural network-based systems should not be considered conventional today. Sonic Visualiser uses the BeatRoot system, which is similar to librosa but older and less accurate. Detectors based on neural networks are not accessible in this software, as they remain as testing tools in the development environment (Sonic Visualiser uses Vamp plugins), although they offer high potential for the musicological analysis. Until they are implemented in a user-friendly environment, musicologists will not use them and their application will depend on the cooperation of musicologists and researchers in the field of MIR.

5 CONCLUSION

This article deals with the application of beat tracking systems for musicological analysis. It tests selected systems on recordings of string quartets, which are generally very challenging for detectors. The test dataset of recordings is well segmented and annotated, which is also confirmed by the additional calculation of the difference between GT and GTT. Madmom seems to be the most suitable system (of the selected detection systems), as it provides relatively high accuracy in determining the average tempo and also estimated the correct number of beats in 25 cases out of 33. The detector from the librosa module, despite the preset start tempo, was able to identify precisely 12 beats in only 3 cases. It should be noted that the system missed only once (i.e. detected 11 or 13 beats) in 21 cases. From a technical point of view, musicological analysis requires minimizing the time spent on manually editing the annotation and ground truth. Accurate determination of tempo, time positions of beats, but also their number is an important factor for testing the validity of detection systems for the musicology analysis and their future use in analysis software.

REFERENCES

[1] *Mazurka Project* [online]. UK: AHRC Research, Centre for the History and Analysis of Recorded Music (CHARM), ©2009. [retrieved 15.01.2021]. Available: `http://www.mazurka.org.uk`

[2] *Machine Intelligence and Music Informatics Research Group* [online]. UK: University of London, Department of Computer Science [retrieved 15.01.2021]. Available: `http://mirg.city.ac.uk/projects`

[3] DAMM, David, FREMEREY, Christian, KURTH, Frank, MULLER, Meinard and CLAUSEN, Michael. Multimodal presentation and browsing of music. In: *ICMI'08*. NY, ACM, 2008, pp. 205–208. DOI: 10.1145/1452392.1452436

[4] EMA. *MIREX HOME* [online]. USA, Illinois: University of Illinois at Urbana-Champaign, ©2005. [retrieved 15.01.2021]. Available: `https://www.music-ir.org/mirex/wiki/MIREX_HOME`

[5] McFEE, Brian, RAFFEL, Colin, LIANG, Dawen, ELLIS, Daniel, McVICAR, BATTENBERG, Eric and NIETO, Oriol. librosa: Audio and music signal analysis in python. In: *Proceedings of the 14th python in science conference*. USA: Texas, 2015, pp. 18–25. DOI: 10.25080/Majora-7b98e3ed-003

[6] *madmom* [online]. Austria, ©2015. [retrieved 15.01.2021]. Available: `https://madmom.readthedocs.io/en/latest/index.html`

IDENTIFICATION OF SLEEP/WAKE STAGES IN ACTIGRAPHY DATA UTILISING GRADIENT BOOSTING ALGORITHM

Marek Mikulec

Doctoral Degree Programme (1st year), FEEC BUT

E-mail: xmikul56@vutbr.cz

Supervised by: Jiří Mekyska

E-mail: mekyska@feec.vutbr.cz

Abstract: Sleep disorders are early markers of various serious diseases that can be treated more effectively when diagnosed in their prodromal stage. Actigraphy is a noninvasive sleep monitoring method for the detection of sleep patterns and determination of sleep parameters that could support the diagnosis of these disorders. This study aims to compare a newly proposed actigraphy-based method of sleep/wake detection with a conventional one in terms of consistency with a polysomnography (PSG) reference. 55 recordings (acquired in 28 subjects) of actigraphy and PSG were modelled by a heuristics-based method and by a new approach utilising a gradient boosting algorithm. In addition, another database (22 subjects, 150 recordings) was used to compare scores of the new method with data reported in sleep diaries. The proposed method achieves 89% accuracy and Mathews correlation coefficient equal to 0.75 when compared to the polysomnography reference. Such results outperform the ones provided by the heuristic technique. The newly proposed method has good consistency with the PSG reference, thus being a good alternative to the golden standard in sleep disorders assessment, especially in decentralised clinical trials.

Keywords: actigraphy, machine learning, polysomnography, sleep, sleep diary

1 INTRODUCTION

Quality of sleep is an essential aspect of a healthy life. Sleep disorders are early indicators of various diseases (e.g. Parkinson's disease, dementia with Lewy bodies [1]) that can be treated much more effectively if they are diagnosed in the prodromal stage, i.e. the stage wherein early symptoms or signs of the disease are present, but a classic clinical diagnosis is not yet possible. The group of sleep disorders includes, for example, insomnia, restless legs syndrome, idiopathic rapid eye movement sleep behaviour disorder (iRBD), parasomnia, or sleep apnoe [1, 2].

