STUDENT EEICT

BRNO UNIVERSITY OF TECHNOLOGY
FACULTY OF ELECTRICAL ENGINEERING AND COMMUNICATION

# PROCEEDINGS I OF THE 30TH STUDENT EEICT 2024

ELECTRICAL ENGINEERING, INFORMATION SCIENCE,
AND COMMUNICATION TECHNOLOGIES

# GENERAL PAPERS

April 23, 2024
Brno, Czech Republic

BRNO UNIVERSITY OF TECHNOLOGY FACULTY OF ELECTRICAL ENGINEERING AND COMMUNICATION


**Title** Proceedings I of the 30th Conference STUDENT EEICT 2024

**Editor** Assoc. Prof. Vítězslav Novák

**Publisher** Brno University of Technology, Faculty of Electrical Engineering and Communication

**Year released** 2024

1st Edition



ISBN 978-80-214-6231-1

BRNO UNIVERSITY OF TECHNOLOGY
FACULTY OF ELECTRICAL ENGINEERING AND COMMUNICATION


# SPONSORS

## Diamond tier

Honeywell

## Golden tier

EATON

onsemi

ThermoFisher
SCIENTIFIC

BRNO UNIVERSITY OF TECHNOLOGY
FACULTY OF ELECTRICAL ENGINEERING AND COMMUNICATION


# SPONSORS

## Silver tier

ABB

ASICentrum
A COMPANY OF THE SWATCH GROUP

Garrett
ADVANCING MOTION

NXP

Valeo

## Honorary tier

AT&T

e·on

SKODA

# TECHNICAL SPONSORS & MEDIA PARTNERS

IEEE
IEEE Czechoslovakia Section

Elektrika.cz

VERONI
kafestánek

BRNO UNIVERSITY OF TECHNOLOGY
FACULTY OF ELECTRICAL ENGINEERING AND COMMUNICATION

# Contents

## Master's Projects

### M1 – Biomedical Engineering



### M2 – Industrial Automation, Instrumentation and Engineering



### M3 – Communication and Information Systems, Network Security I.



### M4 – Communication and Information Systems, Network Security II.



### M5 – Sensors and Measurement Systems and Technology I.

M6 – Sensors and Measurement Systems and Technology II.

## Ph.D. Projects

### D1 – Biomedical Engineering



### D2 – Audio, Speech and Language Processing



### D3 – Communication and Inf. Systems, Network Services and Security



### D4 – Microelectronics and Technology

## D5 – Sensors and Measurement Systems



## D6 – Robotics and Artificial Intelligence



## D7 – Numerical Modelling, Simulation and Measurement

## High School Projects

### S1 – General Technical Focus


# OPENING WORD OF THE DEAN

These Proceedings contain papers presented during the **30th annual STUDENT EEICT conference**, held at the Faculty of Electrical Engineering and Communication, Brno University of Technology, on April 23, 2024. 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.

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, 137 peer-reviewed full papers distributed between 18 sessions were submitted, 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 valuable gifts.

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 30th STUDENT EEICT (2024) 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

# Biometric fingerprint liveness detection

Lukáš Rišian, Ing. Martin Vítek, Ph.D.
*Department of Biomedical Engineering*
*Faculty of Electrical Engineering and Communication*
Brno University of Technology, Czech Republic
240556@vutbr.cz, vitek@vutbr.cz

***Abstract*—This work addresses the problem of biometric recognition of fingerprint liveness to identify and differentiate between real fingerprints and their artificial replicas. The main objective was to identify the features that are crucial for fingerprint liveness recognition and based on these features to propose an efficient classification algorithm. We worked with the LivDet database from 2009, which contains both real and fake fingerprints. This database has been used in a worldwide competition and the results of all implemented algorithms are publicly available for subsequent comparison of success rates. An important part of this work was the preprocessing of the image data, which was crucial for testing the selected features and implementing the algorithms. We analyzed more than 180 different features from which we selected the most relevant ones. We then used the selected features to develop several fingerprint recognition and classification algorithms. Using the selected features, several possible variations of the algorithms have been proposed. Among all the implemented algorithms, we achieved the best result of almost 90%. Compared to other algorithms that have been implemented for the same purpose and have been used and tested on the same database, this can be considered a satisfactory and reliable result. In conclusion, the main objective of this work was to provide an efficient, secure, and reliable solution in the field of biometric fingerprint spoof detection.**

***Index Terms*—Biometric, liveness recognition, fingerprint, image segmentation, machine learning, features, database, algorithm, attributes, security, database, result, identification, classification.**

## I. INTRODUCTION

Today, biometric technologies play a key role in identifying and verifying the identity of individuals. Biometrics uses unique characteristics of a person, such as fingerprints, which are characterized by their universality, uniqueness, and consistency. An example of different fingerprints is shown in Fig. 1.. These characteristics are used as electronic keys or passwords that can be used to identify an individual. With advances in science and technology, the risk of falsification and unauthorized access to these unique features is increasing. The development of new materials offers the possibility of imitating real fingers using artificial imitations. These artificial replicas can be misused for unauthorized access. Therefore, biometric recognition of real fingerprints is now a critical aspect of security and protection. This work presents a software solution to ensure the authenticity of real fingerprints and protect them from unauthorized use through artificial imitations. The proposed algorithms are aimed at the reliable, efficient, and fast detection of fake fingerprints. [1] [2]

Fig. 1. Example of a fingerprints.

## II. DATABASE

The LivDet database was created as part of the worldwide Fingerprint Liveness Detection Competition. The main reason for the selection of this specific data source is a sufficient number of images that include both real and fake fingerprints. The fake fingerprints were made from materials such as silicone, gelatin, and wood glue. This database contains fingerprints from three different sensors (Biometrika, Crossmatch, and Identix) and is divided into training and test sets, which will be useful in designing our fingerprint classification models. The number of real and fake fingerprints in both the training and test sets is shown in TABLE I.

TABLE I
NUMBER OF FINGERPRINTS IN LIVDET DATABASE

| Number of fingerprints | Biometrika | Crossmatch | Identix |
|---|---|---|---|
| Real fingerprints in train set | 520 | 375 | 500 |
| Fake fingerprints in train set | 520 | 375 | 500 |
| Real fingerprints in test set | 1473 | 1125 | 1500 |
| Fake fingerprints in test set | 1480 | 1125 | 1500 |

## III. LIVENESS DETECTION

As already mentioned, one of the main problems with current biometric systems is the attempts at falsification through various artificial imitations, particularly artificial fingers made of silicone. These forgeries are often characterized by imperfections and deviations, which we are trying to identify to distinguish real fingerprints from fake ones. Two main methods are used for liveness detection: hardware and software approaches.

### *A. Hardware methods*

Hardware-based methods represent the first approach, which involves sensors with integrated devices that analyze the finger's liveness based on various parameters such as pulse, temperature, or conductivity of the finger skin. These methods are often considered to be very accurate, but they are complex and challenging to implement. Associated with this accuracy is the high operating cost of these devices.

### B. Software methods

Software methods represent a more cost-effective and attractive way of detecting liveness. These methods are based on preprocessing the input fingerprint and identifying certain characteristics such as textures, edges, and other unique features that allow the recognition of real and fake fingers. Although these methods are often more sensitive to details compared to hardware approaches, they require more computational capacity, which is not a significant problem nowadays. In software-based classification methods, using artificial intelligence and machine learning is the most effective. There are many ways to implement an algorithm suitable for this type of detection, and their effectiveness and accuracy depend on the chosen liveness recognition features and the ability of the algorithms to recognize these differences. An example of a real and fake fingerprint is shown in Fig. 2. [3] [4]

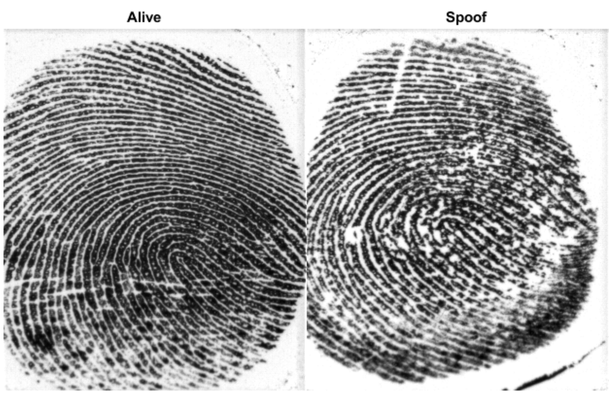



Fig. 2. Example of real and spoof fingerprints.

## IV. Preprocessing

The preprocessing of the input image was performed in a MATLAB environment. This process represents a key step in the classification algorithm since it is necessary to extract the resulting segment (region of interest) of the fingerprint from the input image. The input image includes not only the fingerprint itself, but also the surroundings of the sensor, which we try to minimize or, in the best case, completely remove. The resulting fingerprint segment will then be tested for features that should provide relevant liveness information, and the best of these will be used in the next stage of the algorithm. A key step in the actual preprocessing is to identify the boundary that separates the fingerprint from the unwanted sensor surroundings. This is followed by the process of creating a mask that represents the segment of interest, and then the selected features are tested on the resulting fingerprint mask. The image preprocessing process is shown in Fig 3.

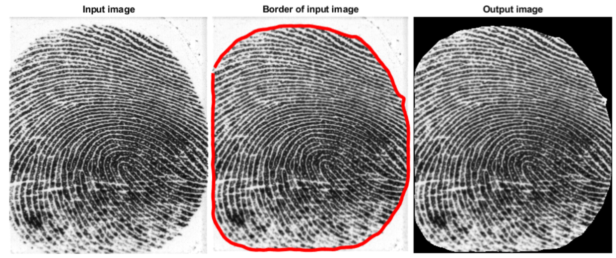



Fig. 3. Preprocessing of input image.

## V. Features

In this work, we have analyzed more than 180 different features to identify fingerprint liveness as best, accurately, and reliably as possible. All the tested features provide numerical values that show some differences between real fingerprints and their artificial replacements. All obtained features are extracted from a segment of our preprocessed fingerprint, and we used various functions in the MATLAB environment and mathematical equations to obtain the resulting values for each feature. These features include, for example, the contrast, entropy, or energy of the fingerprint segment, as well as features resulting from the histogram of our segment, such as the variance, mean, steepness, or skewness of the histogram. In total, we obtain 189 different numerical features for each single sensor. However, not every feature showed significant differences between real and spoof fingerprints; some differences were minimal, even zero. Other features showed relatively large differences in liveness and were appropriate for subsequent use in our algorithms. Accordingly, we decided by appropriate sorting and selection to choose only those features that were most informatively relevant to our purpose.

### A. Sorting

The first step in the selection was to sort all features into informationally significant and non-significant. Many features failed to provide sufficient differences in liveness that would subsequently be used in the algorithm. The next step was to divide the informationally significant features into groups based on their correlation. There were cases where completely different features showed informationally identical values. This whole procedure was repeated for each sensor separately.

### B. Selection

Due to the variability of qualities and parameters for each sensor, it has become challenging to find a universal feature for all three sensor types. From the large set of features that correlated with each other, we selected the best ones to include the same property in at least two sensors. In this way, we have at least partially generalized the feature-based finger liveness recognition process. Despite our attempts to find as many universal features as possible, we also had to select the features that are most appropriate for a given sensor type and thus have the potential to increase the accuracy of our algorithm. [5]

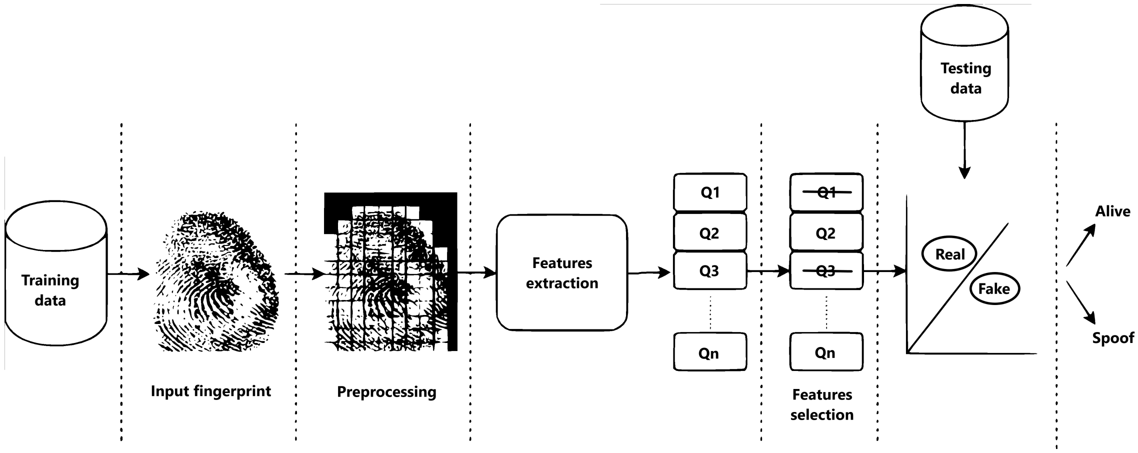



Fig. 4. Block scheme of liveness detection.

## VI. Algorithms

For each sensor type, we developed custom algorithms that exploit the features that provide the most relevant differences in terms of fingerprint liveness. The entire liveness detection process along with image preprocessing and selection of valence features is shown in Fig. 4. We used 7 features for the Biometrika sensor, 10 for the Crossmatch sensor, and 9 for the Identix sensor. We tested several variants of the algorithms. The simplest type of algorithm involved the use of individual thresholds for each feature. Another variant was to include a combination of all and the top three features. Combining the features means that we test all the selected features on a particular fingerprint and if multiple features classify the fingerprint as real, the resulting classification will be marked as a real fingerprint. In contrast, if the majority of the features identify the fingerprint as fake, the resulting classification will identify the fingerprint as fake. Algorithms that already use machine learning to train the model and then classify the fingerprints according to authenticity follow, namely SVM (Support Vector Machine), Decision Tree, and Random Forest methods. For SVM, we have built models without parameters and also with added parameters to optimize our model and contribute to better classification success. The results of all algorithms and individual features for each sensor are shown in Tables II-VII.

TABLE II
Results of individual features for Biometrika sensor

| Individual features | ACC | TP | TN | FP | FN |
|---|---|---|---|---|---|
| MSE (Mean squared error) | 0,8608 | 1392 | 1150 | 330 | 81 |
| Energy of histogram | 0,7650 | 1127 | 1132 | 348 | 346 |
| STD of square segment | 0,5723 | 896 | 794 | 686 | 577 |
| CV of magnitude spectrum | 0,6729 | 871 | 1116 | 364 | 602 |
| GLCM - Correlation | 0,8456 | 1261 | 1236 | 244 | 212 |

Of all the features, only 5 were usable for thresholds in the Biometrika sensor. The SVM algorithm with optimized parameters achieves the most significant result, namely a success rate of 89%, which is a satisfactory result in terms of the security and reliability of the algorithm. Other machine learning-based algorithms and simpler algorithms that combine features have slightly lower success rates. Among the algorithms based on thresholds for individual features, the MSE feature stands out with a detection success rate of 86%.

TABLE III
Results of different algorithms for Biometrika sensor

| Algorithm | ACC | TP | TN | FP | FN |
|---|---|---|---|---|---|
| Combination of all features | 0,8320 | 1253 | 1204 | 276 | 220 |
| Combination of 3 best features | 0,875 | 1348 | 1236 | 244 | 125 |
| SVM without parameters | 0,8619 | 1241 | 1298 | 175 | 232 |
| SVM with parameters | 0,8989 | 1273 | 1375 | 98 | 200 |
| Decision tree | 0,8483 | 1234 | 1265 | 208 | 239 |
| Random forest | 0,8680 | 1287 | 1270 | 203 | 186 |

TABLE IV
Results of individual features for CrossMatch sensor

| Individual features | ACC | TP | TN | FP | FN |
|---|---|---|---|---|---|
| Overall quality | 0,6543 | 1176 | 787 | 713 | 324 |
| CV of magnitude spectrum | 0,7677 | 982 | 1321 | 179 | 518 |
| STD of magnitude spectrum | 0,6913 | 1017 | 1057 | 443 | 483 |
| CV of histogram | 0,8340 | 1304 | 1198 | 302 | 196 |
| VAR of histogram | 0,8403 | 1300 | 1221 | 279 | 200 |
| Energy % of the magnitude spectrum | 0,8120 | 1295 | 1141 | 359 | 205 |
| GLCM - Correlation | 0,7343 | 1369 | 804 | 696 | 131 |
| Gray Level Ratio 2 | 0,8410 | 1296 | 1227 | 273 | 204 |
| Contrast of image | 0,8227 | 1253 | 1215 | 285 | 247 |
| Quadratic mean | 0,8313 | 1354 | 1140 | 360 | 146 |

TABLE V
Results of different algorithms for Crossmatch sensor

| Algorithm | ACC | TP | TN | FP | FN |
|---|---|---|---|---|---|
| Combination of all features | 0,8410 | 1270 | 1253 | 247 | 230 |
| Combination of 3 best features | 0,8477 | 1323 | 1220 | 280 | 177 |
| SVM without parameters | 0,8473 | 1142 | 1400 | 100 | 358 |
| SVM with parameters | 0,8463 | 1135 | 1404 | 96 | 365 |
| Decision tree | 0,8287 | 1108 | 1378 | 122 | 392 |
| Random forest | 0,8463 | 1112 | 1427 | 73 | 388 |

A combination of 3 best features achieves the best results for this sensor. Except for decision trees, all machine learning-based algorithms achieve a success rate of approximately 84%. Algorithms that combine the selected features, either all or only the top 3, achieved almost the same result. Among the individual thresholds, the Gray Level Ratio 2 and Histogram Variance features dominate, also achieving a success rate of 84%.

TABLE VI
RESULTS OF INDIVIDUAL FEATURES FOR IDENTIX SENSOR

| Individual features | ACC | TP | TN | FP | FN |
|---|---|---|---|---|---|
| Contrast | 0,7724 | 719 | 1019 | 106 | 406 |
| Energy of histogram | 0,7604 | 672 | 1039 | 86 | 453 |
| Energy % of the magnitude spectrum | 0,7547 | 977 | 721 | 404 | 148 |
| Gray Level Ratio 1 | 0,7756 | 1058 | 687 | 438 | 67 |
| CV of histogram | 0,7422 | 749 | 921 | 204 | 376 |
| Quadratic mean | 0,8422 | 910 | 985 | 140 | 215 |
| Square segment - mean | 0,7533 | 808 | 887 | 238 | 317 |
| Square segment - STD | 0,6258 | 733 | 675 | 450 | 392 |
| GLCM - Correlation | 0,7707 | 850 | 884 | 241 | 275 |

TABLE VII
RESULTS OF DIFFERENT ALGORITHMS FOR IDENTIX SENSOR

| Algorithm | ACC | TP | TN | FP | FN |
|---|---|---|---|---|---|
| Combination of all features | 0,8302 | 886 | 982 | 143 | 239 |
| Combination of 3 best features | 0,8556 | 949 | 976 | 149 | 176 |
| SVM without parameters | 0,8547 | 915 | 1008 | 117 | 210 |
| SVM with parameters | 0,8471 | 909 | 997 | 128 | 216 |
| Decision tree | 0,8240 | 976 | 878 | 247 | 149 |
| Random forest | 0,8933 | 1026 | 984 | 141 | 99 |

The Random Forest algorithm achieved the best result for the Identix sensor with a success rate of 89%. It was immediately followed by an algorithm where we combined the 3 best features, which achieved a success rate of 85%. The SVM algorithms without parameters also achieved the same successful results. Among the individual thresholds, the Quadratic Average attribute stands out, achieving a success rate of 84%, which is comparable to the results of the more advanced algorithms.

## VII. DISCUSSION AND CONCLUSION

### A. Discussion

In this work, we have analyzed in detail the detection of fingerprint liveness through software methods that exploit image features and corresponding algorithms. Our goal was to find informationally significant features and implement them in the corresponding algorithms. For this purpose, we analyzed 189 different image features in detail and identified the most relevant ones for each sensor. Using these features, we designed different algorithms, ranging from simple ones (single thresholds) to more complex ones that combined multiple features and used artificial intelligence.

An important step in this process was the preprocessing of the input image, which allowed us to isolate the segment of our interest, the fingerprint itself. All the fingerprints used for the feature testing and algorithm development were obtained from the LivDet 2009 database.

We designed separate algorithms for each sensor, with best success rates between 85% and 89%. The machine learning-based algorithms performed better and more accurately than the simpler types of algorithms (separate thresholds and combination of features). The success rate of the individual thresholds was on average 5% lower than that of the machine learning-based algorithms. The combinations of the features also achieved quite reliable results. Their success rates ranged between the results of the individual thresholds and those of the algorithms using machine learning. These results highlight the potential of combination approaches and emphasize the importance of the proper selection of features to achieve optimal results.