Sleep monitoring via actigraphy is a noninvasive monitoring method that records the occurrence and the degree of limb movement. Actigraph is a bracelet mostly worn on the wrist, ankle, or waist. Despite the movement, some actigraph devices additionally track modalities such as temperature, blood pressure, light intensity, etc. The actigraphy data, expressed in the form of time series, are consequently processed using statistical or machine learning methods for the sake of detection of sleep patterns and determination of sleep parameters. Such measures could be digital biomarkers of the sleep disorders mentioned above [3, 4, 5].

Nowadays, polysomnography (PSG) is considered a gold standard measure of sleep. PSG is a systematic procedure that utilizes electroencephalogram (EEG), electrooculogram (EOG), electrocardiogram (EKG), pulse oximetry, and pulse and respiratory effort [2]. The results of polysomnography are precise and reliable, but they depend on the use of expensive specialized equipment, which is not comfortable for a subject. Moreover, the subject's sleep might be affected by an unusual laboratory environment [4, 6]. Actigraphy is not a replacement for PSG, but it could be a valuable complement

that finds its place in decentralised clinical trials. Nevertheless, the outcomes of this technology highly depend on correctly detected time windows, where the subject is asleep/awake. For this purpose, several approaches have been proposed [3]. This study aims to compare a newly proposed method of sleep/wake detection with a conventional one in terms of consistency with a PSG reference.

2 MATERIALS AND METHODS

2.1 DATABASE

For this study, we used the Newcastle polysomnography and accelerometer dataset that contains 28 adult patients who were scheduled for a one night PSG assessment while wearing the GENEActiv actigraph (Activinsights Ltd, Kimbolton, UK) with a tri-axial accelerometer (sampling frequency $f_s = 87.5$ Hz) on both wrists [7, 8]. In total, 55 recordings are available (data from the right wrist of one participant are missing).

In addition, the newly proposed method was validated in a database consisting of 22 subjects, where each one wore the actigraph on the non-dominant wrist for 6–7 nights (the database contains 150 recordings). Besides, the participants filled in sleep diaries. They were enrolled at the St. Annes University Hospital Brno. All of them signed the informed consent. The study was approved by the local ethics committee.

2.2 METHOD 1

Method 1 (M1) is a numeric method based on an estimation of the arm angle relative to the horizontal plane. The arm angle was estimated using equation:

$$angle = \left(\tan^{-1} \frac{a_z}{\sqrt{a_x^2 + a_y^2}} \right) \cdot \frac{180}{\pi},$$ (1)

where a_x, a_y and a_z are the median values of the orthogonally positioned raw acceleration sensors in g-unit derived based on a rolling five-second time window. In the next step, the algorithm identifies periods larger than 5 minutes with the *angle* change smaller than $10°$ – these intervals are marked as sleep periods [7]. Results were replicated with 3-minute and 10-minute time windows and with $3°$ and $5°$ threshold, respectively.

M1 was precisely validated by V. T. van Hees et al. in [7] on a big Whitehall II study dataset (27981 nights from 4094 participants). In addition, M1 was validated against the Newcastle polysomnography and accelerometer dataset as well [7, 8]. Therefore, we will use M1 as a benchmark for our newly proposed method M2 [7].

2.3 METHOD 2

Method 2 (M2) is our novel method based on a supervised machine learning algorithm. In comparison to M1, besides the accelerometer time series, the proposed method uses data from the temperature sensor as well. M2 consists of several steps. Firstly, the signals are decimated to $f_s = 28.5$ Hz and divided into 30-second segments. Next, the data from the accelerometer are converted into a magnitude representation:

$$magnitude = \sqrt{x^2 + y^2 + z^2}.$$ (2)

Consequently, 45 features are extracted from each segment of the acceleration and temperature series, i.e., in total 90 measures per each 30 s time window. To see the whole set of features, we refer to [4, pp. 73–74].

Finally, the features are modelled by XGBoost, which is a state-of-the-art scalable tree boosting classifier. It is a sparsity aware algorithm, theoretically justifying the weighted quantile sketch for approximate learning [9]. The Newcastle polysomnography and accelerometer dataset is significantly smaller than the Whitehall II study database ([8]), therefore the model was trained and its hyper-parameters were optimised employing k-fold cross-validation ($k = 10$). The PSG data were used as targets. In addition, 60% of the total number of data entries were used for training/validation and 40% for testing.

2.4 EVALUATION METRICS

Both methods were evaluated in terms of accuracy, sensitivity, specificity, F_1 score, and the Mathews correlation coefficient (MCC) (positive cases were used for sleep and negative cases for wake). To evaluate consistency with subject-reported data, the second method was also validated with respect to the information reported in the sleep diaries of the second database.

3 RESULTS

Results based on M1 differ from those reported in [7]. The most balanced results, where $F_1 = 73\%$, were achieved using a 5-minute time window and 10° threshold. The highest accuracy (83%) was based on a 3-minute time window and 10° threshold[1].

During the optimisation phase in M2, we identified the following hyperparameters of the model: the number of gradient boosted trees = 1000, boosting learning rate = 0.1, minimum loss reduction = 0.1, maximum tree depth for base learners = 15, subsample ratio of columns for each level = 0.8, subsample ratio of columns when constructing each tree = 0.5. Some other experiments (there were an attempt to balance the dataset with synthetic data and train the model on a balanced dataset to achieve better results, however, the results were worse [4, pp. 55]) could be found in the original work [4, pp. 49].

As could be seen in figure 1, the learning procedure was stopped after approximately 60 epochs. Figure 1 also displays the ten most important features for the task. The 5th percentile of temperature was identified as the most important measure.

	M1	**M1**	**M2**
	5 min, 10°	3 min, 10°	
accuracy [%]	75	77	89
sensitivity [%]	76	83	97
specificity [%]	70	61	73
F_1	0.73	0.70	0.93
MCC	0.40	0.40	0.75

Table 1: Results on the test set

The overall results for each method are shown in table 1. As can be observed, our newly proposed method significantly outperforms the conventional one. Especially the Mathews correlation coefficient, which measures a trade-off between sensitivity and specificity, reaches much higher values (0.40 vs. 0.75).

Finally, the results where we compared the classification of M2 with subject-reported data are summarised in table 2. Although the method has good sensitivity (73%), it turns out that the specificity is much lower (23%).

---

[1]The results and the code of the pipeline are accessible via https://github.com/xsigmu06/actigraphy-processing.
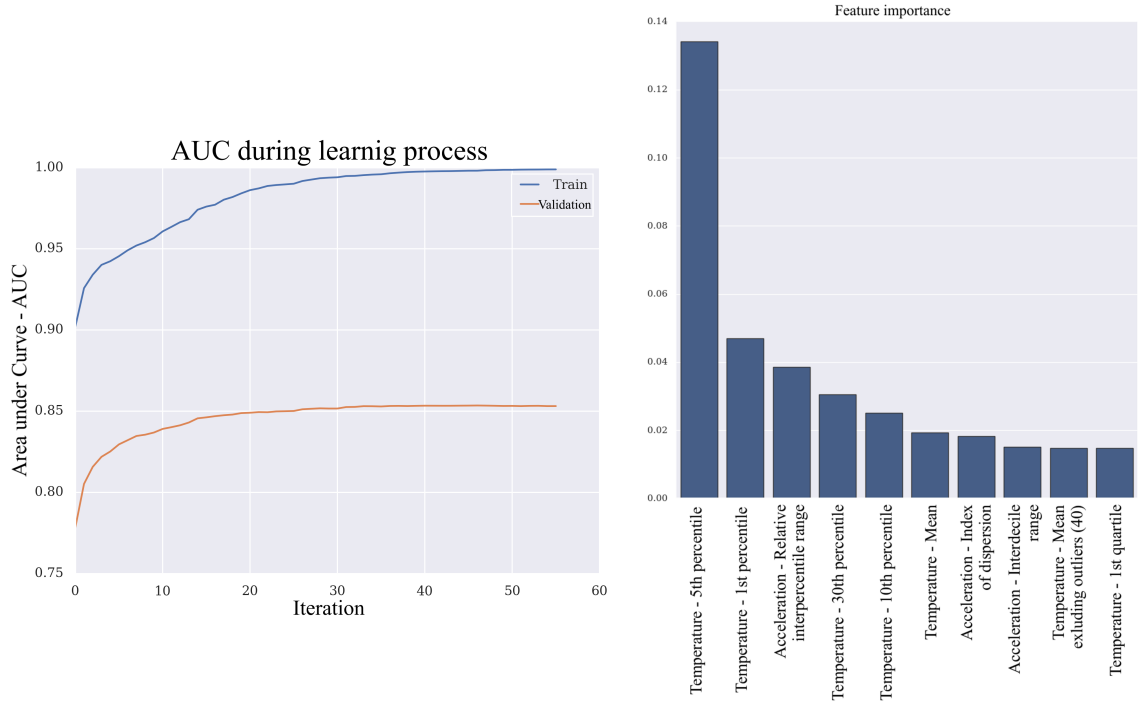
Figure 1: Left part – AUC (area under curve) during the learning process; right part – 10 most important features