Table VIII shows the algorithm success rates of all participants who took part in the competition and their algorithms were tested on the LivDet 2009 database. The success results of each algorithm for each sensor are represented by the FNR (Rate of misclassified live fingerprints) and FPR (Rate of misclassified spoof fingerprints) values. We compared the results of our best algorithms with those of other participants and found that our algorithms perform similarly or even better. As a result, we ranked in the top ranks among the other participants, indicating that our algorithms achieved reliable and satisfactory results.

TABLE VIII
COMPARISON OF THE RESULTS OF ALL PARTICIPANTS

| Participant | Biometrika | | Crossmatch | | Identix | |
|---|---|---|---|---|---|---|
| | FNR | FPR | FNR | FPR | FNR | FPR |
| Dermalog | 74,1% | 1,9% | 7,4% | 11,4% | 2,7% | 2,8% |
| ATVS | 71,7% | 3,1% | 8,8% | 20,8% | 9,8% | 3,1% |
| Anonym1 | 56% | 17,6% | 27,1% | 18,9% | 15,2% | 11,5% |
| Anonym2 | 15,6% | 20,7% | 14,4% | 15,9% | 9,8% | 11,3% |
| Lukáš Rišian | 13,6% | 6,7% | 11,8% | 18,7% | 8,8% | 12,5% |

### B. Conclusion

The goal of our work was to identify suitable features for fingerprint liveness detection and implement them in an efficient algorithm. Given the variability of sensor parameters, we selectively chose the most relevant features for each sensor and developed several separate algorithms. This process was time-consuming but necessary to achieve optimal results. The performance of our algorithms is acceptable in terms of security and liveness detection. We tested in total of 189 features, which maximized the efficiency and success rate of our algorithms. The highest success rate among all the algorithms reached a value of almost 90%, which is considered a satisfactory result in the context of fingerprint detection.

## REFERENCES


[1] JAIN, Anil K.; ROSS, Arun A. a NANDAKUMAR, Karthik, 2011. *Introduction to Biometrics.* Springer Science & Business Media. ISBN 0387773266.

[2] LEE, Henry C. a GAENSSLEN, R. E., 2001. *Advances in Fingerprint Technology.* 2. New York: CRC Press. ISBN 0-8493-0923-9.

[3] GALBALLY, J., F. ALONSO-FERNANDEZ, J. FIERREZ a J. ORTEGA-GARCIA. *A high performance fingerprint liveness detection method based on quality related features*. Future Generation Computer Systems. 2012, roč. 28, č. 1, s. 311-321.

[4] MALTONI, Davide; MAIO, Dario; JAIN, Anil K. a PRABHAKAR, Salil, 2009. *Handbook of Fingerprint Recognition*. Springer. ISBN 978-1-84882-253-5.

[5] GONZALEZ, Rafael C. a WOODS, Richard E., 2008. *Digital Image Processing*. Online. 3. Pearson Prentice Hall. ISBN 978-0-13-168728-8.

# Interconnection of three-dimensional printing, material science and bone tissue engineering for creating ideal bone scaffolds

Veronika Šromová
*Department of Physics*
*FEEC, Brno University of Technology*
Technická 12, 616 00 Brno, Czech Republic
237011@vut.cz

***Abstract*—The interconnection of three-dimensional (3D) printing, material science, and bone tissue engineering holds immense promise in the realm of scaffold development for bone repair. This work delves into the intricate interplay between various factors influencing the success hydroxyapatite biomaterials in bone regeneration, including material selection, porosity, structural design, surface modification, and other mechanical characteristics. Our aim is the pursuit of ideal biomaterials possessing both favorable biological responses and mechanical integrity. Fourier transform infrared spectroscopy (FTIR) is used to analyze the chemical structure of the material, particularly the chemical bonds present. Using SEM, we were subsequently able to analyze the microstructure and nanostructure of the 3D printed hydroxyapatite sample.**

***Keywords—hydroxyapatite, 3D printing, FTIR, SEM***

## I. Introduction

The mineral component of bone is primarily composed of calcium and phosphate, forming hydroxyapatite crystals $Ca_{10}(PO_4)_6(OH)_2$, which are embedded within the collagen matrix. These minerals provide hardness and structural strength and exhibit structural disorder and nonstoichiometry in contrast to synthetic stoichiometric hydroxyapatite (HA). This is because bone mineral contains a significant amount of anionic and cationic species and other incorporations within its crystal lattice [1,2]. This leads to other insights we have about hydroxyapatite. It stands out as the most utilized ceramic biomaterial due to its similarities to bone mineral in terms of structural and mechanical properties. On the other hand, pure hydroxyapatite can exhibit some inadequate mechanical characteristics, including low tensile and compressive strength, which makes it unsuitable for bone tissue engineering. Despite hydroxyapatite's poor mechanical characteristics, its distinctive biological qualities prompt us to consider efforts to enhance its properties rather than entirely substituting it with other biomaterials. This is the reason why hydroxyapatite is frequently combined with various incorporations to enhance its mechanical properties and overall efficacy of implantable biomaterials [3].

In the dynamic field of material science and bone tissue engineering, we examine the potential of hydroxyapatite-based biomaterials, especially their properties, including porosity, surface modifications, rigidity, and other mechanical characteristics, all of which play pivotal roles in facilitating cell attachment, proliferation, and ultimately, bone regeneration.

Central to our exploration are three different samples, differing in their internal and external structure, porosity, as well as mechanical resistance and stability. Dense, compact sample made of pure hydroxyapatite was prepared by hydraulic pressing of hydroxyapatite powder. Porous hydroxyapatite foam was created using foaming methods which utilize ceramic suspensions as a primary material. These methods involve incorporation of gas bubbles into the ceramic suspensions, followed by consolidation of the mixture. Then the sample goes through a process of burning out which leads to a porous structure [4]. A 3D printed sample represents a method of using a 3D additive manufacturing technique. Three-dimensional printing as such has emerged as a favored fabrication method for creating HA-p based scaffolds due to its precise and controllable nature [5].

## II. Fourier Transform Infrared Spectroscopy (FTIR)

FTIR spectroscopy is a non-destructive method that investigates the interaction between matter and electromagnetic radiation, which manifests as a unique spectrum revealing molecular characteristics. By examining specific bands or groups within the sample, FTIR spectroscopy precisely identifies molecular structures, bonding patterns, functional groups, and even intermolecular interactions. The method itself aims to measure the amount of light absorbed by a sample at various frequencies [6].

Infrared radiation spectrum (IR spectrum), part of the electromagnetic spectrum with longer wavelengths than visible light, is divided into near-IR, mid-IR, and far-IR. Medical applications primarily use mid-IR, covering wavelengths from 4000 $cm^{-1}$ do 400 $cm^{-1}$, where absorbed radiation frequencies correspond to vibrational interactions at the subatomic level. Said in other words, absorption of infrared radiation induces vibrational interactions within molecules. Particularly in the mid-infrared spectral ranges, this phenomenon occurs when the frequencies of light and vibration coincide, and when there is a change in the molecular dipole moment during the vibration [6,7].

Thus, when we apply FTIR to hydroxyapatite, we can detect specific bands in HA, specific for $PO_4^{3-}$, and $OH^-$ functional groups. Phosphate functional groups of compacted sample exhibit four different vibrational modes and form intensive peaks at 571 $cm^{-1}$, 599 $cm^{-1}$, 1032 $cm^{-1}$ and at 1112 $cm^{-1}$. The stretching and bending vibrations of adsorbed water result in the formation of peaks at wavenumbers 636 $cm^{-1}$ and 3525 $cm^{-1}$. The 3D printed sample shows slightly lower peaks in that regions compared to the compacted sample, while porous sample exhibits the lowest peaks. Possible differences in the intensities and the area beneath the bands are closely associated with the concentration of individual functional groups [8].

Therefore, FTIR spectrum of presented samples shows characteristic peaks of hydroxyapatite and it follows that detection of these functional groups serves as a main indicator of the formation of the hydroxyapatite structure [8,9].

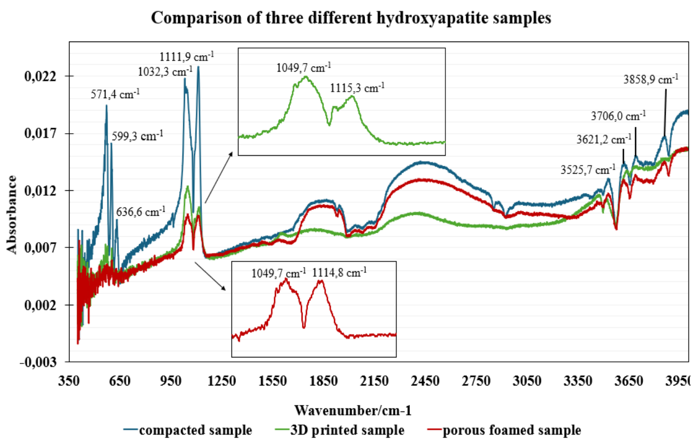



Fig. 1. Graph representing FTIR spectra of three different hydroxyapatite samples.

The bands near 1550 $cm^{-1}$ could be from the carbonate functional group [10].

## III. Scanning Electron Microscopy (SEM) Images of 3D Printed aHydroxyapatite Sample

While optical microscopy relies on lights, scanning electron microscopy (SEM) utilizes electron emissions for imaging. Optical microscopy offers lower magnification, ranging from 400-1000 times original size, while SEM can reach magnifications up to 300 000times and analyze organic and inorganic materials on a nanometer to micrometer scale. This implies that optical microscopy has its limitations in observing small and thin samples. Conversely, SEM provides more detailed grayscale images and offers advantages in detailed imaging and higher magnification capabilities [11].

As mentioned earlier, one of the analyzed samples was prepared by 3D additive manufacturing technique. Our aim was to imitate the microstructure of bone and subsequently analyze this sample using SEM to get more detailed information about its internal structure and microscale and nanoscale pores. These findings will be particularly valuable for guiding further research.

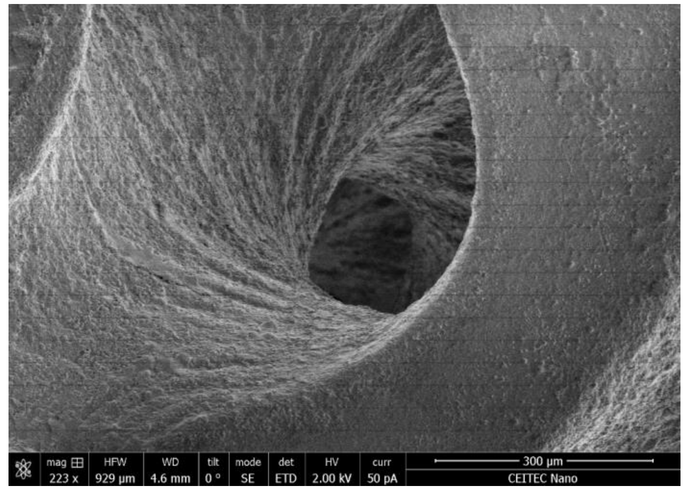



Fig. 2. Porous structure of the 3D printed hydroxyapatite sample.

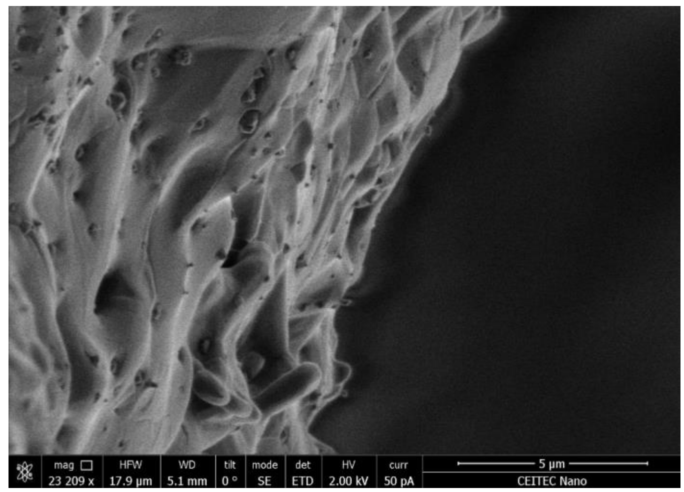



Fig. 3. Rough surface of the sample that could later play a pivotal role in initial attachment of the bone cells.

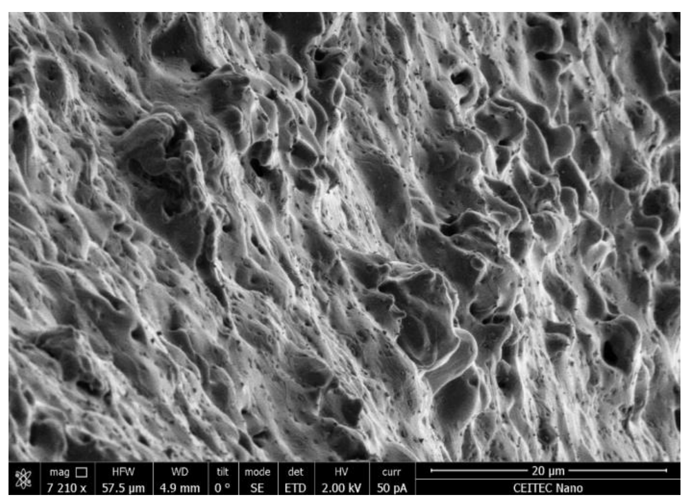



Fig. 4. Surface roughness of the hydroxyapatite sample.

Based on the knowledge we have it is apparent that changes in the surface texture or microtopography of a biomaterial can significantly influence how cells respond to it. For instance, a rougher surface may promote better adhesion and proliferation of cells, while a smoother surface might hinder these processes. Additionally, alternations in surface characteristics can impact the formation of and integration of surrounding tissue, ultimately influencing the overall success of the biomaterial [12].

Further stress tests and other experiments will show whether the sample is suitable for further research.

## IV. Conclusion

Our aim was to show that the interconnection of additive manufacturing, material science and bone tissue engineering offers promising avenues for scaffold development. We highlighted the multifaceted factors influencing the efficacy of biomaterials in bone regeneration and we worked with three different hydroxyapatite samples.

Fourier transform infrared spectroscopy (FTIR) was utilized to analyze the chemical structure of the individual samples, identifying characteristics bands corresponding to phosphate and hydroxyl functional groups. These bands serve as indicators of hydroxyapatite structure and may vary in intensity based on the concentration of the functional groups.

Scanning electron microscopy (SEM) provided detailed imaging of the internal structure and pore characteristics of the 3D printed hydroxyapatite sample, which is essential for further understanding of bone cell attachments and other mechanisms and interactions that could take place on bone-implant interface.

Further examinations of the samples are planned, which will include stress tests and other optimization of their characteristics to ensure their suitability for advanced research and potential clinical applications. Some of the findings will be presented on the student conference EEICT.

## References


[1] Von Euw S, Wang Y, Laurent G, et al. Bone mineral: new insights into its chemical composition. Sci Rep. 2019;9(1):8456. doi:10.1038/s41598-019-44620-6

[2] Šromová V, Sobola D, Kaspar P. A Brief Review of Bone Cell Function and Importance. Cells. 2023;12(21). doi:10.3390/cells12212576

[3] Ielo I, Calabrese G, De Luca G, Conoci S. Recent Advances in Hydroxyapatite-Based Biocomposites for Bone Tissue Regeneration in Orthopedics. Int J Mol Sci. 2022;23(17). doi:10.3390/ijms23179721

[4] Šťastný P. Customized Hybrid Bioscaffolds for Bone Regeneratin. Published online 2020.

[5] Bogala MR. Three-dimensional (3D) printing of hydroxyapatite-based scaffolds: A review. Bioprinting. 2022;28:e00244. doi:https://doi.org/10.1016/j.bprint.2022.e00244

[6] Fadlelmoula A, Pinho D, Carvalho VH, Catarino SO, Minas G. Fourier Transform Infrared (FTIR) Spectroscopy to Analyse Human Blood over the Last 20 Years: A Review towards Lab-on-a-Chip Devices. Micromachines. 2022;13(2). doi:10.3390/mi13020187

[7] Barth A. Infrared spectroscopy of proteins. Biochim Biophys Acta - Bioenerg. 2007;1767(9):1073-1101. doi:https://doi.org/10.1016/j.bbabio.2007.06.004

[8] Kesmez Ö. Preparation of anti-bacterial biocomposite nanofibers fabricated by electrospinning method. J Turkish Chem Soc Sect A Chem. 2019;7:125-142. doi:10.18596/jotcsa.590621

[9] Uysal I, Severcan F, Evis Z. Characterization by Fourier transform infrared spectroscopy of hydroxyapatite co-doped with zinc and fluoride. Ceram Int. 2013;39(7):7727-7733. doi:https://doi.org/10.1016/j.ceramint.2013.03.029

[10] Gheisari H, Karamian E, Abdellahi M. A novel hydroxyapatite – Hardystonite nanocomposite ceramic. Ceram Int. 2015;41(4):5967-5975. doi:https://doi.org/10.1016/j.ceramint.2015.01.033

[11] Abdullah A, Mohammed A. Scanning Electron Microscopy (SEM): A Review. In: ; 2019.

[12] Li L, Crosby K, Sawicki M. Effects of Surface Roughness of Hydroxyapatite on Cell Attachment and Proliferation. J Biotechnol Biomater. 2012;02. doi:10.4172/2155-952X.1000150

# Automating Antibiotic Susceptibility Testing with Machine Learning for Disk Diffusion Test Analysis

Jakub Lepik
*Department of Telecommunications*
*Brno University of Technology*
Brno, Czech Republic
239122@vut.cz

Michal Cicatka
*Department of Telecommunications*
*Brno University of Technology*
Brno, Czech Republic
0000-0002-8119-6679

***Abstract*—Rapid and reliable antibiotic susceptibility testing (AST) methods are imperative in response to the escalating challenges of antimicrobial resistance. This study focuses on enhancing disk diffusion testing, a cornerstone of AST, by integrating machine learning and automation. Leveraging state-of-the-art object detection models, including EfficientDet and Mask R-CNN and image-processing approaches, our methodology addresses the need for standardized evaluation processes across diverse laboratory equipment while enabling the integration of mobile devices into the workflow, democratizing AST, and enhancing its accessibility. We utilize a comprehensive disk diffusion dataset for object detection models captured by devices like mobile phones and professional solutions. Additionally, our experiments lay the groundwork for a web application adopting a device-agnostic approach, promising improved accessibility and efficiency in AST analysis.**

***Index Terms*—antibiotic sensitivity testing, disk diffusion test, machine learning, image processing**

## I. Introduction

In our pursuit of understanding microbial responses to antibiotics, we turn to the power of algorithms and machine learning (ML). Evaluating antibiotic effectiveness against microbial cultures is a crucial aspect of modern medical research, particularly with the rise of antibiotic-resistant bacteria, which presents a significant challenge to global health. Antibiotic susceptibility testing (AST) remains a cornerstone in combating antimicrobial resistance, offering essential guidance for antibiotic prescription.

The need to develop rapid AST methods has become increasingly apparent considering recent advances, such as those explored in [1]. Automation methods and ML, in particular, have seen substantial progress across multiple domains, including clinical diagnostics and pharmaceutical research. However, their utilization in pivotal disk diffusion tests, unlike other methods of AST, remains relatively unexplored [2]. Further research and development in this domain are crucial for enhancing and complementing AST methodologies and effectively combating antibiotic resistance.

By integrating ML and other automation approaches with disk diffusion testing methodologies, we aim to enhance the accuracy, efficiency, and accessibility of this critical part of AST. The current reliance on highly qualified personnel poses significant challenges, as expertise is often scarce and evaluation results can be inconsistent. We seek to standardize the evaluation process across various laboratory equipment, from data captured with expensive professional-grade setups and equipment to phone-captured data, which is significantly less expensive as mobile phones are ubiquitous nowadays. This democratizes the process and enables more laboratories to automate while our methodology ensures reliability and consistency amongst such diverse conditions.

In earlier research, such as in [3], a semi-automated tool for analyzing images from disk diffusion tests was introduced. While demonstrating an agreement of 87 % with expert microbiologists, it focuses solely on using expensive professional-grade laboratory equipment. Similarly, in the studies conducted by [4] and [5], fully automated disk diffusion method evaluation systems were proposed. Despite their success, these methods lack adaptability to resource-limited settings. In 2021, an AI-based mobile application tailored for AST analysis was published, targeting resource-limited environments [6]. While achieving an overall accuracy of 90 %, this application operates exclusively on mobile platforms.

Conversely, we believe that restricting the solution to a single platform limits its overall accessibility and usability. Therefore, we aim to develop our own using a multi-platform approach, allowing for better accessibility of automated AST analysis across different acquisition devices and settings and enabling more laboratories to implement it.

Our methodology relies on an extensive dataset provided by Bruker Daltonics GmbH & Co. KG, comprising over 3200 images captured with different devices, including mobile phones and professional solutions such as the MBT Pathfinder®, colony-picking robot, developed by Bruker Daltonics GmbH & Co. KG. The dataset offers images capturing the disk diffusion approach to AST, specifically images of Petri dishes with agar medium, antibiotic sample disks, and cultured microbes, all ready for evaluation.