Table 2: Evaluation of M2 based on the sleep diaries

accuracy [%]	sensitivity [%]	specificity [%]	F_1	MCC
67	73	23	0.34	-0.03

4 DISCUSSION

In comparison to the statistical method M1, which is based on simple heuristics, our newly proposed approach employing machine learning provides better results in terms of all metrics. Nevertheless, M1 was validated on the big Whitehall II dataset, where it proved to be reliable. Thus, to be fair and before claiming strong conclusions, we are going to apply for the database as well and make further comparisons.

Although our method provided a very good performance on the Newcastle polysomnography and accelerometer dataset, it was much worse when comparing its scores with the self-reported data. Vice versa, V. T. van Hess et al. stated in [7], that M1 has a strong correlation with sleep diary outcomes but a weaker correlation with the PSG data. In other words, M2 seems to be complementary to M1, thus its combination (e.g., inputting M1 scores in the model of M2) could probably enhance the results in both scenarios.

The most important features in M2 are mainly temperature-based. Significant correlations between core-body temperature and sleep/wake patterns were mentioned in previous studies [10]. In the first dataset, the information about the temperature is reliable, because it was recorded in a laboratory environment. However, in the case of the second dataset, we measured the temperature of the wrist which could be easily affected by an environment condition, e.g. by covering the wrist with a duvet. This could be the reason we observed a poor correlation with the subject-reported outcomes.

5 CONCLUSION

We proposed a new method of sleep/wake identification, which is based on actigraphy data with consequent modelling by a gradient boosting algorithm. The method achieves 89% accuracy and Mathews correlation coefficient equal to 0.75 when compared to the polysomnography reference. The new approach outperforms the heuristic method proposed by V. T. van Hess et al. [7], nevertheless, it has a lower correlation with subject-reported outcomes.

ACKNOWLEDGEMENTS

REFERENCES

[1] S. Chokroverty, "Overview of sleep & sleep disorders," Indian J Med Res, vol. 131, no. 2, pp. 126–140, 2010.

[2] K. Crowley, "Sleep and Sleep Disorders in Older Adults," Neuropsychology Review, vol. 21, no. 1, pp. 41–53, Mar. 2011, doi: 10.1007/s11065-010-9154-6.

[3] M. T. Smith et al., "Use of Actigraphy for the Evaluation of Sleep Disorders and Circadian Rhythm Sleep-Wake Disorders: An American Academy of Sleep Medicine Systematic Review, Meta-Analysis, and GRADE Assessment," Journal of Clinical Sleep Medicine, vol. 14, no. 07, pp. 1209–1230, doi: 10.5664/jcsm.7228.

[4] M. Mikulec, "Systém zabezpečeného přenosu a zpracování dat z aktigrafu," Brno, 2020.

[5] V. Farrahi et al., "Calibration and validation of accelerometer-based activity monitors: A systematic review of machine-learning approaches," Gait & Posture, vol. 68, pp. 285–299, nor 2019, doi: 10.1016/j.gaitpost.2018.12.003.

[6] J. V. Rundo and R. Downey, "Chapter 25 – Polysomnography," in Handbook of Clinical Neurology, vol. 160, K. H. Levin and P. Chauvel, Eds. Elsevier, 2019, pp. 381–392.

[7] V. T. van Hees et al., "A Novel, Open Access Method to Assess Sleep Duration Using a Wrist-Worn Accelerometer," PLOS ONE, vol. 10, no. 11, p. e0142533, Nov. 2015, doi: 10.1371/journal.pone.0142533.

[8] V. T. van Hees, S. Charman, and K. Anderson, "Newcastle polysomnography and accelerometer data," Jan. 2018, doi: 10.5281/ZENODO.1160410.

[9] T. Chen and C. Guestrin, "Xgboost: A scalable tree boosting system," in Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining, 2016, pp. 785–794.

[10] M. Uchiyama et al., "Altered phase relation between sleep timing and core body temperature rhythm in delayed sleep phase syndrome and non-24-hour sleep–wake syndrome in humans," Neuroscience Letters, vol. 294, no. 2, pp. 101–104, Nov. 2000, doi: 10.1016/S0304-3940(00)01551-2.