In the first experiment of this study, we utilize state-of-the-art object detection models such as EfficientDet and Mask R-CNN to identify antibiotic disks within microbial cultures. In the second experiment, we explore the assessment of inhibition zones of antibiotic samples using an image processing approach. Both experiments provide a foundation for integrating our methodology into a cross-platform web application that will adopt a device-agnostic approach.

## II. Materials and methods

### A. Object Detection Experiment

Our approach primarily adopts an ML framework for its capacity to generalize effectively across diverse scenarios. That is crucial, especially since our goal is to utilize both mobile and professional platforms. Devices like mobile phones frequently generate lower-quality images, particularly due to overall poorer acquisition conditions. As mentioned earlier, the foundation of our methodology lies in the dataset provided by Bruker Daltonics GmbH & Co. KG, which includes more than 3200 high-resolution disk diffusion test images. This dataset consists of images with varying proportions taken from different capturing angles under various lighting and other acquisition conditions. Pairing the diversity of our dataset with the strength of ML empowers us to achieve high-precision detection across these greatly varying acquisition conditions met in real-world laboratory and clinical settings.

For the purpose of training the object detection models and further evaluation, each of the images in the dataset was manually annotated. This process involves labeling the positions of antibiotic sample disks by bounding boxes and then translating them to coordinates. Another step is measuring the radii of inhibition zones generated by the antibiotic samples in pixels. The output of this process can be seen in the Fig. 1. Although we avoid manual solutions by facilitating ML, this labor-intensive process remains essential. It significantly pays off as it contributes to the superior generalization capability of ML solutions compared to hard-coded approaches.

Fig. 1 further illustrates the conditions and challenges our approach must address by presenting ideally inferred example images. The top row showcases images captured by the aforementioned MBT Pathfinder® professional solution, while the bottom row features images captured with mobile phones. This section focuses on detecting sample disks highlighted by green bounding boxes. Antibiotic disks typically appear in oval shapes, as depicted in the figure. However, their overall appearance and color vary significantly across different acquisition settings, particularly in mobile-captured images, where the capturing conditions can often be inferior.

To achieve accurate detection, which is the foundation for our further experiments, we utilize TensorFlow framework [7] and pre-trained models from the TensorFlow Model Detection Zoo, such as EfficientDet D0 and D2 from the EfficientDet family [8], alongside Mask R-CNN with Inception Resnet v2 backbone [9] (referred to as just Mask R-CNN). Working with pre-trained models offers clear advantages over developing them from scratch. Trained on diverse datasets, they capture key features for object detection, improving performance, robustness, and resource efficiency.

We utilize transfer learning to adapt these pre-trained models, which is computationally intensive, especially for high-precision large image models. Training may span several days, necessitating access to powerful computing resources. To accelerate this, we leverage the computational capabilities of Bruker Daltonics GmbH & Co. KG server infrastructure equipped with NVIDIA RTX 4090 graphic cards.

### B. Antibiotic Effectivity Assessment Experiment

Another part of our methodology incorporates an algorithm dedicated to assessing the effectiveness of antibiotics on microbial cultures. Its primary objective is determining the radii of circular inhibition zones surrounding antibiotic sample disks. It integrates multiple image processing techniques, including methods from the OpenCV library [10]. The desired outcomes of this experiment are further illustrated in Fig. 1, where inhibition zones, demarcated by a ground truth circle, are highlighted in cyan alongside their corresponding ground truth radii measured in pixels.

The effectiveness assessment process involves several image preprocessing steps. We employ a denoising technique and a Gaussian filter to refine the images for further processing.

As hinted in section II-A, this stage heavily depends on precise antibiotic sample disk detection, as our algorithm chooses several seed positions surrounding the closest proximity of each antibiotic disk. The average pixel brightness differences for each color channel of the given seeds are calculated. Moreover, a coefficient, determined based on the overall brightness of the image with less impact on the images with the lowest and highest values, is applied to the brightness differences. Subsequently, every seed and its adjusted brightness differences are sequentially fed into separate OpenCV flood-fill algorithms. The output of this process resembles a joint flood-filled area converted to a binary mask.

After employing the abovementioned methods, supplementary algorithms are essential to interpret flood-filled areas and convert them into specific radial measurements. To address this challenge, we utilize a morphological ellipsoid filter on the final joint flood-filled mask. This ensures that small holes or gaps are closed and the mask is smoothed before being put into our custom adaptive circle-enlarging algorithm.

The adaptive circle-enlarging algorithm scans the binary mask, gradually expanding itself and the inhibition zone circle until reaching a brightness deviation threshold. Suppose this threshold is not met within a set number of iterations. In that case, it suggests the image is fully flood-filled, implying the selected antibiotic sample likely did not generate any inhibition zone, which is vital for preventing false results.

## III. Results

The efficiency of our trained object detection models was assessed using the Mean Average Precision (mAP) metric defined in [11], with a script rewritten in Python specifically developed for [12]. Along with mAP, we also measured the average inference time of each model. (Average time to process one image by the model.) The mAP metric offers a comprehensive evaluation, considering average precision across various Intersections over Union (IoU) thresholds, ranging from 0.5 to 0.95, with gradual increments of 0.05.

Table I evaluates trained object detection models using the abovementioned metrics. Models were tested on our dataset's

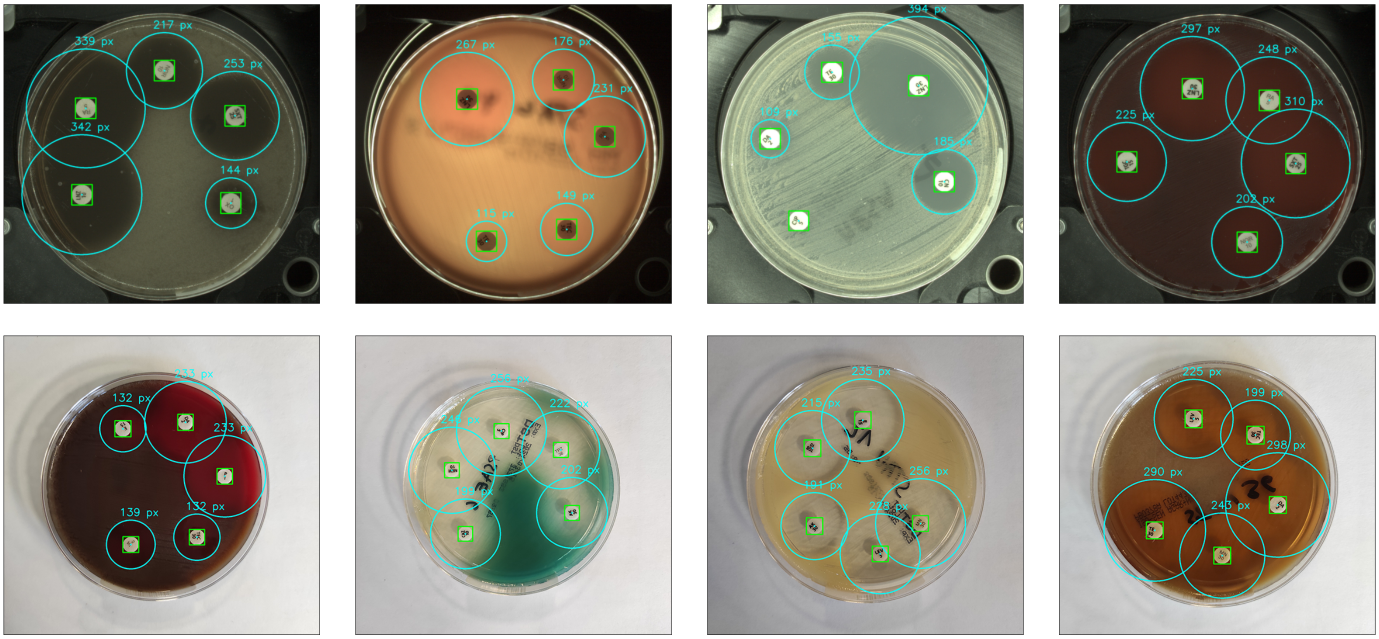


Fig. 1. Examples of ideally processed images: top images taken with professional lab equipment (MBT Pathfinder®), bottom images with mobile phones.

test subset, consisting of 10 % of the total number of images. We assessed each model's performance based on its mAP score and the inference time using a laptop equipped with an AMD Ryzen 5800H CPU.

From Table I, it is apparent that the Mask R-CNN model achieved the highest mAP score of 94.37 %, indicating its superior performance in accurately detecting various antibiotic disks. However, this improved accuracy comes at the cost of increased inference time, with Mask R-CNN requiring 3896 ms per inference on the specified hardware.

TABLE I
COMPARISON OF ANTIBIOTIC SAMPLE DETECTION PERFORMANCE USING DIFFERENT TRAINED MODEL ARCHITECTURES

| Architecture | Inference time [ms] | mAP [%] |
|---|---|---|
| EfficientDet D0 | 237 | 87.43 |
| EfficientDet D2 | 454 | 88.62 |
| Mask R-CNN | 3896 | 94.37 |

Conversely, the EfficientDet models, while exhibiting slightly lower mAP scores compared to Mask R-CNN, showcased significantly faster inference times. EfficientDet D0 and D2 achieved mAP scores of 87.43 %, and 88.62 % respectively, with inference times of 236 ms and 453 ms.

Our dataset images are large, so they must be downsampled before inferencing. We would suspect that models with greater input sizes would yield more precise results. However, in our case, the EfficientDet D2, despite having a greater input size of 768 $\times$ 768 pixels, gains minimal improvements to D0, which operates on 512 $\times$ 512 pixel images.

The radii measuring algorithm was designed to operate on images captured by professional-grade equipment and non-professional solutions such as mobile phones. Despite this, the variability in data quality between images captured by professional-grade equipment and those obtained from mobile phones necessitated a nuanced approach to evaluation. In this part, we present the results based on two distinct datasets: images captured by professional-grade equipment, specifically MBT Pathfinder®, and images obtained from mobile phones. Each dataset contains 60 varying images, with more than 250 antibiotic samples in each of them. As the evaluation of our radii measuring algorithm is completely separate from the evaluation of the object detection experiment, we utilize the ground truth locations of the disks instead.

Given the inherent differences in image capture conditions and data quality between professional-grade equipment and mobile phones, we recognize the need to evaluate algorithm performance separately for each dataset. This allows for accurately assessing the algorithm's capabilities under diverse capture conditions and ensures the results' reliability.

Table II summarizes the comparison of radius detection performance across the two types of datasets. Given the significant variation in the radii of inhibition zones in pixels between datasets captured on mobile devices and MBT Pathfinder®, with average values of 399.78 pixels and 270.74 pixels, respectively, the metrics used are relative to the average ground truth radii sizes in pixels. Specifically, we utilize Relative Mean Absolute Error (RMAE), Relative Root Mean Square Error (RRMSE), and Relative Standard Deviation (RSTD).

As we delve into the results, notable differences in algorithm performance between the two datasets are observed. While inferring the one captured by the professional solution, the algorithm demonstrates fairly decent performance, as evidenced by lower error metrics: RRMAE (12.05 vs. 18.49 %), RRMSE (16.45 vs. 26.29 %), and RSTD (14.13 vs. 25.53 %),

TABLE II
COMPARISON OF RADIUS DETECTION PERFORMANCE ACROSS DATASETS ACQUIRED IN VARYING ACQUISITION CONDITIONS

| | **Acquisition device** | |
|---|---|---|
| **Metric** | **MBT Pathfinder®** | **Mobile** |
| **RMAE [%]** | 12.05 | 18.49 |
| **RRMSE [%]** | 16.45 | 26.29 |
| **RSTD [%]** | 14.13 | 25.53 |

indicating closer alignment between predicted and ground truth inhibition zone sizes. Conversely, the algorithm performs worse on the dataset acquired from mobile phones. Higher error metrics indicate less accurate detection of inhibition zone sizes than on the professional-grade equipment dataset. For example, the RSTD for inferring on the mobile dataset is more than 80 % higher than on the professional-captured dataset (25.53 vs. 14.13 %), indicating a greater degree of inconsistency in the algorithm's performance between each of the predictions. This further highlights the algorithm's struggle to generalize across datasets with more varying capture conditions.

## IV. CONCLUSION

In this study, we endeavored to automate the evaluation of antibiotic effectiveness in the disk diffusion method across a wide range of laboratory equipment. Human evaluation of antibiotic effectiveness requires specialized training and expertise, which may not always be readily available. The standardized approach we propose enhances reliability while enabling laboratory personnel to focus on more critical aspects of their work. With the help of ML and its great generalization capabilities, we capitalize on the widespread availability and ubiquity of image-capturing devices like mobile phones. Utilizing both professional and mobile-capturing devices, we democratize the evaluation process and make it accessible to a broader range of laboratories.

The Object Detection Experiment involved training machine-learning object detection models to detect antibiotic disks. After reviewing the results, the models trained on our dataset exhibit robust performance in detection accuracy. During this process, we also demonstrated that different models vary in computation efficiency and accuracy, emphasizing the importance of selecting the most suitable model for future work. The results of this experiment underscore the effectiveness of the ML approach as the ability of the models to generalize on data of different acquisition conditions and devices is high. While the experiments provided promising results, there is room for improvement. Exploring additional state-of-the-art object detection models could lead to better balances between efficiency and accuracy.

Conversely, while yielding satisfactory results, the Antibiotic Effectivity Assessment Experiment revealed challenges in generalization across mobile phone data. This disparity underscores the limitations of hand-coded algorithms in accommodating varied acquisition conditions. In future experiments, we may explore additional machine-learning approaches to address these challenges and enhance the adaptability and robustness of our methodology. Additionally, the varying conditions under which the photos were taken highlight the importance of standardization in the image acquisition process. Especially for non-professional solutions, improving lighting and stabilization conditions could dramatically enhance the quality and consistency of images.

The next step in our methodology involves integrating the experiments into a valuable application for use in laboratories that could run on a broad range of architectures. This endeavor will provide more insights into the effectiveness of object detection algorithms versus their accuracy and how much resources can be allocated accordingly.

## ACKNOWLEDGMENT

This article was made in cooperation with Bruker Daltonics GmbH & Co. KG. The company generously provided a disk diffusion dataset and other valuable resources, including access to their server. We sincerely appreciate their support and collaboration in this research endeavor.

## REFERENCES

[1] D. Song, Y. Lei, Mini-review: Recent advances in imaging-based rapid antibiotic susceptibility testing, Sensors and Actuators Reports 3 (2021) 100053. doi:10.1016/j.snr.2021.100053.

[2] O. Dauwalder, F. Vandenesch, Disc diffusion AST automation: one of the last pieces missing for full microbiology laboratory automation, Clinical Microbiology and Infection 26 (5) (2020) 539–541. doi:10.1016/j.cmi.2020.01.021.

[3] C. Alonso, C. Domínguez, J. Heras, E. Mata, V. Pascual, C. Torres, M. Zarazaga, Antibiogramj: A tool for analysing images from disk diffusion tests, Computer Methods and Programs in Biomedicine 143 (2017) 159–169. doi:10.1016/j.cmpb.2017.03.010.

[4] M. Hombach, M. Jetter, N. Blöchliger, N. Kolesnik-Goldmann, E. C. Böttger, Fully automated disc diffusion for rapid antibiotic susceptibility test results: a proof-of-principle study, Journal of Antimicrobial Chemotherapy 72 (6) (2017) 1659–1668. doi:10.1093/jac/dkx026.

[5] A. Cherkaoui, G. Renzi, N. Vuilleumier, J. Schrenzel, Performance of Fully Automated Antimicrobial Disk Diffusion Susceptibility Testing Using Copan WASP Colibri Coupled to the Radian In-Line Carousel and Expert System, Journal of Clinical Microbiology 59 (9) (Aug. 2021). doi:10.1128/jcm.00777-21.

[6] M. Pascucci, G. Royer, J. Adamek, M. A. Asmar, D. Aristizabal, L. Blanche, A. Bezzarga, G. Boniface-Chang, A. Brunner, C. Curel, G. Dulac-Arnold, R. M. Fakhri, N. Malou, C. Nordon, V. Runge, F. Samson, E. Sebastian, D. Soukieh, J.-P. Vert, C. Ambroise, M.-A. Madoui, AI-based mobile application to fight antibiotic resistance, Nature Communications 12 (1) (Feb. 2021). doi:10.1038/s41467-021-21187-3.

[7] T. Developers, TensorFlow (2024). doi:10.5281/ZENODO.10798587.

[8] M. Tan, R. Pang, Q. V. Le, EfficientDet: Scalable and Efficient Object Detection (2019). doi:10.48550/ARXIV.1911.09070.

[9] K. He, G. Gkioxari, P. Dollár, R. Girshick, Mask R-CNN (2017). doi:10.48550/ARXIV.1703.06870.

[10] G. Bradski, The OpenCV Library, Dr. Dobb's Journal of Software Tools (2000).

[11] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, A. Zisserman, PASCAL VOC2012 Challenge Results, http://www.pascal-network.org/challenges/VOC/voc2012/workshop/index.html.

[12] J. Cartucho, R. Ventura, M. Veloso, Robust Object Recognition Through Symbiotic Deep Learning In Mobile Robots, in: 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2018, pp. 2336–2341.

# NANOBLAST: Python Tool for Raw Nanopore Signal Processing

1st Martin Suriak
*Department of Biomedical Engineering*
*Faculty of Electrical Engineering and Communication*
Brno University of Technology, Brno, Czech Republic
240566@vut.cz

2nd Markéta Nykrýnová
*Department of Biomedical Engineering*
*Faculty of Electrical Engineering and Communication*
Brno University of Technology, Brno, Czech Republic
nykrynova@vut.cz

***Abstract*—Oxford Nanopore Technologies' sequencers enable direct real-time DNA/RNA sequencing. While numerous tools aid in analyzing the nanopore output data, offering functions such as visualizing raw signals with highlighted nucleotide positions, none provide a complete solution for exporting analyzed data into a clear, comprehensive file. In response, a Python tool has been developed to streamline various tasks. This includes searching for specific nucleotide sequences using BLAST, plotting raw signals with detected nucleotide bases, and generating a comprehensive file containing all essential information. The tool integrates components for handling raw nanopore data, extracting crucial information from the basecalling process using SAM file handlers, and utilizing a BLAST search engine. Employing a comprehensive algorithm, it can handle both the old FAST5 and the novel POD5 formats, enabling the identification of any nucleotide sequence and its corresponding signal.**

***Index Terms*—squiggle, sequencing, signal processing**

## I. Introduction

Sequencers from Oxford Nanopore Technologies (ONT) enable direct, real-time DNA/RNA sequencing by operating at the level of individual molecules, thus optimizing data library preparation processes. These sequencers employ nanopores for sequence detection, utilizing a molecular motor to transport molecules through the pore. The passage of nucleotides through the pore modifies the measured electric current, generating a signal termed a 'squiggle', representing the electric ion current over time. One of the significant advantages of ONT sequencing is the generation of reads up to 2.3 MB in length with sequencing accuracy of up to 95% for the R9.5 version using the $1D^2$ chemistry [1]. Additionally, the system allows for real-time sequencing, enabling the suspension of the process after reading a targeted region. Data are continuously transmitted to a connected computer throughout the sequencing process and stored in FAST5 (support ended in the first quarter of 2023) or POD5 (current, up-to-date format) files for subsequent analysis.

FAST5 files comprise organized groups and datasets, often containing additional information in the form of metadata encapsulated as attributes. Each file contains numerous reads placed within distinct groups. Within these groups, subgroups are structured, including '*Raw*' (containing the raw signal from sequencing as a dataset), '*channel_id*' (providing data for signal conversion to picoamperes), '*context_tags*', and '*tracking_id*' (both containing general details regarding the sequencing run) as depicted in Fig. 1. [2]


Fig. 1. Hierarchy scheme of a FAST5 file containing 4 reads, the first read showing the groups "Raw", "channel_id", "context_tags" and "tracking_id"

Due to the continuous increase in the volume of output data from nanopore sequencing, a concerted effort has been made to improve and optimize the process of writing and reading information. In 2023, ONT introduced a novel data format, POD5, intending to replace the older, less efficient FAST5 format completely. POD5 represents a collection of Apache Arrow tables stored within the Apache Feather 2 format and packaged within a container format bearing the extension .pod5. The tables *'Reads'*, *'Signal'*, and *'Run Info'* contain all data gathered from the sequencing procedure. [3]

A critical component of nanopore sequencing is basecalling, a time-consuming process wherein raw squiggle data are translated into a nucleotide sequence. Historically, the predominant basecalling tool has been Guppy [4]. However, with the introduction of the innovative POD5 format, there was a demand for a new basecaller. In 2023, Dorado [5] was introduced as a new basecaller optimized explicitly for the structure of the POD5 format, offering expedited processing times and enhanced accuracy. The default output from basecalling is a BAM file, a binary version of the commonly used SAM format for storing various data, including the obtained nucleotide sequence, read length, sequencing initiation point, move table, etc. This process often introduces some inaccuracies, where the accuracy of Guppy basecaller was 90-96% for high-accuracy basecalling and 85–92% for fast basecalling [6]. In comparison, the novel basecaller Dorado shows mean accuracy of 91% for the fast sequencing mode and 97.1% for the most accurate one. [7]

Despite the abundance of existing analytical tools offering a wide array of procedures for data analysis, their utility is threatened by the arrival of the POD5 format, rendering them obsolete. Consequently, in response to this challenge, a Python-based tool has been developed to handle both the old FAST5 and the novel POD5 formats, ensuring data handling capabilities across platforms. This tool facilitates tasks ranging from searching a desired nucleotide sequence using BLAST [8] to plotting squiggles with detected nucleotides, finally generating a comprehensive file containing all necessary information.