DETECTION OF ROAD SURFACE DEFECTS FROM DATA ACQUIRED BY A LASER SCANNER

Vojtech Myska

Doctoral Degree Programme 3nd year, FEEC BUT

E-mail: xmyska04@vutbr.cz

Supervised by: Radim Burget

E-mail: burgetrm@feec.vutbr.cz

Abstract: Research in the field of automatic detection of road surface defects has been relatively widespread in recent years. Most of the existing works solve this issue by processing the image acquired by camera technology. The contribution of this study is the proposal of the LRS-CNN algorithm for the detection of defects on road surfaces based on their laser scans. The advantage of LRS-CNN is the ability to detect so-called microcracks, which can not be recognized from camera recordings. We have also found that transfer learning methods are not suitable for the use of road defect detection from their laser scans. Our LRS-CNN algorithm has been trained on unique nonpublic data and is able to achieve up to 99.33 % of success depending on the type of task.

Keywords: road damage detection, road surface laser scan, deep learning

1 INTRODUCTION

The density and quality of the road network is one of the driving forces of the world's advanced economies. Therefore, there must be an effort to create quality transport infrastructure in countries that aim to grow economically. These countries include the Czech Republic and, with a 55 000 km long road network, it is one of the top 25 countries with the densest road network [1, 2]. However, this also entails certain risks, such as environmental and people living in the neighborhood impact, the relatively high financial costs required to build new roads or to repair existing roads to maintain the required quality.

The most serious problems that cause a reduction in road life include excessive traffic congestion or adverse weather conditions. This causes the formation of microcracks. Subsequently, water may flow into them and in the event of a drop in temperature below freezing, the defect gradually increases. One of the possible ways to achieve higher quality roads and at the same time reduce their maintenance costs is the early detection of incipient defects in the road surface, such as microcracks. This can be achieved under 2 basic conditions, the first is a technology capable of scanning the surface of roads and an algorithm for automatic processing of acquired data. Scanning of road surfaces in the required quality is already performed, however, the data are evaluated manually. The aim of this study is to create a basic concept of an algorithm for evaluating the quality of road surfaces from data acquired by laser technology. The result is an LRS-CNN algorithm capable of evaluating road surface defects.

2 RELATED WORKS

In recent years, several papers have been published that deal with the issue of road defect detection. The presented approaches are primarily based on image processing techniques using artificial intelligence algorithms. The study [3] deals with the task of detecting more extensive road damage caused by earthquakes from satellite images using the SVM algorithm. At first glance, it may seem relatively distant from the topic of our work, however, the only difference is the point of view on the extent

of road damage. Efforts to use laser sensors to detect road damage resulted in the development of a mobile robotic device in 2013. In this case, the data have been evaluated using the proposed threshold algorithm [4].

In the following years, works using deep learning techniques [5, 6] have been published. Another study [7] deals with the detection of road defects by a mobile phone. Subsequently, several studies have been published in which approaches based on transferred learning methods have been used [8, 9, 10, 11]. The [12] study published this year presents a GAN-based approach to generating pseudo-real road surface defects to solve the problem of insufficient training data.

3 EXPERIMENT

The data set used in this study is composed of nonpublic unique data acquired by laser technology. Specifically, these are the scanned surfaces of the Czech concrete-cement motorways D0, D1 and D11. The high quality of the data is evidenced by the fact that they were scanned with a resolution of 1 mm.

In order to determine the feasibility, areas of 600 x 600 px size containing various severe defects were selected from the original images. Another assumption was the selection of areas without joints or curbs that could have a negative effect on the result of the experiment due to the limited amount of data. Examples of data samples can be seen in figure 1. In this way, 60 samples were generated from each of the monitored classes. The ratio of the distribution of the training to the test set is 9 : 1.
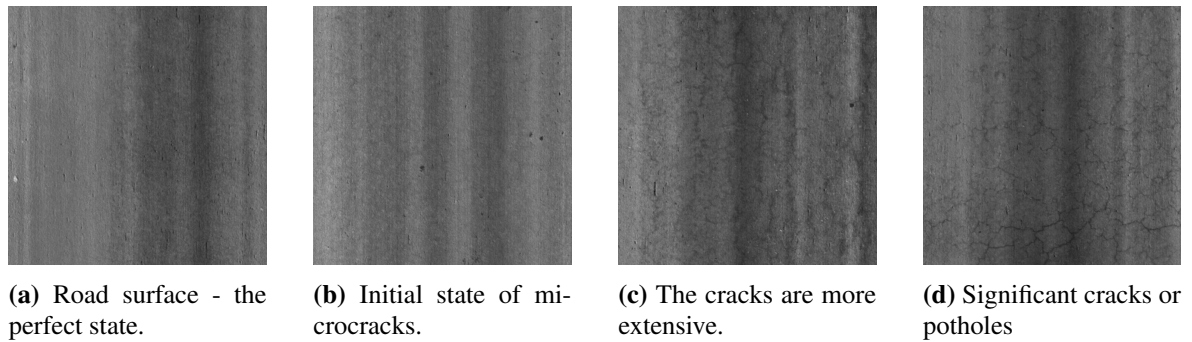

(a) Road surface - the perfect state.