## II. Methods

### A. Dataset

Sequencing was performed using the ONT MinION platform. Initially, the old chemistry R9.4.1 was utilized; thus, sequencing data in FAST5 files were obtained. The second chemistry used was the newer R10.4.1, which, thanks to software updates, produces POD5 files. Subsequently, it was necessary to obtain a nucleotide sequence using the basecalling process. For FAST5 files, the Guppy (v6.5.7) tool was used to generate multiple BAM files, which needed to be merged into one and then reformatted into SAM. The POD5 files were basecalled using the new Dorado (v4.2.0) tool. One of the notable advantages of the new basecaller is its capability to generate a singular, exhaustive SAM file, if the option is activated, thereby skipping the intermediate steps.

### B. Handling of the raw data

The primary objective of this section is to retrieve the signal data from a read effectively. Two distinct methodologies are employed, given the utilization of two different formats. Initially, the FAST5 handler utilizes the h5py library, specially designed to handle HDF5-like formats such as FAST5. It opens a file, extracts the desired signal and sampling rate for subsequent analysis, and employs the attributes from the group *channel_id* to convert the signal values to picoamperes. The formula for this conversion is as follows:

$$signal\_in\_pA = (raw\_signal + offset) \times \frac{range}{digitisation} \tag{1}$$

where *raw_signal* contains the signal values, *offset* sets the ADC offset error, *range* indicates the full-scale measurement range, and *digitisation* establishes the number of quantization levels in the Analog to Digital Converter.

The procedure remains largely analogous when dealing with POD5 files. However, notable distinctions arise in utilizing the pod5 library, reflecting the fundamentally different architecture of POD5. Unlike FAST5, POD5 files inherently contain signal values in picoamperes, which creates an opportunity to skip the conversion step.

Therefore, the presented tool demonstrates the ability to seamlessly read any of the files mentioned above formats and then extract the necessary data, such as signal values and sampling rates, to visualize the analyzed readings.

### C. Plotting the signal

The SAM files produced from basecalling differ from standard SAM files because they lack alignments to other sequences. Consequently, they do not include the ”@SQ” line in their header, rendering libraries like Samtools unsuitable for their handling. Instead, these files must be parsed and searched as if they were plain text files.

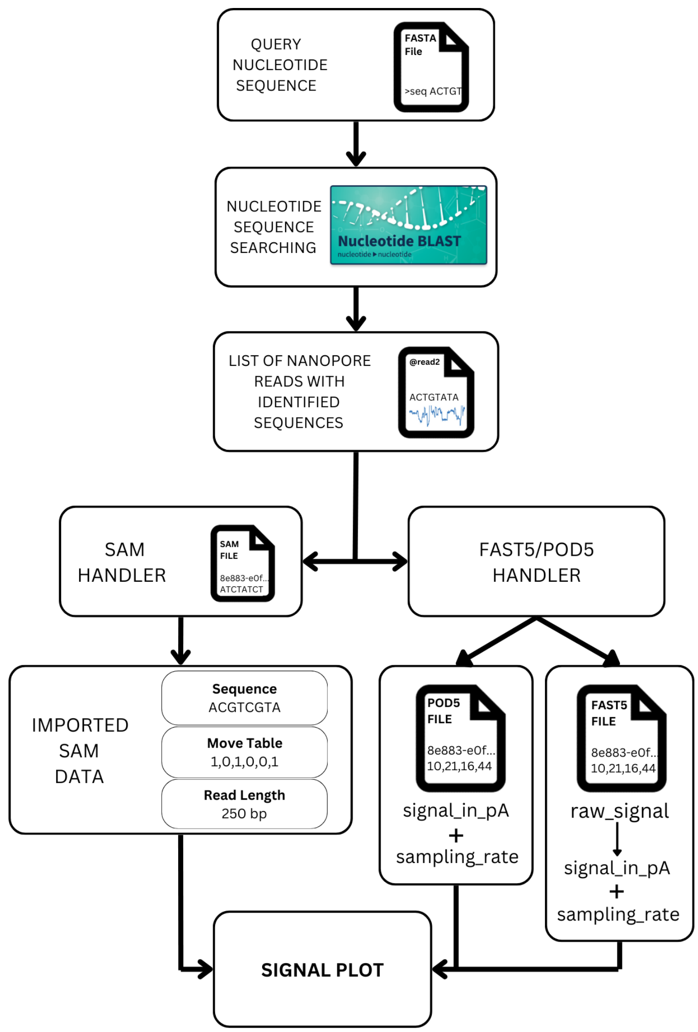



Fig. 2. Development scheme of the plotting function

The SAM handler initially locates the required read in the SAM file based on its ID. Subsequently, this line is segmented into fields separated by tabs, enabling the extraction of necessary information. Certain information consistently occupies specific positions in SAM files, facilitating the extraction of data such as *nucleotide sequence, read length, move table, and number of trimmed samples* from the start of sequencing. The *move table* comprises a series of ones and zeros, where the number of ones corresponds to the length of the basecalled sequence. Each move table begins with a special value termed as *stride*, denoting the down-sampling factor employed in the neural network. A number one indicates the detection of a nucleotide in the signal, while zeros indicate the distance between consecutive nucleotides. The fundamental formula for this mathematical process resembles the following:

$$signal\_point = (number\_of\_zeroes + 1) \times stride \quad (2)$$

The variable *number_of_zeroes* indicates the count of consecutive zeroes (0) identified in the move table, thereby providing insight into the width of the segment and the distance of each nucleotide from the next one.

Once a list of moves is generated, a cumulative sum is computed. This segmentation step is pivotal for aligning the original signal with the identified nucleotides. These segments are visually distinguished based on their nucleotide assignment: A is represented in red, T in blue, C in yellow, and G in green. When ambiguous nucleotides like Y, N, or W are present, the corresponding segment is coloured grey. The entire data acquisition process and plotting function is depicted in Fig. 2.

### D. Searching for a specific nucleotide sequence

One of this tool's primary functionalities is its capability to search for a specified nucleotide sequence within a SAM file generated through basecalling. BLAST executes this task utilizing the API from the Biopython package. Since BLAST's input format is FASTA, the SAM file must undergo a conversion step integrated into the tool's architecture.

Subsequently, a BLAST database is created, encompassing all reads from a sequencing run. Following the execution of a search, the results can be refined according to user specifications and presented in tabular format. The identified reads are then cross-referenced with the raw signal data files, enabling the extraction of signal values corresponding to the searched sequence.

The resulting product is a table generated in PARQUET format, encapsulating the discovered reads alongside their corresponding signal values. This output can subsequently be visualized utilizing other functionalities offered by the tool.

## III. Results

The Graphical User Interface (GUI) is a core part of the tool and offers quick access to all the information needed for its efficient operation and service delivery as illustrated in Fig. 3 and 4. Considerable effort was given on developing a tool that not only prioritizes functionality but also takes user-friendliness into account. The GUI is divided into two parts: the initial part facilitates BLAST searches and creates a PARQUET file containing the identified reads together with their associated signals. The subsequent part allows users to plot desired signals.

The most important ability of this tool is plotting a signal of any analyzed read. The nucleotide bases are highlighted,

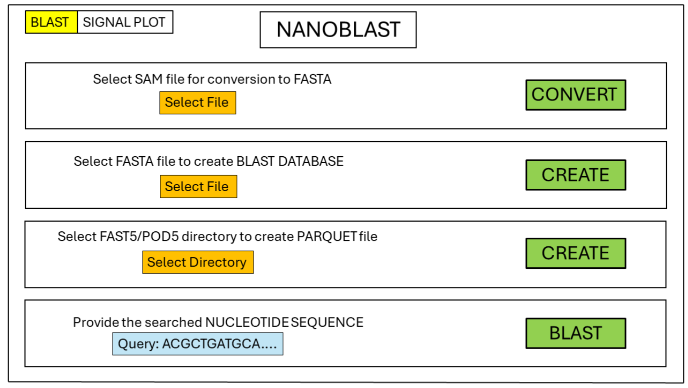



Fig. 3. The first part of the GUI contains fields for a user to provide the necessary information and also provides a user with the tool's various functionalities

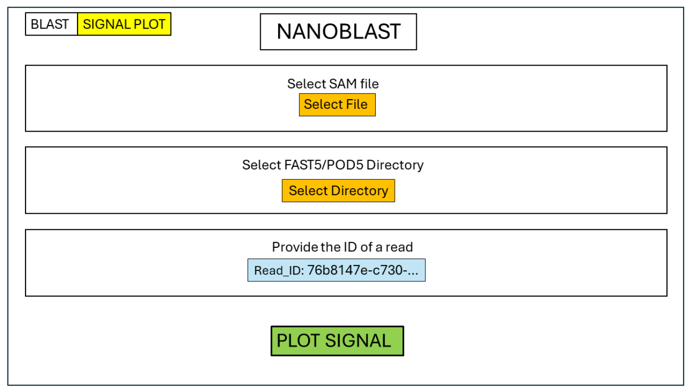



Fig. 4. The second part of the GUI provides the ability to create a squiggle of an analysed read with highlighted nucleotide positions

resulting in a self-explanatory image. The signal values are depicted in picoamperes, and the axis is transformed to show time in milliseconds. Fig. 5 illustrates the resultant output derived from three distinct reads of the same nucleotide sequence. Evidently, despite the sequences being identical, discernible variations are observed within the signals.

## IV. Conclusion

In conclusion, the development of this Python tool represents a significant simplification of data analysis, particularly in the context of Oxford Nanopore Technologies' sequencing technology. This tool seamlessly integrates various functionalities, including the conversion of SAM files into FASTA format if necessary, searching for specific nucleotide sequences using BLAST, plotting raw signals with detected nucleotide bases, and generating comprehensive PARQUET files containing the discovered reads alongside their corresponding signal values. Importantly, its versatility includes compatibility with both legacy (FAST5) and modern (POD5) sequencing data formats.

Furthermore, developing an intuitive graphical user interface enhances usability, providing researchers with efficient access to critical data and analysis tools. By empowering users to interact with sequencing data in real-time and facilitating

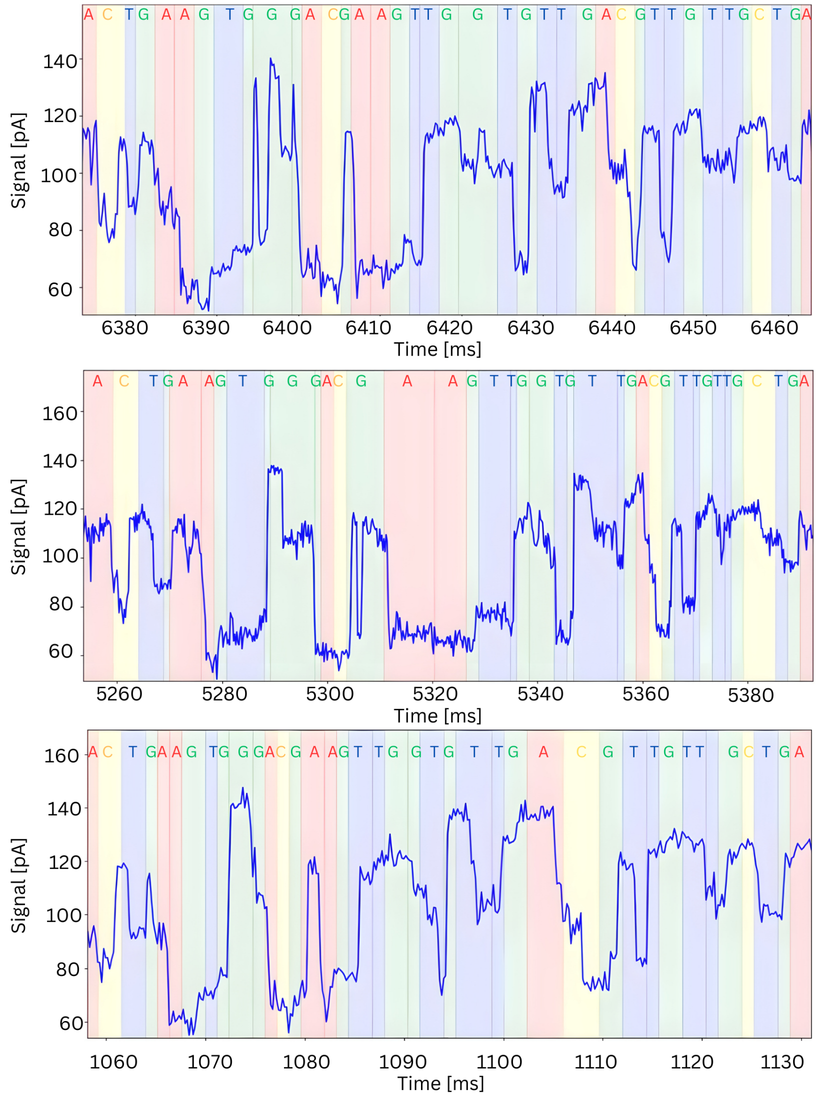



Fig. 5. Identical nucleotide sequences of length 41 bp from different reads searched by BLAST. Signals are not aligned, as the basecalled sequence is not equidistant per each base (colors shown represent nucleotide bases: A - red, T - blue, C - yellow, G - green)

rapid identification of target sequences, this tool represents a significant step towards expediting the analysis process and improving overall accuracy.

In essence, this Python tool not only addresses the challenges associated with traditional basecalling methods but also opens new possibilities for more efficient and accurate analysis of nanopore sequencing data. As the field of DNA/RNA sequencing continues to evolve, tools like this will play a crucial role in unlocking the full potential of nanopore sequencing technology.

## Acknowledgment

This work was supported by a grant project from the Czech Science Foundation [GA23-05845S].

## References


[1] WANG, Yunhao; ZHAO, Yue; BOLLAS, Audrey; WANG, Yuru and AU, Kin Fai, 2021. Nanopore sequencing technology, bioinformatics and applications. Online. *Nature Biotechnology*. Vol. 39, no. 11, pp. 1348-1365. ISSN 1546-1696. Available from: https://doi.org/10.1038/s41587-021-01108-x.

[2] *Supplementary Note 1. FAST5 format demystified*, 2022. Online. Available from: https://hasindu2008.github.io/slow5specs/fast5_demystified.pdf.

[3] *POD5 File Format Design Details*, 2023. Online. Available from: https://github.com/nanoporetech/pod5-file-format/blob/master/docs/DESIGN.md.

[4] *Basecalling using Guppy*, 2018. Online. Available from: https://timkahlke.github.io/LongRead_tutorials/BS_G.html

[5] *Dorado*, 2023. Online. Available from: https://github.com/nanoporetech/dorado.

[6] PEREŠÍNI, Peter; BOŽA, Vladimír; BREJOVÁ, Broňa and VINAŘ, Tomáš, 2021. Nanopore basecalling on the edge. Online. *Bioinformatics*. Vol. 37, no. 24, pp. 4661–4667. Available from: https://doi.org/10.1093/bioinformatics/btab528.

[7] *ONT-only accuracy: 5 kHz and Dorado*, 2023. Online. Available from: https://rrwick.github.io/2023/10/24/ont-only-accuracy-update.html

[8] KORF, Ian; YANDELL, Mark a BEDELL, Joseph, 2003. *BLAST*. O'Reilly Media. ISBN 9780596002992.

# Py/ExplorReg: Exploration of Transcriptome for Potential Regulon Structure Detection

1st Patrícia Janigová
*Department of Biomedical Engineering, Faculty of Electrical Engineering and Communication, Brno University of Technology*
Brno, Czech Republic
231027@vut.cz

2nd Wolfram Weckwerth
*Molecular Systems Biology (MOSYS), University of Vienna*
Vienna, Austria
*Vienna Metabolomics Center (VIME), University of Vienna*
Vienna, Austria
wolfram.weckwerth@univie.ac.at

3rd Jana Schwarzerová
*Department of Biomedical Engineering, Faculty of Electrical Engineering and Communication, Brno University of Technology*
Brno, Czech Republic
*Department of Molecular and Clinical Pathology and Medical Genetics, University Hospital Ostrava*
Ostrava, Czech Republic
*Molecular System Biology (MOSYS), University of Vienna*
Vienna, Austria
Jana.Schwarzerova@vut.cz

***Abstract*—This study introduces Py/ExplorReg, a tool designed for exploring transcriptomic landscapes and identifying potential regulon structures, with a particular focus on its utility in the context of *Arabidopsis thaliana* research. Developed using Python, Py/ExplorReg demonstrates its effectiveness in identifying potential regulons through a pioneering approach rooted in the analysis of gene expression data. Leveraging publicly available datasets from the EMBL-EBI project PRJNA779072, this case study highlights its adaptability to *Arabidopsis thaliana* datasets.**

***Keywords* — *Arabidopsis thaliana*, Gene expression analysis, Regulon, Transcription units**

## I. Introduction

Understanding the intricate regulatory mechanisms governing gene expression is fundamental to deciphering the complexities of biological systems [1]. Regulatory networks, composed of interconnected genes and their regulatory elements, orchestrate precise control over cellular processes in response to internal and external cues. However, accurately detecting the regulon across the organism is a pivotal step for the precise derivation of regulatory networks [2].

In the past, comparable tools were predominantly utilized for identifying operons in bacterial systems [3], [4]. However, in this case study, our aim is to repurpose a similar methodology by recalibrating its parameters. Our objective is to apply this adapted approach to analyze regulatory structures within eukaryotic organisms, including plants. This endeavor aims to pave the way for novel avenues within ecological engineering.

Py/ExplorReg represents a novel approach to interrogating gene expression data, particularly focusing on its applicability within the context of *Arabidopsis thaliana* research. By harnessing methodologies such as correlation coefficients in gene expression analysis [5], Py/ExplorReg pioneers a new frontier in the investigation of gene regulatory networks from an in-silico perspective.

Py/ExplorReg, implemented in Python, is capable of analyzing transcriptomic data to elucidate regulatory relationships within *Arabidopsis thaliana* and beyond.

## II. Dataset

This study, centered on *Arabidopsis thaliana*, employed the Columbia (Col-0) ecotype and specific mutants sourced from the Arabidopsis Biological Resource Center. All specimens were cultivated under controlled conditions to ensure data reliability.

By leveraging publicly available datasets, such as those from the EMBL-EBI project PRJNA779072 [6], Py/ExplorReg demonstrates its adaptability and efficacy in uncovering regulon structures. This introduction provides an overview of Py/ExplorReg's significance in advancing our understanding of gene regulation and highlights its potential contributions to molecular biology research.

Various genomic and epigenomic techniques, including Whole-genome bisulfite sequencing (BS-seq), Transcription Start Site sequencing (TSS-seq), and Chromatin Immunoprecipitation sequencing (ChIP-seq), are available to unveil the dynamics of gene expression regulation and DNA methylation patterns [6]. Subsequently, our analysis primarily focused on gene expression data derived from ChIP-Seq and RNA-Seq (ID GSE188493) [6] as the used dataset for this study.

Different phenotypic expressions in various *Arabidopsis thaliana* mutants may arise from different regulons influencing plant traits. This underscores the significance of regulons in shaping plant morphology and adaptive characteristics. Fig. 1 highlights morphological differences in *Arabidopsis thaliana* resulting from distinct regulon structures. By comparing variants such as hira-1, asf1a-1, and asf1b-1 with the Col-0 control, distinct phenotypic expressions were observed [6].

These differences are crucial for understanding how regulons modulate specific plant features and adaptations.

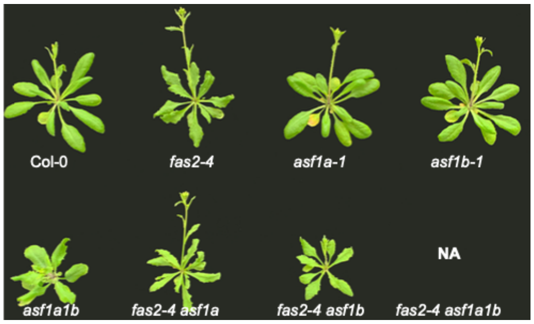



Fig. 1. Morphology of Col-0, hira-1, asf1a-1, asf1b-1, asf1a1b, hira-1 asf1a-1, hira-1 asf1b-1, and hira-1 asf1a1b. The figure was taken over from [6].