(b) Initial state of microcracks.

(c) The cracks are more extensive.

(d) Significant cracks or potholes

Figure 1: The perfect state is denoted as value 0, whereas the worst as 3.

3.1 PROPOSED NEURAL NETWORKS MODELS

Since we are aware of the benefits provided by the so-called transfer learning methods, we have also experimented with this approach. Unfortunately, the expected result has not been achieved. Therefore, it has been necessary to propose own neural network. The result is a basic design of an architecture called Laser Road Surface - Convolutional Neural Networks (LRS-CNN). In figure 2 is depicted the architecture with settings of layers. The proposed model use only ReLU activation functions. It also includes a number of regularization techniques to minimize overtraining. Namely, a dropout layers, L1 and L2 kernel or activity regularization of the appropriate convolution or fully interconnected layers. Specific setting values vary and are listed for each experiment.

4 RESULTS

The aim of this preliminary study is to determine the feasibility of automatic categorization of laser scans of road surfaces. However, as can be seen from the above examples, the adjacent defect classes are very similar. Thus, several types of experiments have been demonstrated. They mainly differ in the number of distinguished classes.
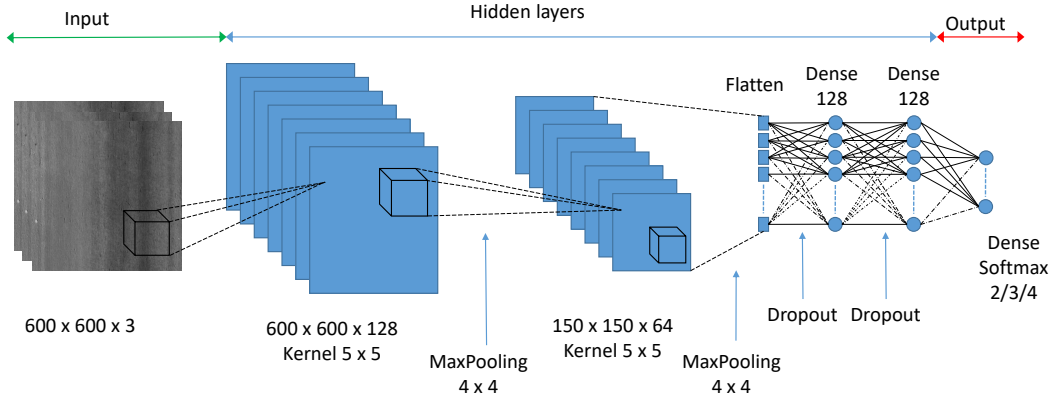
Figure 2: Proposed architecture of the LRS-CNN for the road surface defect detection.

In order to reproduce the results, the TensorFlow and Keras seed variables have been set. Furthermore, due to the amount of available data, we have used cross-validation method. This experiment has been performed on a PC with Windows OS, Intel i7-8700k, 32 GB RAM and Nvidia 1080 Ti using high-level API Keras 2.4.1 and TensorFlow 2.2.0.

The learning rate has been set to 6.67×10^{-5}. In all experiments, Adam optimization algorithm have been used. Based on the monitoring of the loss metric, the callbacks Early Stopping (patient 20) and Reduce Learning Rate (patient 10, factor 0.667) have been set.

4.1 ROAD SURFACE DETECTION TASKS

The basic experiment aims at the simplest task, namely to distinguish a surface on which no defects occur and which contains the most serious defects. The motivation is that in case of failure, it is not necessary to train LRS-CNN in order to have a finer classification.

In practically all groups of cross-validation except the best, overtraining is evident, despite the setting of the regularization setting. We have added to the first convolutional layer kernel regularization $L1 = 0.01$, to the first dense layer kernel regularization $L1 = 0.001, L2 = 0.001$ and dropout rate 0.4. By this settings, our approach has reached up to 93.33 % classification accuracy (the best k-group). The average is 77.03 %. Figure 3 depicts graphical course of the training processes of all three sub-experiments, the binary classification training process is marked in blue. For the sake of clarity, the value of the loss function in the graph is limited to a maximum of 5. Relatively high values of the loss function were achieved precisely thanks to the use of L1 and L2 regularization techniques.