## III. Methods

The identification and analysis of regulon structures within the transcriptome of *Arabidopsis thaliana* constitute the primary focus of Py/ExplorReg. The potential regulons, defined as functional clusters of genes regulated under a shared promoter region and controlled by specific transcription factors, play a pivotal role in understanding gene regulatory networks [7].

Regulatory networks are essential for the organism's response to various environmental cues and developmental processes. Unlike operons, which are characterized by their continuous genomic location, regulons are not confined to a specific definitive order and may be distributed in different regions throughout the genome, Fig. 2 [8].

This dispersal allows for a coordinated response to external stimuli, reflecting the complexity of gene regulation in eukaryotic organisms [2]. This regulatory feature permits genes within a regulon to be coordinately activated or repressed, even when dispersed throughout the genome.

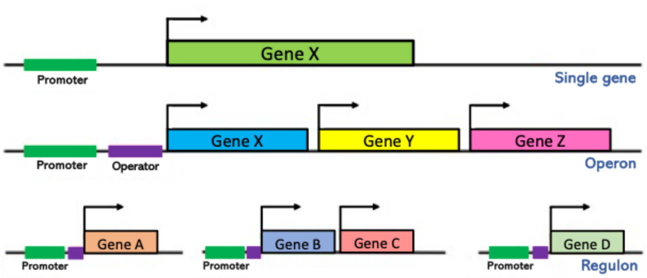



Fig. 2. Illustration of three types of regulatory structures: Single Gene (top), Operon (middle), and Regulon (bottom). In the Single Gene structure, a gene is controlled by its own promoter. The Operon depicted includes genes X, Y, and Z, which share a common promoter and operator, indicative of co-regulated gene expression. The Regulon consists of genes A, B, C, and D, each with individual promoters but collectively regulated, exemplifying the concept of a regulon.

## IV. Py/ExplorReg: Exploration of Transcriptome for Potential Regulon Structure Detection

Py/ExplorReg presents a computational case study developed to analyze gene expression data with the objective of identifying potential regulons — groups of co-expressed genes confirmed post-hoc by literature and regulon databases such as [9]. Utilizing Python packages such as Pandas [10], NumPy [11], SciPy [12], and Scikit-learn [13], Py/ExplorReg navigates the complex landscape of gene regulation by combining Pearson's correlation coefficient [5] and mutual information [14], followed by visualization using a heatmap and cluster dendrogram.

The correlation coefficient calculates the linear dependency between gene expressions, identifying genes that co-vary under various conditions. This measure is crucial for identifying genes with synchronized expression changes, suggesting they might belong to the same regulon. Meanwhile, mutual information extends this analysis to capture non-linear relationships, revealing deeper layers of gene regulation and providing insights into more complex gene interactions that correlation alone might miss.

Py/ExplorReg combines these two metrics to create an intersection matrix that normalizes and integrates the calculated correlation coefficients and mutual information scores. By applying a threshold (set at 0.8 for this study), the tool identifies the most correlated or dependent gene pairs, suggesting a high likelihood of belonging to the same potential regulon. The final step involves visualizing these relationships through a binary image representation, where values above the threshold are marked, highlighting the identified potential regulons.

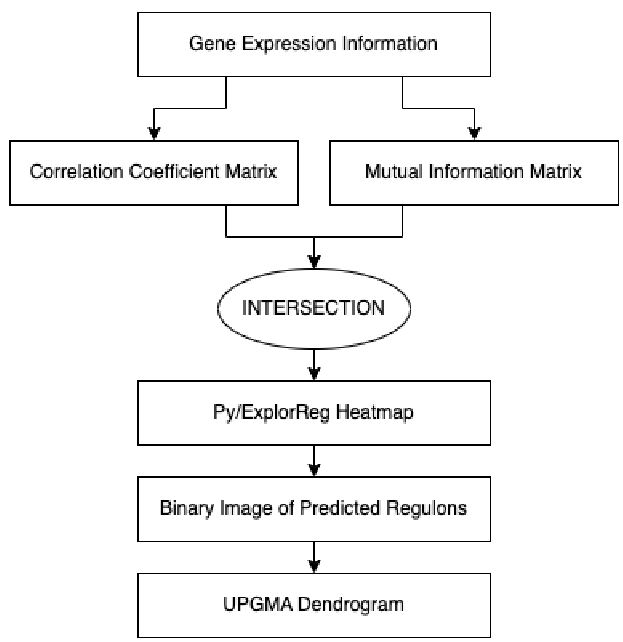



Fig. 3. Flowchart illustrating the workflow of Py/ExplorReg, such as a computational analysis for transcriptome analysis and potential regulon identification.

## V. Results and Discussion

Computational exploration using Py/ExplorReg identified a strong correlation between specific genes in *Arabidopsis thaliana*, revealing key insights into their genetic coordination. Specifically, the algorithm revealed that the strongest linkage was notably observed between AT4G28520 and AT4G27150. This correlation is slightly less significant when involving AT4G25140, indicating a nuanced hierarchy of interaction within the potential regulon. These gene relationships were illustrated in a detailed heatmap, Fig. 4, providing a compelling graphical representation of their interconnectedness.

The heatmap visualizes the intersection of correlation coefficients and mutual information, ranging from 0 to 1. Values close to 1 indicate a strong linear correlation or dependency, suggesting significant regulation between genes. Importantly, the heatmap's diagonal always shows values of 1, as it represents each gene's correlation with itself, demonstrating the heatmap's utility in depicting complex gene interactions within biological systems. This concise visualization facilitates understanding of the intricate co-expressed networks influencing gene expression in *Arabidopsis thaliana*, serving as a preliminary step before exploring regulatory gene networks, which can be confirmed based on literature information.

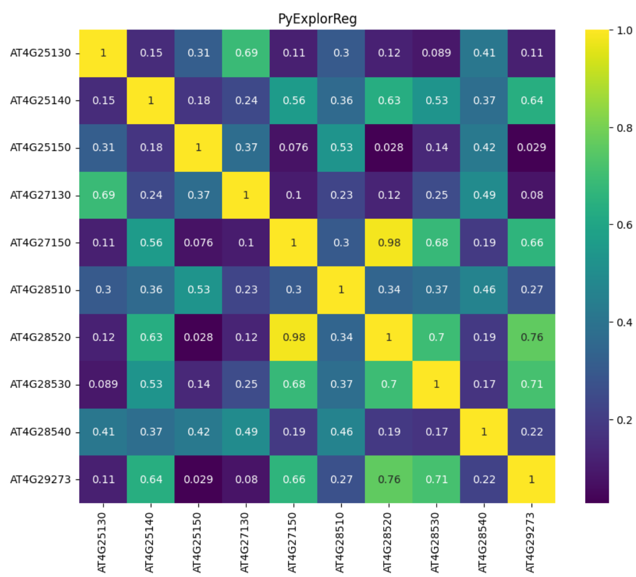



Fig. 4. Visualization of correlation coefficients and mutual information scores in gene expression analysis using Py/ExplorReg – a heatmap representation.

After the heatmap visualization highlighted the intricate dependencies among genes AT4G28520, AT4G27150, and AT4G25140, this complex interaction is further discernible in the dendrogram, Fig. 5, generated through the Unweighted Pair Group Method with Arithmetic Mean (UPGMA) algorithm [15]. UPGMA is a hierarchical clustering method that builds a phylogenetic tree by sequentially merging pairs of clusters based on similarity, determined by the average distance between elements in each cluster. Within the dendrogram, these genes form a distinct cluster, illustrating their close regulatory relationship and mutual dependency.

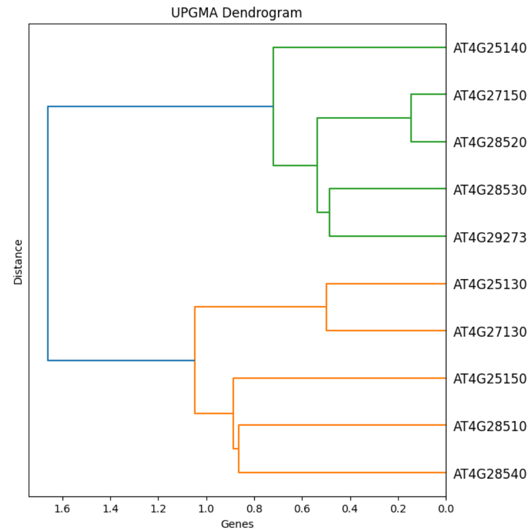



Fig. 5. Visualization of gene regulatory relationships using UPGMA dendrogram: analysis of gene dependency among AT4G28520, AT4G27150, and AT4G25140.

This finding is supported by insights from research on the ABI3 regulon [9] and validates Py/ExplorReg's capability in identifying complex gene regulatory networks. The ABI3 regulon, a pivotal factor influencing these genes, plays a central role in seed development and maturation processes. ABI3 is known for its comprehensive involvement in regulating genes critical for seed storage, desiccation tolerance, and embryo development.

## VI. Conclusion

In conclusion, our case study has introduced Py/ExplorReg for exploring gene expression and regulation. By incorporating both Pearson's correlation coefficient and mutual information, Py/ExplorReg offers a unique approach to uncovering potential regulons, providing a deeper understanding of gene regulatory networks. Unlike traditional methods that rely solely on database searches, Py/ExplorReg integrates gene expression data, thereby enhancing the probability of regulon prediction.

Through testing on meaningful datasets and implementation in Python, Py/ExplorReg sets a new standard in algorithmic analysis for potential regulon structures. It can explore and identify a significant number of co-expressed genes, which can subsequently be confirmed through literature, database searches, or new wet-lab experiments as genes belonging to a single regulon. This capability provides a new computational approach for easier data mining within gene expression and enables faster focus on specific *in-silico* potential regulon structures. This study presents Py/ExplorReg, which, by reproducing potential regulon structures, opens up new possibilities for further research unifying cis-regulatory elements in *Arabidopsis thaliana*.

Overall, Py/ExplorReg is applicable for researchers studying gene expression and regulation, offering enhanced capabilities for exploring complex regulatory networks. Its availability as an open-source implementation on GitHub (https://github.com/pa3cka/PyExplorReg) ensures accessibility within the field of computational biology.


## ACKNOWLEDGMENT

This work has been supported by grant project FEKT/FIT-J23-8274.

Computational resources were provided by the e-INFRA CZ project (ID:90254), supported by the Ministry of Education, Youth and Sports of the Czech Republic.


## REFERENCES


[1] V. M. Weake and J. L. Workman, “Inducible gene expression: diverse regulatory mechanisms”, *Nature Reviews Genetics*, vol. 11, no. 6, pp. 426-437, 2010.

[2] H. Zare, D. Sangurdekar, P. Srivastava, M. Kaveh, and A. Khodursky, “Reconstruction of Escherichia coli transcriptional regulatory networks via regulon-based associations”, *BMC Systems Biology*, vol. 3, no. 1, 2009.

[3] J. Schwarzerová, “Operon-Expresser: The Innovated Gene Expression-Based Algorithm For Operon Structures In-Ference”, in *Proceedings I of the 27st Conference STUDENT EEICT 2021: General papers*, 2021, pp. 301-305.

[4] J. Nejezchlebová and J. Schwarzerová, “Operon identifier: Identification of operon structures in the whole genome”, in *Proceedings II of the 28st Conference STUDENT EEICT 2022: Selected papers*, 2022, pp. 80-83.

[5] A. Almudevar, L. B. Klebanov, X. Qiu, P. Salzman, and A. Y. Yakovlev, “Utility of correlation measures in analysis of gene expression”, *NeuroRX*, vol. 3, no. 3, pp. 384-395, 2006.

[6] Z. Zhong, Y. Wang, M. Wang, F. Yang, Q. A. Thomas, Y. Xue, Y. Zhang, W. Liu, Y. Jami-Alahmadi, L. Xu, S. Feng, S. Marquardt, J. A. Wohlschlegel, I. Ausin, and S. E. Jacobsen, “Histone chaperone ASF1 mediates H3.3-H4 deposition in Arabidopsis”, *Nature Communications*, vol. 13, no. 1, 2022.

[7] *Regulon.* Online. National Library of Medicine. 1993.

[8] RODIONOV, Dmitry A. Comparative Genomic Reconstruction of Transcriptional Regulatory Networks in Bacteria. Online. *Chemical Reviews.* 2007, s. 3467-3497.

[9] G. Mönke, M. Seifert, J. Keilwagen, M. Mohr, I. Grosse, U. Hähnel, A. Junker, B. Weisshaar, U. Conrad, H. Bäumlein, and L. Altschmied, “Toward the identification and regulation of the Arabidopsis thaliana ABI3 regulon”, *Nucleic Acids Research*, vol. 40, no. 17, pp. 8240-8254, Sep. 2012.

[10] J. Bernard and J. Bernard, “Python Data Analysis with pandas”, in *Python Recipes Handbook*, Berkeley, CA: Apress, 2016, pp. 37-48.

[11] C. R. Harris, K. J. Millman, S. J. van der Walt, R. Gommers, P. Virtanen, D. Cournapeau, E. Wieser, J. Taylor, S. Berg, N. J. Smith, R. Kern, M. Picus, S. Hoyer, M. H. van Kerkwijk, M. Brett, A. Haldane, J. F. del Río, M. Wiebe, P. Peterson, P. Gérard-Marchant, K. Sheppard, T. Reddy, W. Weckesser, H. Abbasi, C. Gohlke, and T. E. Oliphant, “Array programming with NumPy”, *Nature*, vol. 585, no. 7825, pp. 357-362, Sep. 2020.

[12] P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S. J. van der Walt, M. Brett, J. Wilson, K. J. Millman, N. Mayorov, A. R. J. Nelson, E. Jones, R. Kern, E. Larson, C. J. Carey, İ. Polat, Y. Feng, E. W. Moore, J. VanderPlas, D. Laxalde, J. Perktold, R. Cimrman, I. Henriksen, E. A. Quintero, C. R. Harris, A. M. Archibald, A. H. Ribeiro, F. Pedregosa, P. van Mulbregt, A. Vijaykumar, A. P. Bardelli, A. Rothberg, A. Hilboll, A. Kloeckner, A. Scopatz, A. Lee, A. Rokem, C. N. Woods, C. Fulton, C. Masson, C. Häggström, C. Fitzgerald, D. A. Nicholson, D. R. Hagen, D. V. Pasechnik, E. Olivetti, E. Martin, E. Wieser, F. Silva, F. Lenders, F. Wilhelm, G. Young, G. A. Price, G. -L. Ingold, G. E. Allen, G. R. Lee, H. Audren, I. Probst, J. P. Dietrich, J. Silterra, J. T. Webber, J. Slavič, J. Nothman, J. Buchner, J. Kulick, J. L. Schönberger, J. V. de Miranda Cardoso, J. Reimer, J. Harrington, J. L. C. Rodríguez, J. Nunez-Iglesias, J. Kuczynski, K. Tritz, M. Thoma, M. Newville, M. Kümmerer, M. Bolingbroke, M. Tartre, M. Pak, N. J. Smith, N. Nowaczyk, N. Shebanov, O. Pavlyk, P. A. Brodtkorb, P. Lee, R. T. McGibbon, R. Feldbauer, S. Lewis, S. Tygier, S. Sievert, S. Vigna, S. Peterson, S. More, T. Pudlik, T. Oshima, T. J. Pingel, T. P. Robitaille, T. Spura, T. R. Jones, T. Cera, T. Leslie, T. Zito, T. Krauss, U. Upadhyay, Y. O. Halchenko, and Y. Vázquez-Baeza, “SciPy 1.0: fundamental algorithms for scientific computing in Python”, *Nature Methods*, vol. 17, no. 3, pp. 261-272, Mar. 2020.

[13] O. Kramer and O. Kramer, “Scikit-Learn”, in *Machine Learning for Evolution Strategies*, Cham: Springer International Publishing, 2016, pp. 45-53.

[14] I. Priness, O. Maimon, and I. Ben-Gal, “Evaluation of gene-expression clustering via mutual information distance measure”, *BMC Bioinformatics*, vol. 8, no. 1, 2007.

[15] I. Gronau and S. Moran, “Optimal implementations of UPGMA and other common clustering algorithms”, *Information Processing Letters*, vol. 104, no. 6, pp. 205-210, 2007.

# Evaluating BPL Communication: A Simulation Approach

Matěj Baranyk, Ing. Lukáš Benešl, Ing. Martin Rusz
*Department of Telecommunications*
*Faculty of electrical engineering and communication*
Brno, The Czech Republic
237853@vut.cz, lukas.benesl@vut.cz, martin.rusz@vut.cz

***Abstract*—In recent years BPL (Broadband over Power-line) technology is becoming more used. Despite the technology not needing any additional cable installation it is important to make sure that using the technology itself makes sense or what the needed requirements will be. Due to this reason, we opt to simulate almost real scenarios and compare them with outputs from network simulator ns-3. This way we can recognize the usefulness of our efforts even before we make any plans to install anything, which saves time and money. Furthermore, we can use the same simulator for diagnostics of already installed BPL routes.**

***Index Terms*—Communication, BPL, economy, diagnostics, simulation, ns-3, maintenance, emulation, testing, SNR.**

## I. Introduction to Broadband over Power-line

The ability to communicate is one of the most important skills we possess as the human race. This applies to telecommunication technologies which we are constantly trying to improve, for example, improving living conditions in developing countries. This however isn't an easy task to accomplish because we need to increase the amount of bandwidth used which leads to a further increase in price. To save funds, we are looking for new means that we can use to achieve our goals. Compared to standard types of transmission media used such as wired LAN, fiber optic cable connections, or Wi-Fi wireless connections we opt to use technologies that make use of existing power lines [1].

### A. The essentials of Power-line Communication

An example of such a technology can be Power-line Communication (PLC). The idea of using PLC emerged on the edge of the 19th and 20th centuries, where the original purpose was remote meter reading or remotely controlling the amount of transmitted power [2]. PLC is a technological process that allows data to be distributed over power lines. This makes it easy to provide internet coverage for locations where it would be difficult to install any other transmission media. This could be the case in old buildings, factories, or underground complexes (e.g. underground car parks). This way, we will increase the availability of internet connectivity in a way that is not too economically challenging and is equal in capability to other options [3]. Newly developed technologies allow us to modulate signals into different frequency bands. This process greatly expands the frequency band, which we can use (up to megahertz). One way to achieve this goal is to use Orthogonal Frequency Division Multiplexing (OFDM). This modulation uses frequency division channel technology. The transmitted signal is sent over several mutually orthogonal frequencies that do not affect each other and thus do not interfere with each other. Another problem that arises here is attenuation. As the distance over which the signal is transmitted increases, so does the attenuation that occurs on the line. The varying impedance of the power network, caused by the disconnection and connection of other appliances such as large motors or compressors that are capable of generating large interference, can also be a problem. Taps on low-voltage networks also create attenuation and there is a possibility of interference [4].

### B. Diverse Facets of Power-Line Communication

We can categorize PLC into three different sections based on the frequency range used. Depending on the frequency range used the actual practical use of the technology changes too. The first type is Ultra Narrowband (UNB). This technology usually uses a frequency range below 3 kHz. Its data transmission speed is very low and it is usually used for long-range data monitoring. The second type is narrowband. Its frequency range is 3–500 kHz. Its usage is similar to UNB, although we can also use it to automate households, and industries and to control various Smart Grid applications. Unlike UNB, this technology already has developed standards for it. The last technology is Broadband over power-line. This technology uses a frequency band from 0,5–100 MHz.

### C. Deeper Look at Broadband Power-line Communication

PLC has many variations based on the frequency we decide to use. Simulations are created for broadband. This variation of PLC serves to provide last-mile connection. In recent years it has been developing for Smart Buildings, Smart Cities, and Industry 4.0, which allow humanity easier automation and provide us with higher control over our household appliances. Another example could be HVAC (Heating, ventilation, and air conditioning) purposes. The problem with BPL communication is its vulnerability to interference, due to its high bandwidth (1,6 MHz–100 MHz, depending on the standard used). [5]

It is important to select the proper frequency range when designing our BPL circuit to prevent any kind of jamming.

Another huge risk of using BPL communications is the instability of the electrical grid itself. For example, the quality of the signal can degrade depending on other appliances in the electrical grid. The same can be said for other passive devices in the aforementioned grid, for example, DC–DC transformations. This problem has to be solved with a bypass device, to transfer the signal correctly.