Based on the results obtained in the previous experiment, it has been decided to increase the number of detected classes. Due to the extremely small difference between the fault-free state and the beginning microcracks, we decided to omit class 1 from the detection. Therefore, the data set has been extended by class 2. The detection success has been expected to be lower than in the previous case. Nevertheless, with 66.55 % average classification accuracy (the best k-group 77.77 %) the LRS-CNN algorithm exceeded our expectations. Unfortunately, the model also struggles with overfitting despite all the settings. We have increase the L1 kernel regularization of the first dense layer to the value $L1 = 0.01$. Again, as in the previous experiment, satisfactory results have not been obtained via ResNet model.

In the last experiment, the aim has been to verify the success of the detection of all 4 monitored classes. Unfortunately, the last experiment showed more significant overtraining despite all efforts to tune the LRS-CNN hyperparameters. This phenomenon is most likely the result of very subtle

differences between the individual classes and the limited number of training samples. The LRS-CNN have reached up to 70.83 % accuracy in the best k-group. Unfortunately, the average is much lower - 55.41 %. Detailed results can be found in table 1. As mentioned several times, the ResNet model has not been able to learn to solve the problem of laser scans.

Table 1: Achieved results all experiments. The proposed approach is marked in bold.

| Task	k-group	LRS-CNN			ResNet 50		
		Train acc.	Test acc.	Test loss	Train acc.	Test acc.	Test loss
	-	-	-	-	-	-	-
Binary detection	The best	0.9629	0.9333	0.9702	0.5200	0.5000	2.2380
	The worst	0.8981	0.5333	1.3768	0.5000	0.5000	2.8324
	Average	0.9444	0.7703	1.2089	0.5033	0.5000	2.5879
Leave the value 1	The best	0.8333	0.7777	0.9916	0.3457	0.3333	2.3409
	The worst	0.7777	0.5000	1.9366	0.2901	0.2222	3.1639
	Average	0.8125	0.6555	1.3624	0.3086	0.2989	2.6879
All classes	The best	0.8395	0.7083	1.3130	0.2639	0.2500	3.2113
	The worst	0.7530	0.3750	1.9633	0.2901	0.2500	4.0100
	Average	0.7967	0.5541	1.5853	0.2407	0.2500	3.8336
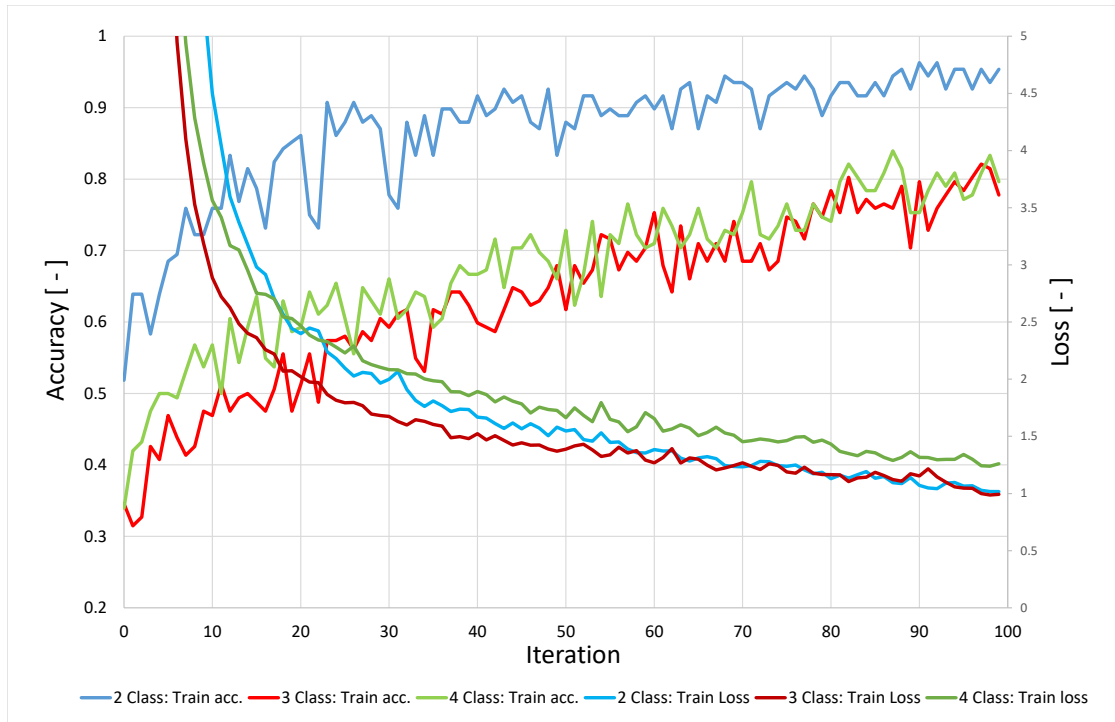
Figure 3: Graphical dependence of the course of the training process on iterations.

5 CONCLUSION

In this study, we have laid the foundation for the research of an automatic system for detecting the state of the road surface from their high-resolution laser scans. We have presented a different view of the issue than studies focused on the same topic. We have designed the LRS-CNN architecture, which has been trained in several experiments on a unique non-public dataset. In the basic experiment, it has achieved 93.33 % (77.07 % average) accuracy in binary classification task. With a finer division

of the surface defects, the LRS-CNN has achieved 77.77 % (65.55 % average) accuracy. In the last experiment, where microcrack samples have been included, the lowest accuracy has been achieved 70.83 % (55.41 % average).

We believe that as part of further research, we can improve the LRS-CNN algorithm and thus contribute to the automation of the evaluation of defects occurring on the road surface.

REFERENCES

[1] Czech statistical office *Transport infrastructure - time series*. [online]. Available from: `https://www.czso.cz/csu/czso/transport_infrastructure_time_series`

[2] OECD *Indicators : Transport infrastructure*. [online]. Available from: `https://stats.oecd.org/Index.aspx?DataSetCode=ITF_INDICATORS`

[3] LI, Peijun; XU, Haiqing; SONG, Benqin. *A novel method for urban road damage detection using very high resolution satellite imagery and road map.* Photogrammetric Engineering & Remote Sensing, 2011, 77.10: 1057-1066.

[4] T. Yamada, T. Ito and A. Ohya, *Detection of road surface damage using mobile robot equipped with 2D laser scanner*, Proceedings of the 2013 IEEE/SICE International Symposium on System Integration, Kobe, Japan, 2013, pp. 250-256, doi: 10.1109/SII.2013.6776679.

[5] L. Zhang, F. Yang, Y. Daniel Zhang and Y. J. Zhu, *Road crack detection using deep convolutional neural network*, 2016 IEEE International Conference on Image Processing (ICIP), Phoenix, AZ, USA, 2016, pp. 3708-3712, doi: 10.1109/ICIP.2016.7533052.

[6] ZHANG, Allen, et al. *Automated pixel-level pavement crack detection on 3D asphalt surfaces using a deep-learning network.* Computer-Aided Civil and Infrastructure Engineering, 2017, 32.10: 805-819.

[7] MAEDA, Hiroya, et al. *Road damage detection and classification using deep neural networks with smartphone images.* Computer-Aided Civil and Infrastructure Engineering, 2018, 33.12: 1127-1141.

[8] L. Ale, N. Zhang and L. Li, *Road Damage Detection Using RetinaNet*, 2018 IEEE International Conference on Big Data (Big Data), Seattle, WA, USA, 2018, pp. 5197-5200, doi: 10.1109/BigData.2018.8622025.

[9] W. Wang, B. Wu, S. Yang and Z. Wang, *Road Damage Detection and Classification with Faster R-CNN*, 2018 IEEE International Conference on Big Data (Big Data), Seattle, WA, USA, 2018, pp. 5220-5223, doi: 10.1109/BigData.2018.8622354.

[10] Y. J. Wang, M. Ding, S. Kan, S. Zhang and C. Lu, *Deep Proposal and Detection Networks for Road Damage Detection and Classification*, 2018 IEEE International Conference on Big Data (Big Data), Seattle, WA, USA, 2018, pp. 5224-5227, doi: 10.1109/BigData.2018.8622599.

[11] ARYA, Deeksha, et al. *Transfer learning-based road damage detection for multiple countries.* arXiv preprint arXiv:2008.13101, 2020.

[12] MAEDA, Hiroya, et al. *Generative adversarial network for road damage detection.* Computer-Aided Civil and Infrastructure Engineering, 2021, 36.1: 47-60.

BRNO UNIVERSITY OF TECHNOLOGY
FACULTY OF ELECTRICAL ENGINEERING AND COMMUNICATION
TECHNICKÁ 3058/10, 616 00 BRNO, CZECH REPUBLIC

www.eeict.cz

www.fekt.vut.cz