### *D. Navigating Standards in Power-line Communication*

Over the course of many years many standards have been developed and experimented with. The first developed standard was HomePlug 1.0. This standard served as the very base which was further improved upon. It can cover a distance of 200 m with transfer speeds reaching up to 14 Mb/s. The successor of HomePlug 1.0 is called HomePlug AV. Serving as an improvement over its previous version, its transfer speed reaches up to 200 Mb/s with its main usage being HDTV applications. Improvement doesn't stop there, as it has been replaced by standard HomePlug AV2. This standard is backward-compatible with its predecessor. It provides us with low latency and high transfer speeds, reaching up to 2 Gbps. Its example use could be online video gaming. Standard IEEE 1901 is used for transmitting data over a distance of $<$1500 m. Another usage is its usage in vehicles or Smart Grids with distance $<$100 m. The last mentioned standard will be standard G.hn. Transfer speeds reaching up to 1000 Mb/s. Its design was made to be fully compatible with all household appliances. [6]–[12]

TABLE I
COMPARISON OF STANDARDS

| | HomePlug AV2 | IEEE 1901 | G.hn |
|---|---|---|---|
| PHY [Mb/s] | 1000 | 500 | 2000 |
| Frequency range [MHz] | 1,8–86,13 | up to 100 | up to 100 |
| Encryption | Yes (AES-128) | Yes (AES-128) | Yes (AES-128) |
| Coverage range [m] | 200 | 1500 | 200 |

## II. SIMULATING SUCCESS: THE ROLE OF NETWORK SIMULATORS

### *A. Understanding Network-Simulator 3*

To obtain accurate simulation data, it is necessary to choose a suitable simulation environment. For our purposes, we chose the ns-3 environment with an add-on module that allows us to work with PLC models. It is not too complex for installation, extensive documentation, simplicity, and modularity make it an ideal tool for the needs of creating the project. NS-3 also includes pre-installed sample programs to draw from.

The first release of ns-3 software was made in 2008. The software was designed to allow the user to be able to use C++ and Python scripts. For our purposes, version ns-3.25 was installed on the Ubuntu 14.04 LTS operating system. The ns-3 simulator is unlicensed (it is under the GNU GPLv2 license), so it is an open-source program that all parts of it can be changed or additional modules can be added to enrich its functionality.

### *B. Enhancing Simulations: The PLC Module*

The module is mainly developed for the development and testing of designs with the possibility of realistic emulation of conditions commonly encountered in practice. All these parameters are specified by the user, similarly to the geometric design of the topology, the type of cables used, or the impedance and location of the PLC device in the network. The module can even define the elements of a two-port network (transformer units). As such, the simulations also take into account time-dependent conditions that could affect the state of the variables under investigation in the emulation procedures performed. The module is also suitable for emulating situations that contain different types of noise and time variations, as it can obtain the Signal-to-Interference-to-Noise Ratio (SINR). The different communication and interference signals can be represented by approximate Power Spectral Densities (PSD).

### *C. Network Simulators: Picking Your Virtual Battlefield*

There are other simulators, that we can consider using. Other simulators could for example be MiniNET, which we opted not to use due to its lack of documentation and support of BPL communication. Another option could be OPNET, but there is no available PLC module that would prove to be useful for our purposes and it would be rather difficult to make our one. The last option was OMNeT++. This simulator has PLC capabilities but is not as useful due to its difficulty of defining user-selected cable types. Available to use are also older versions of ns, for example, ns-2 and ns-1, these are heavily outdated and lack any PLC capabilities. Therefore we chose to use ns-3 with PLC module, as it is the most customizable and well-documented option of them all. It also comes with a user guide [13] and throughout the documentation. Many other works have been published with it, making it our preferred choice.

### *D. Mastering the Simulation Software*

During our simulation, we need to consider multiple things. First, we need to consider what cable we're using. For this purpose, we have to study the physical attributes of the simulated cable. Defining topology is also needed. We have to define the distance between nodes, how many conjunctions, and possible interference. After everything has been inputted in and thoroughly checked, the software is going to output a set of values. Arithmetic mean shall be acquired from this set of numbers and compared to a laboratory test that we have conducted or compared to a dataset that has been provided to us.

## III. BLUEPRINTS OF EVALUATION: METHODOLOGICAL INSIGHTS

Various simulation environments have been created to develop these technologies. ns-3 (network simulator 3) is going to be the main tool used for testing paired with an additional module for working with PLCs. By programming various scripts, it is possible to achieve output that will be compared with real values from which a conclusion will be drawn.

### A. Crafting the Methodology

The procedure is as follows. First, a physical measuring in a lab setting is conducted. We achieve this goal by using a dual computer setup with iPerf 3 software. SW is designed to measure the throughput through the communication medium. In this process, gathering as many results for the most accurate measurement results is of utmost importance. Only then shall we proceed and perform any simulation through ns-3. Here, we try to define the scenario so it is similar to the physical measurement.

## IV. In the Lab: Conducting Practical Tests

Laboratory measurements were taken in two different settings. We have conducted multiple tests based on the distance of the circuit to gather as many values as possible. This was conducted in a less professional setting than the other measurements, which took place at a different laboratory. We have tried to prevent any outer interference from being involved in these measurements, however, it is impossible to say we have eliminated it.

### A. The Toolbox: Equipment Under Scrutiny

We have measured two different BPL modems. PPC 4BBPL1L1B and Xingtera MOD957. The cable type was AlFe 36Al/1Fe. PPC type modem fully supports data transmission on low and medium-voltage power grids in accordance to IEEE1901 FFT OFDM Access standard. Its transmission rate from or to the low-voltage network is 200 Mb/s each. Typically its net transmission rate is 5–50 Mb/s. Its frequency range is 2–30 MHz. Xingtera modem fully supports all ITU–T G.hn standards with a throughput of up to 800 Mb/s. Its frequency range is 2–100 MHz. The second BPL modem was Xingtera MOD957 [12]. The primary difference between the modems used is in their standard and the bandwidth used, which is also related to the maximum physical (PHY) throughput. Another important difference is the way the signal is injected into the communication medium. The first modem used an inductive coupler that must be threaded onto the MV cable, including the cable that is connected directly to the modem. Xingtera has a coupling member formed by a coupling capacitor. The aim was to compare the two standards (G.hn and IEEE1901) with each other.

### B. Numbers Speak: Unveiling Acquired Data

During our laboratory test, we have conducted two measurements. One at max cable length, which was 3 m, and then another test, where we lowered the distance by 0,4 m. Our tests were run for 60 seconds. The throughput was in the range of 23,4–26,0 Mb/s. The average during our 60-second long segment was 25,1 Mb/s. The total amount of data transmitted was 180 MB. After reducing the distance measured the speed went up as expected. During this measurement, the lowest speed achieved was 28,40 Mb/s. The highest speed achieved was 30,1 Mb/s. The average for this segment is 28,9 Mb/s. The total amount of data transmitted was 207 MB. It is hard to create a simulation for this type of cable as there is a lack of documentation to be found for precise results.

TABLE II
Acquired Values

| | | | |
|---|---|---|---|
| Distance | [m] | 2,6 | 3,0 |
| Lowest throughput | [Mb/s] | 28,40 | 23,39 |
| Highest throughput | [Mb/s] | 30,10 | 28,90 |
| Average value | [Mb/s] | 28,90 | 25,10 |
| Total data transmitted | [MB] | 207,0 | 180,0 |

## V. Signal Integrity Analysis: Comparison

### A. The Signal-To-Noise Paradigm

SNR serves as a ratio between the strength of our signal and the sum of the strength of the interference and noise. As the value of this parameter increases, the properties of the network improve. For this example, the NAYY–J cable and a dataset from Germany from a project by Fühler-im-Netz (FiN) will be used. In this project, the SNR value was recorded every 15 minutes for 22 months. This dataset provides a wide range of data against which accurate comparisons can be made. The tables obtained contain 1536 carriers, of which only 917 are useful. The dataset also contains notches in which other technologies operate. For this reason, they are not used as they would interfere with each other. After emulating this situation, comparisons of the individual averages will be made from which a conclusion can be drawn.

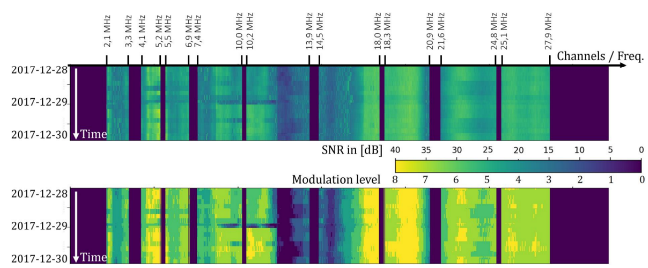



Fig. 1. Notches in conducted testing by FiN [14]

### B. Insights and Implications: Analyzing the Data

The track that we simulated in this case and that was measured is 169,3 m long. The effect of junctures is negligible as there are no junctures on the track. After using the simulator, the output of the simulation is a number where the SNR is equal to 10,814 dB. The value from the provided dataset after averaging and rounding is equal to 7,358 dB. This difference equals to 3,456 dB is not very large, but it is still present. There are several possible explanations for why this phenomenon occurs. As might be expected, measurements made using the ns-3 simulator produce more ideal values. This effect can occur for several reasons. The first reason may be the emulation itself, which may not take into account all the physical properties of the cable we are working with, which can negatively affect our result. Another factor may be the age of the cable itself. The actual weather in which the readings were recorded may also affect the measurement, as the cabling may be affected by the external environment. It is also important to consider the time range at which the readings were recorded. The human factor itself may have used these PLC connections during active measurement times. There are also uncontrollable influences on the measurement that can negatively affect it. The process of calculating the arithmetic mean from the emulated values also plays a role in our comparison, as we are working with all the values provided (a very wide dataset compared to the simulation). Another factor that affects the comparison is the dissimilarity of the cable. In reality, we work with the NAYY–J cable, whereas in the simulation we only work with the NAYY cable. The cable that is used in reality contains an additional GN–YE protective conductor, which could also affect our measurement to a certain degree.

TABLE III
SIMULATION TOPOLOGY

| Parametr | | Practical situation | Simulation |
|---|---|---|---|
| Route length | [m] | 169,3 | 169,3 |
| Conjunction count | [-] | 0 | 0 |
| Cable type | [-] | NAYY50-J | NAYY50 |
| PSD | [dBm/Hz] | -55 | -55 |
| **SNR** | **[dB]** | **7,358** | **10,814** |

## VI. CONCLUSION

The project serves as an example of how we can easily simulate a scenario, which will be comparable in practice. This way, we can save money on planning and installing new cables to a location. We can predict how a connection will behave before its installation and testing. We can then easily state if it is worth installing a BPL circuit in our place of interest or if it is better to opt in for different solutions, for example, due to connectivity issues.

### *A. Methodologies in Perspective: A Comparative Analysis*

We can state that we have conducted multiple tests and simulations to see, what's useful and yields the best results. The simplest way to acquire a good comparison was to use an already provided dataset and make our simulation as similar as possible. This has given us the best results, where we can see the difference. We can state, that if we were given a cable of NAYY type we could reliably simulate its behavior before its installation. The second test we have conducted has also yielded decent results, which we sadly cannot compare to the simulation due to the lack of documentation found for this type of cable. The third conducted test yielded results that were deemed useless, therefore there is no reason even to try and make a simulation to see, how this scenario would transpire. The error caused too much signal radiation into space, which means the induction of the signal on the cable is incorrect. We can therefore consider this measurement as not relevant.

### *B. Understanding Variance: The Why Behind the Numbers*

The answer is rather simple. Conducting tests in an already working environment is easier than using experimental, untested technology. During our tests, we figured out that it might be impossible to conduct any measurements on the experimental BPL modem designed in our colleague's master's thesis.

### *C. Pioneering Contributions of Our Study*

The contribution of the paper is the comparison of measurement and simulation and the use of publicly available data (the FiN project). This paper lets us see if it is possible to compare real-life data with simulated data to improve said BPL route. We can see that the simulation doesn't differ much from the real-life situation, therefore we could use these simulations to diagnose already planted cables. Furthermore, we could simulate planned BPL routes before they are even planted to see if it's worth our time, money, and effort. We could also diagnose problems (junctions, noise, interference) before they even arise. If it's deemed not worth using BPL, we can use different technologies such as LTE, 5G, or many others.

## REFERENCES


[1] S. K. Mohanty, R. K. Giri, The analysis of broadband communication over indoor powerline channel, in: 2014 International Conference on Communication and Signal Processing, IEEE, 2014, pp. 1293–1299. doi:10.1109/ICCSP.2014.6950059.

[2] E. Kabalci, Y. Kabalci, From Smart Grid to Internet of Energy, Elsevier, 2020.

[3] ACRYLIC SUITE, Repeater, extender, plc, or wifi mesh: which is better?, https://www.acrylicwifi.com/en/blog/repeater-extender-plc-or-wifi-mesh/ (2023).

[4] K. Dostert, Powerline communications, Prentice Hall, 2001.

[5] P. Mlýnek, M. Rusz, L. Beneš, J. Sláčik, P. Musil, Possibilities of broadband power line communications for smart home and smart building applications, Sensors (Basel, Switzerland) 21 (1) (2021) 240. doi:10.3390/s21010240.

[6] M. K. Lee, R. E. Newman, H. A. Latchman, S. Katar, L. Yonge, Homeplug 1.0 powerline communication lans-protocol description and performance results, International journal of communication systems 16 (5) (2003) 447–473. doi:10.1002/dac.601.

[7] K. Afkhamie, S. Katar, L. Yonge, R. Newman, An overview of the upcoming homeplug av standard, in: International Symposium on Power Line Communications and Its Applications, 2005, IEEE, Vancouver, BC, Canada, 2005, pp. 400–404. doi:10.1109/ISPLC.2005.1430539.

[8] S. Katar, M. Krishnam, R. Newman, H. Latchman, Harnessing the potential of powerline communications using the homeplug av standard, R. F. Design, accessed: 2023-10-22 (2006).
URL https://shorturl.at/nzMZ2

[9] L. Yonge, J. Abad, K. Afkhamie, et al., An overview of the homeplug av2 technology, Journal of electrical and computer engineering 2013 (2013) 1–20. doi:10.1155/2013/892628.

[10] Ieee standard for broadband over power line networks: Medium access control and physical layer specifications, IEEE Std 1901-2010 (2021). doi:10.1109/IEEESTD.2021.9329263.

[11] X. Trulls, D. Mateo, A. Bofill, A high dynamic-range rf programmable-gain front end for g.hn rf-coax in 65-nm cmos, IEEE transactions on microwave theory and techniques 60 (10) (2012) 3243–3253. doi:10.1109/TMTT.2012.2207913.

[12] M. Mahút, Broadband over powerline modem, Ph.D. thesis, Vysoké učení technické v Brně. Fakulta elektrotechniky a komunikačních technologií (2023).

[13] F. Aalamifar, L. Lampe, Plc software user guide, https://shorturl.at/blRT5 (2013).

[14] C. Balada, S. Ahmed, A. Dengel, M. Bondorf, N. Hopfer, M. Zdrallek, Fin: A smart grid and powerline communication dataset, Zenodo (2022). doi:10.5281/zenodo.5948716.

# Simulator of IEC 60870 communication

1st Tomáš Calábek
*Department of Telecommunications*
*Brno University of Technology*
Brno, Czech Republic
237881@vutbr.cz

2nd Antonín Bohačík
*Department of Telecommunications*
*Brno University of Technology*
Brno, Czech Republic
Antonin.Bohacik@vut.cz

***Abstract*—The power industry's evolution hinges on efficient and reliable transmission, distribution, and control of electricity, increasingly facilitated by remote monitoring and control systems. Among these, SCADA technologies stand out, enabling centralized supervision and management of energy infrastructure. Central to these systems are communication protocols such as protocols from IEC 60870 standard, facilitating data exchange within power networks. This paper introduces a communication simulator based on IEC 60870-5-101/103/104 protocols. Our simulator for IEC 60870 is capable of simulating all three mentioned protocols on the client-server principle. It can simulate both balanced and unbalanced transmissions, adjust transmitted data, set quality bits, and much more. Our simulator was tested in a virtual environment.**

## I. Introduction

Today, the power industry is a key sector that is constantly evolving to ensure efficient and reliable transmission, distribution, and control of electricity [1]. With the development of technology and automation in the power sector, the remote control is becoming an increasingly important system, enabling centralized monitoring and control of power equipment, including distribution networks [2]. In the field of remote control of power systems, modern technologies such as SCADA (Supervisory Control and Data Acquisition) are used, which enables operators to efficiently monitor and control energy equipment remotely [3]. These controls are key to ensuring the safe and optimized operation of power plants, transmission networks, and distribution systems [4]. In the context of remote control and SCADA systems, communications play a key role. Protocols that enable the transfer of information between devices in a power system. These protocols include the IEC 60870 standard, which defines the communication method between devices in the energy sector [5].

The aim of this paper was the presentation of the communication simulator based on IEC 60870-5-101/103/104 protocols. This simulator provides a variety of functions such as changing the transmitted data, the period of sent messages, or communication protocol, also sending various queries, or creating communication datasets. The simulator also includes a user interface that makes it easier to work with our developed simulator. The final simulator was tested in a virtual environment using 4 VMs (Virtual Machine), where 3 were individual servers of the mentioned protocols and one VM where all clients were stored.

### *A. State of the Art*

In the domain of operational communication generators/simulators, a diversity of manufacturers and services is evident. While the generators generate almost random data, simulators generate data with a sort of logic behind them. Numerous companies engage in the development of such tools, each offering its unique solution. Examples of these manufacturers include National Instruments with their NI VeriStand software, Doble Engineering Company with their simulators, and OPAL-RT Technologies with their Real-Time Simulator.

However, it is crucial to note that not all these products sufficiently cover a broad spectrum of ICS (Industrial Control Systems) protocols. Some companies focus solely on specific protocols or provide limited support for various industrial control standards. Examples of companies specializing in ICS protocol simulators include ABB Group with their System 800xA simulator, Honeywell International Inc. with their Experion PKS simulator, and Schneider Electric with their EcoStruxure Control Expert simulator.

Particular attention must be paid to simulators operating with the standard 60870 protocol, which is crucial for many industrial applications, especially in the energy sector. Unfortunately, only a limited number of companies offer protocol generators of this specification. Among them are Siemens AG with their SIMATIC WinCC simulator for the 60870 protocol and ABB Group with their RTU500 series, which supports the 60870 protocol.

The following Table I provides a concise overview of companies involved in communication data generation and analyzes the solutions they offer, including their protocol coverage and other key functionalities.

TABLE I
Overview of Companies Engaged in Data Generation and Analysis of Provided Solutions.

| **Offered solution** | **Supported protocols** | **Price** |
|---|---|---|
| NI VeriStand | Modbus, CANopen | Unspecified |
| Simulators ICS | DNP3, IEC 61850 | 275 € |
| EE Power RTS | IEC 60870, IEC 61850, DNP3 | 18.300 € |
| ABB System 800 | IEC 61850, DNP3, Modbus | Unspecified |
| Experion PKS | IEC 61850, Modbus, PROFIBUS | 155 € |
| EcoStruxure | IEC 61850, Modbus, PROFINET | 6.510 € |
| SIMATIC WinCC | IEC 60870, Modbus, PROFINET | 1.000 € |
| **Our IEC 60870 Sim.** | IEC 60870-5-101/103/104 | $<$ 100 € |

## II. SCADA

The standard provides functionality for remote controlling SCADA systems. The term SCADA refers to a set of software and equipment used to monitor and control technical or industrial equipment. The SCADA system works on the principle of control and server stations. In this respect, IEC 60870 provides SCADA systems with support for the implementation of functions that are crucial for the proper functioning of SCADA systems. The most important functions that IEC 60870 provides to SCADA systems include the following [4]:

- Ensuring interoperability – IEC 60870 is designed to ensure interoperability between the various components of SCADA, facilitating seamless communication across different devices and systems.
- Improving reliability and safety – IEC 60870 makes SCADA systems implement an authentication and access control process, preventing unauthorized access to critical control points. At the same time, IEC 60870.
- Real-time communication – IEC 60870 can handle requests. communication in real-time. By using efficient communication protocols the standard facilitates the rapid exchange of data between field devices and the central SCADA control center, enabling rapid decision-making

### *A. IEC 60870*

IEC 60870 is a set of international standards, developed by the IEC (International Electrotechnical Commission), which addresses the need for efficient and reliable communication in power systems [4]. IEC 60870 is referred to in the Czech Republic as CSN EN 60870. It is the name for the whole group of standards called *Remote control systems and equipment*. These standards were originally developed to support systems for SCADA in the energy sector. The IEC 60870 standard is based on the EPA architecture, as well as protocol DNP3 [5].

### *B. IEC 60870-5*

IEC 60870-5 also referred to as *Communication Protocols* defines detailed descriptions of the useful functions of systems for remote control and the control of geographically extensive processes. These functions include three of the most important – Polling, Report by Exception, and time stamp assignment [6]. Polling involves the master unit in a communication system sending requests to multiple server units, each responding as required, while Report by Exception allows server stations to communicate with the master independently, transmitting crucial information promptly. Timestamps, automatically attached to events, help identify the initial event amidst subsequent occurrences, requiring accurate time synchronization between the master and server units for effective usage.

The IEC 60870-5 standard provides information on a set of information elements that are suitable for a wide range of SCADA applications, especially for power distribution system applications [4]. This data is transmitted within ASDUs (Application Service Data Units) carrying application data, which may contain one or more information, see Fig. 1 [7]. Each data unit has a unique type identification number. Only one data type is included in each ASDU, located in the first field of the ASDU. The types of information elements are defined by the standard and grouped by direction, either monitoring or control, and by the type of information that is transmitted, such as process information, system information, or file transfer parameters. Furthermore, messages can be classified according to the actual reason that caused the message to be sent, referred to as COT (Cause Of Transmission). This classification serves to improve clarity when processing received messages. The next section in the ASDU message header is the sequence number SQ (Sequence number), which indicates the order of the messages received. It also contains the sender address of the message known as OA (Object Address) and an information element, which is a key element for the transmission of information.

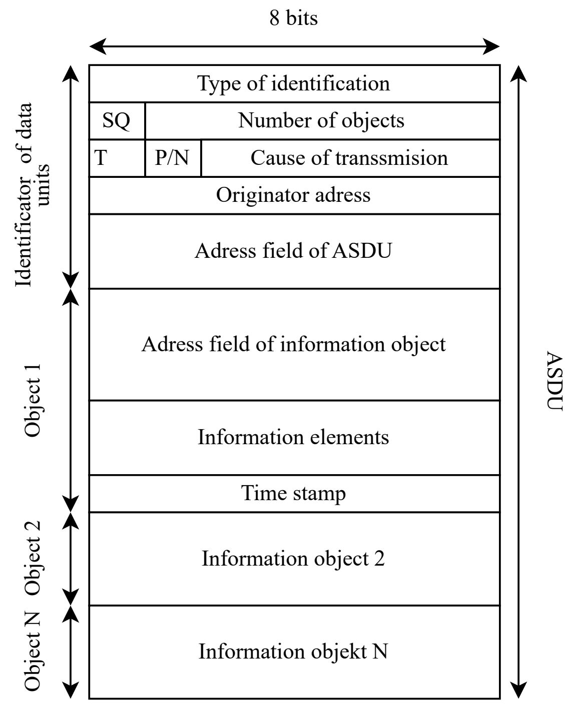



Fig. 1. Structure of ASDU format [7]

*1) IEC 60870-5-101:* This is a companion standard known as *Transmission Protocols – Companion standard especially for basic telecontrol tasks* that defines a communication profile for the transmission of basic messages of remote control messages between the central control station (client) and the remotely controlled stations (server) [8]. The companion standard works on the 2nd layer, i.e. link layer data. IEC 60870-5-101 allows two basic ways of transmitting communication: (1) unbalanced transmission and (2) balanced transmission [7].

In unbalanced transmission, the control station initiates all communication by polling the controlled outputs, supporting services such as SEND/NO REPLY for global messages and cyclic commands, SEND/CONFIRM for control commands and setup commands, and REQUEST/RESPOND for calling data from controlled stations.

On the other hand balanced transmission, each station can initiate message transmission, acting as both control and controlled stations, with supported services like SEND/CONFIRM and SEND/NO REPLY limited to point-to-point and multipoint configurations. IEC 101 is used in telecontrolling power systems, electric power systems, teleprotection, etc. It also serves as the basis for the IEC 104 standard, which is an extension of it that can communicate via TCP/IP protocol, i.e. wirelessly.

*2) IEC 60870-5-103:* This standard specifies the information interface of protection systems in the field of automation and control of power systems. This standard aims to ensure compatibility between protection devices and control system devices in server stations [9]. The VDEW protocol, incorporated into the IEC 60870-5 protocol suite, specifically the discussed IEC 60870-5-103, is a five-tone selcall mode defined by the German Verband der Elektrizitätswirtschaft. Within IEC 60870-5-103 there are two methods of information exchange. First, it uses standardized messages and application procedures to transfer standardized ASDUs. The second method uses generic services to transfer a wide range of information. The standardized messages cover only a limited set of protection functions, meaning that protection devices can only support a subset of the functions defined by this standard. Predefined messages and application procedures are mandatorily used where possible, and generic services are used otherwise. Overall, IEC 60870-5-103 and the VDEW protocol provide a framework for efficient communication and information exchange between protection devices and control systems in power systems, with an emphasis on compatibility and the ability to use different transmission protocols, including TCP/IP.

*3) IEC 60870-5-104:* As mentioned before IEC 104 is an extension standard that expands IEC 101 with a TCP/IP interface so it's able to suit complete network access [7]. The application layers of these two standards are the same with the difference that some of the data types are not implemented so they cannot be used. The IEC 104 data contain mechanisms for synchronization. One of the most common devices where IEC 104 is used is the MT-151 which is a telemetry mobile controller.

## III. IEC 60870 COMMUNICATION SIMULATOR

While existing solutions cater to specific protocols or lack comprehensive support, our simulator fills a critical void by offering a versatile platform encompassing the full spectrum of ICS protocols, particularly IEC 60870. Featuring a modular architecture and user-friendly interface, our simulator facilitates testing, research, and education in the energy and automation sectors. Its capabilities extend to simulating diverse communication models, enhancing its applicability for developers, researchers, and practitioners. Furthermore, its potential integration into physical devices bridges the gap between theoretical exploration and practical implementation. In summary, our IEC 60870 communication simulator represents a significant advancement, addressing the challenges of communication protocol testing and development in the power industry. For implementation purposes, the practical part will focus on the analysis and verification of the functionality of the IEC 60870-5-104 protocol. This part focuses on the protocol implementation. The communication simulator will serve as testing software for people working with telecommunication protocols in the energy industry, but also as a kind of insight into this standard for any non-expert user.

The simulator for the IEC 60870 communication standard significantly enhances scientific research and development in the energy and automation sectors through diverse approaches. It serves as an important tool for testing and validating implementations by providing a simulated environment that aligns with the standards outlined in IEC 60870. Developers can thoroughly test various device and system implementations, facilitating error identification, resolution, and performance optimization. Additionally, the simulator supports research and development efforts by providing a platform for experimentation without the constraints of physical deployment, thereby reducing both the time and costs associated with developing new technologies or refining existing systems. Moreover, the simulator is pivotal in education and training initiatives within the energy and automation domains. It furnishes students, technicians, and professionals with a practical environment for hands-on work and experimentation with communication protocols and devices, fostering deeper understanding and proficiency without the risks associated with real-world deployment. In discussing the IEC 60870 communication standard simulator, it's crucial to highlight its potential for implementation within real, physical devices. This feature not only ensures seamless integration between virtual simulations and hardware but also extends the utility of the simulator in the context of practical use.

### A. Software design

The simulator comprises two core components: (1) client and (2) server. Each component is intricately crafted to adhere to the nuances of the protocol version, facilitating seamless extension with additional protocols via modular add-ons. The client module not only accommodates interaction with simulated servers but also interacts with real-world energy or industry devices. This integration offers distinct advantages, including the ability to test and validate device behavior, as well as generate datasets from live equipment. Such versatility enhances the utility of the simulator for both testing and real-world application scenarios.

### B. Simulator Design

Fig 2 illustrates the testing topology of the created IEC 60870 communication simulator. The implementation of this simulator involves modular architecture, utilizing appropriate programming languages and frameworks for protocol handling and message parsing. Key components include modules for protocol handling, message parsing, data simulation, and a GUI (Graphical User Interface). The GUI provides users with intuitive interaction, allowing configuration of simulation parameters, monitoring of communication traffic, and analysis of protocol behavior. It will offer features such as configuration settings, message monitoring, simulation controls, data visualization, and error handling.

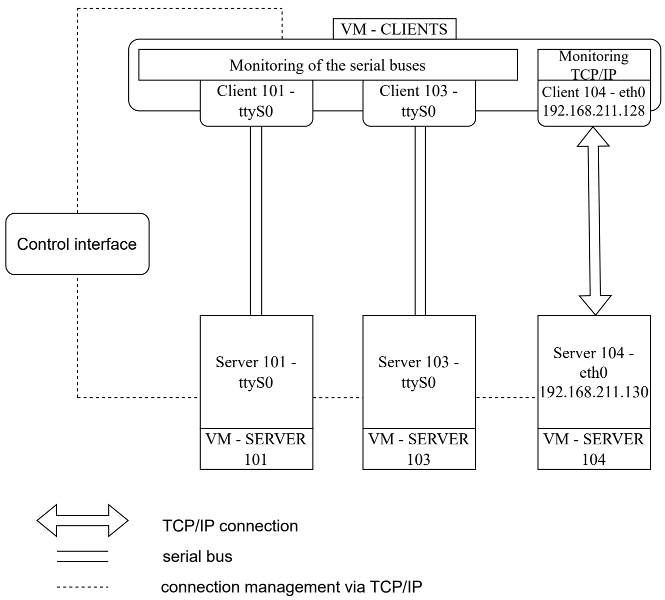



Fig. 2. Test scheme of the IEC 60870 protocol simulator

Figure shows a diagram of the operation of the communication simulator. This diagram meets all the objectives, as it contains 4 virtual stations – 1 virtual station where clients of all mentioned protocols run, and 3 virtual stations, where one of the servers operates in each of them. These virtual stations are shown in the diagram as VM – CLIENTS, VM – SERVER 101, etc. Servers are shown in the diagram as Server 101, Server 103, etc. They also described which communication interface they are working on. All protocol clients are running on the VM – CLIENTS virtual station. At the same time, network monitoring is connected to the clients, but for the IEC 60870-5-101/103 protocol, the monitoring of serial buses is involved, whereas, for the IEC 60870-5-104, it is monitoring TCP/IP connection. For clients, as well as for servers, it is also noted on which communication interface they operate. Here only exception is Client 104, which operates on the *eth0* interface. For better and clearer functionality, a *Control Interface* is also added, which will subsequently manage all simulators of each protocol and will be the GUI of the simulator. The double-sided arrow indicates a TCP/IP connection, the two parallel lines mark the serial bus connection, and the dashed line is used to connect the Control Interface to the individual clients and servers.

The GUI will provide features such as configuration settings where the user will be able to specify parameters such as communication modes (client/server), message types, data objects, and network settings. It will also provide data visualization where the GUI will offer graphical representations of simulated data, such as charts, tables, and plots, to facilitate data analysis and interpretation. The last important thing is going to be error handling where will GUI provide feedback on simulation errors and issues, helping users troubleshoot and debug their implementations.

## IV. Conclusion

In the realm of operational communication generators and simulators, numerous manufacturers offer a variety of solutions. While companies provide unique tools, not all adequately cover a broad spectrum of ICS protocols. Some focus solely on specific protocols or offer limited support for various industrial control standards. Few companies specialize in simulating the IEC 60870 protocol, crucial for many industrial applications, especially in the energy sector.

In this paper, we introduced the concept of an IEC 60870 communication simulator. This simulator offers a versatile platform for testing, research, and education in the energy and automation sectors. Its modular architecture, with components for protocol handling, message parsing, data simulation, and a user-friendly graphical interface, ensures flexibility and usability. Our simulator represents a significant contribution to the field a comprehensive solution to the challenges of communication protocol testing and development. Our simulator's potential for implementation in industry devices enhances its utility, bridging the gap between theoretical exploration and practical deployment. In the context of future development, the communication simulator will be optimized for better performance and efficiency. Additionally, support for communication protocols such as IEC 61850, DNP3, Profinet, EtherNet/IP will be expanded to enable communication simulation across various devices, including ABB REC615 and PQ monitor MEg44PAN. The simulator's GUI will be enhanced with new features such as data visualization, communication statistics, and interactive elements for scenario creation.

## References


[1] C. Rohmingtluanga, S. Datta, N. Sinha, and T. S. Ustun, "Scada based intake monitoring for improving energy management plan: Case study," *Energy Reports*, vol. 9, pp. 402–410, 2023.

[2] T. Cherifi and L. Hamami, "A practical implementation of unconditional security for the iec 60780-5-101 scada protocol," *International Journal of Critical Infrastructure Protection*, vol. 20, pp. 68–84, 2018.

[3] G. Cheng, Y. Lin, A. Abur, A. Gómez-Expósito, and W. Wu, "A survey of power system state estimation using multiple data sources: Pmus, scada, ami, and beyond," *IEEE Transactions on Smart Grid*, 2023.

[4] G. Clarke and D. Reynders, *Practical modern SCADA protocols*. Oxford: Newnes, 2004.

[5] V. Skoko, B. Atlagic, and N. Isakov, "Comparative realization of iec 60870-5 industrial protocol standards," in *2014 22nd Telecommunications Forum Telfor (TELFOR)*. IEEE, 2014, pp. 987–990.

[6] Y. RUDZINSKI and P. VLADYKA, "Komunikační protokoly pro dálkové ovládání iec/iso 60870-5," 2010.

[7] P. Matoušek, *Description and analysis of IEC 104 Protocol*. Faculty of Information Technology BUT, 2017. [Online]. Available: https://www.fit.vut.cz/research/publication/11570

[8] M. Uzair, "Communication methods (protocols, format & language) for the substation automation & control."

[9] A. Sudhakar, R. C, and A. S, "Implementation architecture of iec 60870-5-103 communication protocol on arm platform running on rtos for industrial ieds," in *2022 4th International Conference on Inventive Research in Computing Applications (ICIRCA)*, 2022, pp. 80–86.


PROČ ČEKAT
NA BUDOUCNOST,
KDYŽ JI MŮŽETE
VYTVÁŘET?
Honeywell
Objev kariérní příležitosti v největším evropském R&D centru společnosti Honeywell v České republice. Nabízíme široké možnosti uplatnění v oblasti výzkumu, vývoje a IT se zaměřením na letectví a vesmírné technologie, řešení pro zvýšení bezpečnosti práce a produktivity.
www.honeywell.com
careers.honeywell.com/Brno
careers.honeywell.com/EarlyCareersCZ
THE FUTURE IS WHAT WE MAKE IT
Honeywell

# IoT Monitoring system for 3D Printer

1st Samuel Vaculík
*Brno University of Technology*
*Faculty of Electrical Engineering and Communication*
Department of Control and Instrumentation
Brno, Czech Republic
240462@vutbr.cz

2nd Ing. Jakub Arm, Ph.D.
*Brno University of Technology*
*Faculty of Electrical Engineering and Communication*
Department of Control and Instrumentation
Brno, Czech Republic
arm@vut.cz

***Abstract*—This paper consists of the design and implementation of a 3D printing related data collection system. The aim is to define the variables suitable for measurement, then to select suitable sensors for given measurements and implement them on the printer. The next point is the design and implement system for collecting print data from the printing firmware.**

***Index Terms*—Klipper, Moonraker, Mainsail, MQTT, Raspberry Pi**

## I. Introduction

Nowadays, 3D printers based on the FDM principle are quite popular. The reason for this is the increasingly simple operation of newly released printers, which allows even inexperienced users to use the printer. At the same time, the new printers are equipped with advanced features that automate the entire 3D printing process, so there is less and less human intervention in the printer. Examples of such features include algorithms for automatic calibration of the first print layer, X and Y axis orthogonality checking, or nozzle cleaning of tightly before printing begins.

However, all these compensations focus only on the print itself (or the printer), and the printing problem may not be caused by a bad printer setting, but by parasitic influences acting on it. From this point of view, it is therefore advisable to monitor all variables that could adversely affect printing and then, according to their values to compensate or stop the printing.

The open-source printing firmware Klipper was chosen to control the printer.
As the printer is ready for an automatic print farm thanks to the pull-out system, it is desirable to be able to access the printer remotely without the need for to upload files to the printer using an SD card each time it is printed. Octoprint is a frequently used environment, but another pair of environments have been created specifically for Klipper - Mainsail and Fluidd. Mainsail was chosen for this purpose.

## II. Main idea

Measuring external quantities on the printer is particularly suitable for 3D print farms. In a fully automated line, printers operate without a single human intervention (with respect to Industry 4.0 concept, even the printing itself can be started without operator intervention, for example when an order is created) and the operator only checks the status of the entire line. From this already implies the use of sensors to monitor the printer's status.

## Architecture

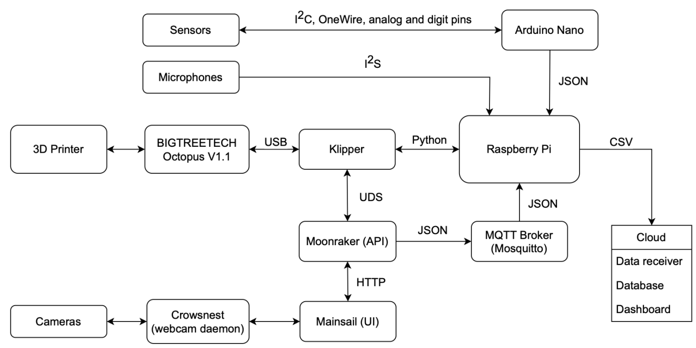



Fig. 1. Designed architecture of the entire system [1]

## III. Measurement of quantities

The following section will briefly mention the physical variables that should be measured to properly evaluate the printer's condition.

The **ambient temperature** is measured by <u>SHT-31</u> and <u>DHT-22</u> sensors. Temperature fluctuations are a significant problem for certain materials. In particular, ABS and ASA materials can flex and deform if not printed in a closed box. This then makes it impossible to use the print. It is therefore advisable to measure temperature fluctuations at the printer and report them to the operator if necessary.
The SHT-31 sensor is connected to the $I^2C$ bus. Unfortunately, the module has only one address and it was not possible to use two of these modules at the same time.

TABLE I
Comparison of SHT-31 and DHT-22 sensors

| Parameter | SHT-31 | DHT-22 |
|---|---|---|
| Supply voltage | 2,4 - 5,5 V | 3,3 - 6 V |
| Current draw during measurement | 0,8 - 1,5 mA | 1 - 1,5 mA |
| Idle current consumption | 45 - 70 $\mu$A | 40 - 50 $\mu$A |
| Temperature measurement range | -40 - 125 °C | -40 - 80 °C |
| Temperature measurement accuracy | ±0,3 °C | ±0,5 °C |
| Moisture measurement range | 0 - 100 % | 0 - 100 % |
| Moisture measurement accuracy | ±2 % | ±2 % |

**Ambient humidity** under normal operating conditions does not effect the printer itself, but rather the used filaments. Each filament is hydroscopic to some degree, which naturally absorbs moisture from the air. Moisture from the filament then evaporates when printing in the hotend, and the printstring thus contains a number of voids that cause uneven printing.

The most commonly used materials such as PLA, PET-G, ASA or ABS suffer the least from moisture. On the other hand, PVA is a very hydroscopic material, which must be kept in a closed box when printing, otherwise the surrounding moisture could start to have a negative effect within a few hours.

The same sensors were used for humidity measurements as for temperature measurements.

**Motor temperatures** should be measured for two reasons. The first reason is to detect motor overheating, since despite high operating temperatures, stepper motors usually do not exceed the temperature of 100 °C on a conventional 3D printer. The second reason is to detect different motor temperatures. The printer on which the work was carried out uses 4 motors for Z-axis motion. These motors should have the same temperatures when printing as they are loaded identically. Different temperatures could then indicate a failure on one of the motors.

Temperature measurement of all motors is realized by temperature sensors DS18B20 from Dallas.

TABLE II
DS18B20 TEMPERATURE SENSOR PARAMETERS

| Parameter | Value |
|---|---|
| Supply voltage | 3 - 5,5 V |
| Current draw during measurement | 1 - 1,5 mA |
| Idle current consumption (up to 70 °C) | 0,7 - 1 $\mu$A |
| Idle current consumption (at 125 °C) | 3 $\mu$A |
| Temperature measurement range | -55 - 125 °C |
| Temperature measurement accuracy (-10 - 80 °C) | ±0,5 °C |
| Temperature measurement accuracy (-55 - 125 °C) | ±2 °C |

**Current flow in motor** can together with knowledge of motor temperature can lead to the detection of failure. All motor control takes place in the motor driver, which also store all data related to the running of the motor. When using TMC drivers, the current values are stored in the MSCURACT register, which is divided into two 9-bit sections to store the values from windings A and B.

**Processor temperatures** can contribute to the optimization of operating conditions. Temperature monitoring can identify potential overheating problems, which can lead to performance limitations. Internal thermistors are used to measure the temperatures of the Raspberry Pi CPU board and printer control board. Therefore, no additional sensors were needed.

**Gases** are released when most materials are melted. The same is true for 3D printing. In particular, carbon dioxide, carbon monoxide or ammonia may be released when the printing strings are melted. Other gases released are already in very low concentrations and therefore have no impact on humans. In addition, more accurate sensors would be needed to measure them.

Carbon dioxide was measured with the sensor MQ-135, for carbon monoxide the sensor MQ-7 was chosen. The placement of the sensors on the printer was realized according to the physical properties of the gases. Carbon dioxide is heavier than air and was therefore placed in the lower part of the printer, while carbon monoxide is lighter and was therefore placed in the upper part of the printer.

TABLE III
PARAMETERS OF MQ-135 AND MQ-7 GAS SENSORS

| Parameter | MQ-135 | MQ-7 |
|---|---|---|
| Supply voltage | 5 V | 3,3 - 6 V |
| Current draw during measurement | 160 mA | 70 mA |
| Measurement range [ppm] | 10-300 ($CO_2$) | 20-2000 ($CO$) |

**Vibrations** occur with every mechanical movement. In certain cases they can be desirable, but in 3D printing the aim is to eliminate them. Mechanical vibration can cause print layers to shift, resulting in uneven print walls. Two MPU-6050 modules were used to measure vibration with $I^2C$ communication interface. The modules have a temperature sensor in addition to a 3-axis accelerometer and gyroscope. Its accuracy is very low and is therefore it is not recommended to use.

TABLE IV
MPU-6050 SENSOR PARAMETERS

| Parameter | Value |
|---|---|
| Supply voltage | 2,375 - 3,46 V |
| Current draw during measurement (max) | 3,9 mA |
| Idle current consumption | 20 $\mu$A |
| Gyroscope measurement range [°/sec] | ±250, ±500, ±1000, ±2000 |
| Accelerometer measurement range [g] | ±2, ±4, ±8, ±16 |

**Filament weight** is a measure of the amount of material remaining, as running out of filament is one of the common problems in 3D printing. Especially in print farms, it is therefore necessary to weigh the remaining material and check if there is enough for the print.

Here you can use the MMU (Multi Material Unit), which not only supports multi-colour printing, but can also automatically switch to a second ready spool when the material runs out.

However, even when using the MMU unit, it is not possible to prematurely detect running out of filament. Therefore, it is advisable to use a weight sensor to weigh the material in real time. Thus, the filament weighing is realized by the YZC-131 sensor together with AD converter HX-711.

TABLE V
YZC-131 SENSOR PARAMETERS

| Parameter | Value |
|---|---|
| Supply voltage | 5 - 10 V |
| Measurement accuracy | 0,05 % (Full Scale) |
| Hysteresis | 0,03 % (Full Scale) |
| Nonlinearity | 0,05 % (Full Scale) |
| Temperature compensation | -10 - 40 °C |
| Safe overload | 120 % |
| Maximum overload | 150 % |

TABLE VI
HX-711 ADC PARAMETERS

| Parameter | Value |
|---|---|
| Supply voltage | 2,6 - 5,5 V |
| Current draw during measurement | 1,5 mA |
| Idle current consumption | 1 $\mu$A |
| Number of bits | 24 |
| Gain | 32, 64, 128 |

**Filament diameter** cannot be due to manufacturing tolerances maintained along the entire length of the string, and it is therefore advisable to measure the possible ovality of the filament. The measured values are used both to display the actual filament diameter, but more importantly, they can influence extrusion. Ovality of the printing string could result in uneven laying of the print layers.

Two linear Hall probes SS49E, connected in differential mode to compensate for temperature drift, were used for the measurements.

TABLE VII
SS49E HALL PROBE PARAMETERS

| Parameter | Value |
|---|---|
| Supply voltage | 3 - 6,5 V |
| Current draw during measurement | 4,2 - 8 mA |
| Sensitivity | 1,8 mV/G |
| Output voltage (min) | 0,86 V (B = -1500 G) |
| Output voltage (max) | 4,21 V (B = 1500 G) |

**Sound recording** is combined with vibration sensors for a more comprehensive picture of the condition of mechanical components. However, for proper sensing, the printer should be in enclosed box so that the microphones do not pick up sounds around the printer.

Two microphones SPH-0645LM4H with $I^2S$ interface were used for detection. Although the microphones are on a common bus, each of the microphones at the bottom and top of the printer senses into a different audio channel. In this way, data resolution is ensured.

TABLE VIII
SPH-0645LM4H MICROPHONE PARAMETERS

| Parameter | Value |
|---|---|
| Supply voltage | 1,62 - 3,6 V |
| Current draw during measurement | 600 $\mu$A |
| Idle current consumption | 3 - 10 $\mu$A |
| Frequency range | 50 - 15 000 Hz |
| SNR (Signal to Noise Ratio) | 65 dBA |
| PSR (Power Supply Rejection) | -86 dBA |
| THD (Total Harmonic Distortion) | 1 % (SPL 110 dB) |

**Camera** serves as a remote monitoring of the printer, but using machine learning algorithms, it can also automatically detect print crashes and therefore print to free space. At such a time, it is advisable to stop printing or inform the operator in order to save material. Two USB webcams with a focal length of 3.85 mm and 10x digital zoom were used for image recording.

## IV. PLACEMENT OF SENSORS

The placement of the sensors on the printer was made taking into account the physical properties of the given quantities and also the efficiency of the measurement. All sensors were selected according to affordability and sufficient accuracy.

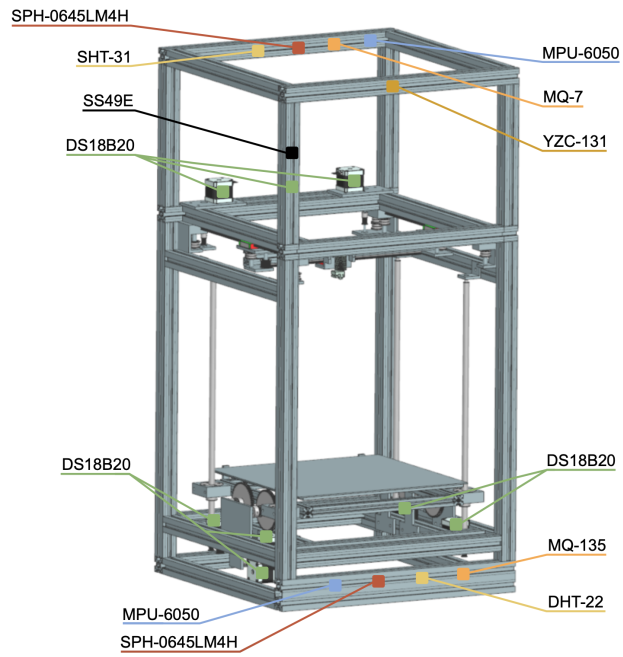



Fig. 2. Example of sensor placement on the 3D model fo the printer [1]

## V. HARDWARE USED FOR DATA COLLECTION

Data acquisition from the sensors was divided into two parts due to the missing $I^2S$ bus on the Arduino Nano board. Data from most of the sensors are read by the Arduino Nano microcontroller. Although this module has the today's relatively slow 8-bit ATmega328 processor, it is sufficient for the purpose of data acquisition. The reason for using it was the price and also the possibility of programming in the Arduino IDE environment.

The reading of the audio data from the microphones is then implemented using a Raspberry Pi 3B.

## VI. COLLECTING DATA FROM THE PRINTER FIRMWARE

Although the goal was to collect data from the printing firmware, Klipper is not responsible for this transmission. The Moonraker communication interface is used for proper interaction between Klipper and the Mainsail user environment. This is a Python 3 based API server. For the transfer of print data, the MQTT protocol was chosen, for which Moonraker has a ready implementation.

To make the data transfer work, the MQTT broker had to be installed on the Raspberry Pi. This is the link between senders and receivers, and its purpose is to forward messages to the devices. Each message contains a so-called *topic*. Clients will only receive the sent messages if they receive the *topic*.

Then it was necessary to write the *mqtt* section into the printer configuration file, in which the broker address, port, protocol type, but above all the data to be sent were defined (see the following table for an example).

```
heater_fan Hotend_fan = speed, rpm
gcode_move = position, speed, speed_factor
hall_filament_width_sensor = Diameter
extruder = target, temperature, power
heater_bed = target, temperature, power
print_stats = total_duration, print_duration
print_stats = filament_used, info.total_layer
temperature_sensor raspberry_pi = temperature
temperature_sensor board_sensor = temperature
```

## VII. Results

The data collection system has been tested on a test print and it is functional. A sample of the measured data is below.

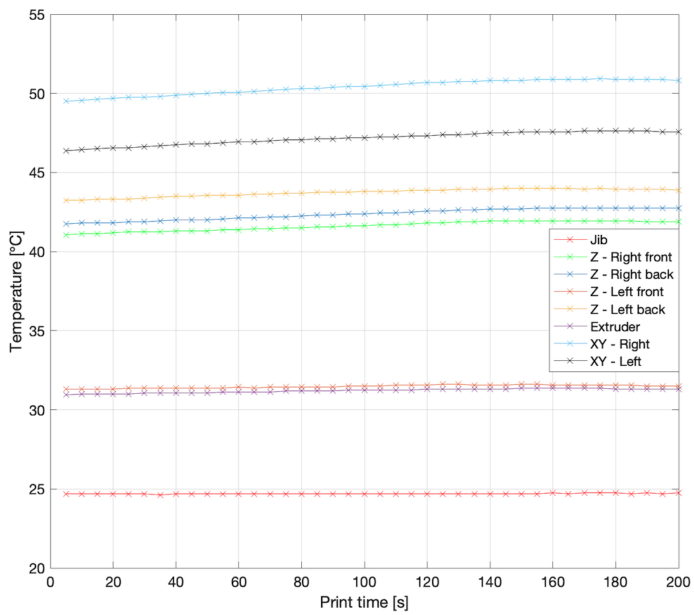



Fig. 3. Measured temperatures of stepper motors during printing [1]

The jib motor reached the lowest temperature during the test print as it was not active. On the other hand, the highest temperatures were achieved by the X and Y motors, where an increase in temperature over time can be observed. The extruder motor reached a temperature of around 32°C throughout the printing process. Although the measured temperature is quite low (extruders usually reach temperatures above 40°C), it could correspond to reality as the extruder is fitted with a heat sink to prevent overheating.

A spoken word test of the microphones was performed on a sample audio recording for 2-5 s. At time 6 s, an impact to the bottom of the printer structure was then performed (red waveform). The same impact was then made to the upper part of the structure (blue waveform). It can be observed that the impacts to the upper part were also picked up by the lower microphone (and vice versa).

Currently, the system is designed to measure motor temperatures with a 5 second period and a 22 $\mu$s audio data reading period. The system will be modified if these values prove to be inadequate.

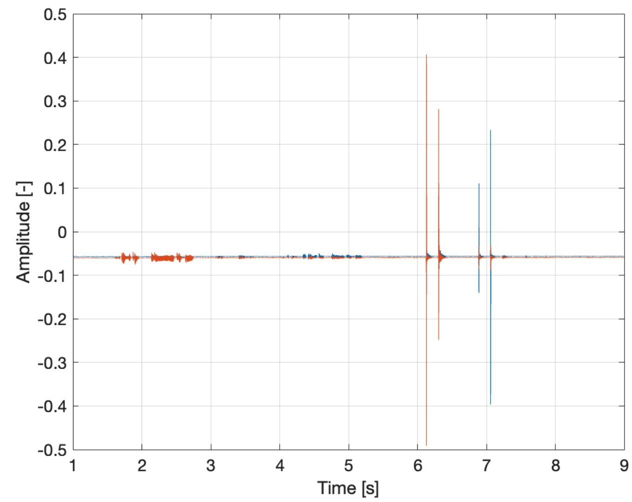



Fig. 4. Test sound recording [1]

## VIII. Conclusion

The aim of the paper was to describe the physical variables affecting 3D printing. Each variable was described how it can affect the printing process and why it is useful to measure it. In addition, the sensors for the measurements were selected, followed by the placement on the printer structure. The paper also discussed the Klipper print firmware and the design of the print data acquisition system. The frequently used IoT protocol MQTT with Mosquitto broker was chosen for this transmission.

## Acknowledgement


The completion of this paper was made possible by the grant No. FEKT-S-23-8451 - “Research on advanced methods and technologies in cybernetics, robotics, artificial intelligence, automation and measurement” financially supported by the Internal science fund of Brno University of Technology.


## References


[1] VACULÍK, Samuel. IoT systém pro zpracování dat z 3D tiskárny. Online, Semestrální práce, vedoucí Ing. Jakub Arm Ph. D. Brno: Vysoké učení technické v Brně, 2024. Dostupné z: https://www.vut.cz/www_base/zav_prace_soubor.php?file_id=262045&zp_id=155205. [cit. 2024-03-07].

[2] Klipper API. Online. 2016. Dostupné z: https://www.klipper3d.org/Status_Reference.html?h=status. [cit. 2024-03-08].

[3] Moonraker MQTT. Online. C2024. Dostupné z: https://moonraker.readthedocs.io/en/latest/configuration/#mqtt. [cit. 2024-03-08].

[4] Mainsail - Crownsnest. Online. Cca 2021. Dostupné z: https://crowsnest.mainsail.xyz. [cit. 2024-03-08].

# Wide area tracking system based on LoRa devices

1st David Holík
*Department of Control and Instrumentation*
*Brno University of Technology*
Brno, Czech Republic
230074@vut.cz

2nd Ing. Jakub Arm, Ph.D.
*Department of Control and Instrumentation*
*Brno University of Technology*
Brno, Czech Republic
jakub.arm@vut.cz

***Abstract*—The work deals with wireless communication using LoRa (Long Range) technology categorized as LPWAN (Low Power Wide Area Network). The technoloy enables wireless data transmission over long distances more than 5 kilometers in urban or rural areas. The technology is used as the communication between ESP32-based devices providing their GPS data and time synchronization. The system also consists Raspberry Pi which provides the central LoRa station for near area. The communication between central and peripheral devices was established using open-source libraries. The central device might need to receive packets from surrounding devices concurrently, whichi is not possible using the designed device, thus a protocol needs to be established as discussed. Moreover, the communication with the public TTN network has been tested. Regarding the communication testing, test scenarios were designed.**

***Keywords*—LoRa, LoRaWAN, IoT, LPWAN, GPS, Spreading Factor, The Things Network**

## I. Introduction

The main motivation for the proposed system is that it will be used to implement an interactive story game at a summer camp. The game will take place in the varied terrain around the campground, from where it will be possible to receive location information and also exchange information about the time of a given device. The architecture of the entire network was chosen with this in mind. That is, a private network without the need to connect to the Internet network, which, in fact, may not even be available in a given location. The connection is point to multipoint. The main point is a stationary device located at the base and directly powered by the mains. This device is considered as a reference in terms of time. The end devices in the surrounding range then synchronise their time according to it. These devices are wearable and therefore change their position in real time, which they then feed back into the system. They are powered by an external battery.

The first communication was through the creation of an application within the public network with which the end device then connected. In this way it was possible to send data and receive acknowledgement of the delivery. Point to point communication was achieved by establishing a connection between two LoRa modules located in close proximity. Publicly available libraries for the hardware were used. The devices are now capable of exchanging data with each other.

## II. LoRa technology overview

Long Range (hereafter referred to as LoRa), is a wireless modulation technology developed by Semtech Corporation. It belongs to the Low Power Wide Area Network (LPWAN) group. LoRa uses CSS modulation technology, commonly used for sonar and radar.

The maximum size of one message is 51-242 Bytes. The data rate (DR) ranges from 0.25-50 kbps [1]. The distance at which a transmitted message can still be received is up to 5 kilometres in built-up areas of a city and up to 15 kilometres in areas outside the city [3]. These two parameters are defined by the spreading factor, hereafter referred to as SF. There are 6 designations used in the European communication system (SF7-SF12). The duration of one chirp for each SF7-12 ranges from 1-32 ms [4]. That is, the size of the message and at what speed it is sent. It should be remembered that a higher transmission rate reduces the distance between the transmitter and receiver required to transmit the message [1]. The transmission of messages takes place on internationally unlicensed frequency bands (ISM) reserved for industrial, scientific and medical purposes. Typical bands for LoRa are 433 MHz, 868 MHz and 915 MHz. The 433 MHz and 868 MHz frequencies are used in Europe and 915 MHz in the United States. The bandwidth in which it transmits is 125-500 kHz. The specific bandwidth is determined by the legislation for each of the world regions. For example, in Europe the maximum bandwidth allowed is 250 kHz [2]. Authentication and encryption of communication between end devices and the application server is achieved by the AES 128b standard.

The LoRa Alliance was formed to manage and further develop LoRa communications. It is developing its global standard for wireless communication - LoRaWAN. The LoRa Alliance currently has over 500 members (companies and enterprises) [1].

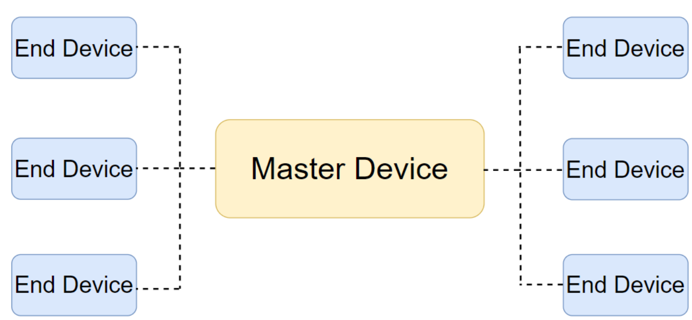



Fig. 1. The chosen network architecture

### III. Design of the System

#### A. Network architecture

Fig. 1 shows the network architecture that will be used. As already mentioned the central node is the master device, which holds the reference timing data and provides it to the end devices. In return, it receives data from them about their geographical location. Both the game itself and the measurements during the implementation of the test scenarios will be conducted in this spirit.

#### B. Design of master and end device

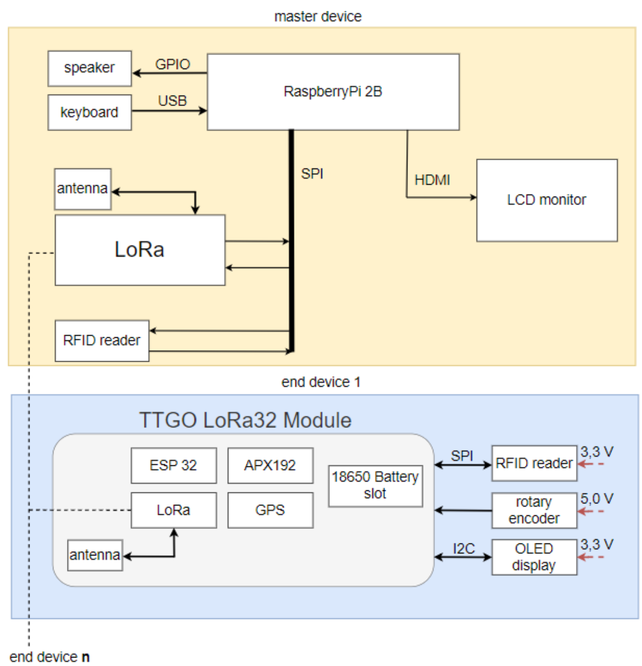



Fig. 2. Block diagram of the proposed system

The specific design of the system can be seen in Fig. 2. It shows the selected components that should meet the requirements to ensure proper functionality. The core components are a microcontroller, for the control of the device, a LoRa communication module, and then other peripheral devices connected to the microcontroller.

*1) Master Device:* It will be located at a base within range of the electrical grid connection on which its functions will depend. The main function will be to control and evaluate wireless communication, interact with a user via a clear display of available information, and data backup.

RaspberryPi 2B is used as an available COTS device. Its task will be to control the whole system. It will be possible to connect a keyboard, LCD monitor, RFID reader, LoRa module and speaker. The power supply will be done by a mains adapter.

The RFID reader will be used to authorize access to the profile of a given user in the system and subsequent modifications.

LoRa module is currently using Ra-01h. It mediates the actual wireless communication of the main device with the secondary devices deployed in the field. The main requirements are reliability, low power consumption and communication range up to 5 kilometers. Transmission will be at 868 MHz.

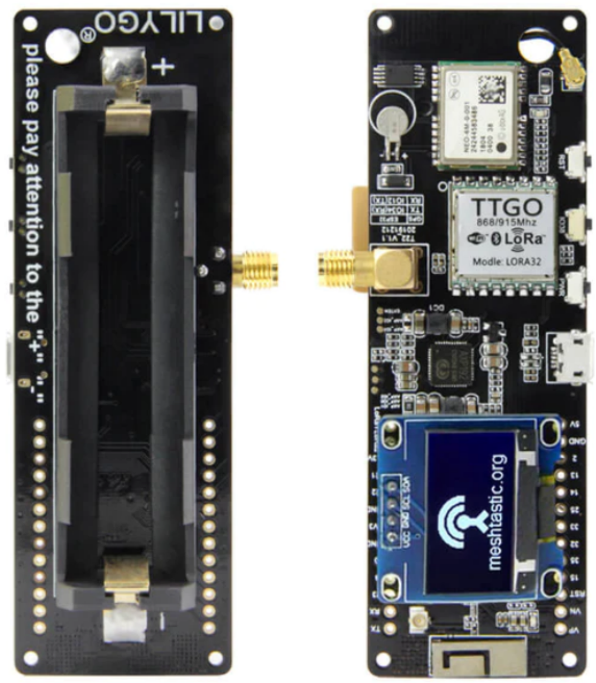



Fig. 3. ESP32 TTGO T-Beam board [6]

*2) End Device:* It will be away from the mains most of the time and will therefore be powered by a battery cell. For this reason, the resulting consumption of each end device is also important to us. Its main task will be to transmit information to the main device about the time adjustments of each user and to project them into the running application. Furthermore, reporting its current geographical location. Thus, it will be possible to keep track of where a device is located. Therefore, it would be useful if the resulting device is smaller in size and easily portable.

For the implementation of the end device, a very well equipped TTGO board from LilyGO was chosen for this specific use case. This is because it has ESP32, LoRa32 modules, GPS and battery management chip integrated on it. There is a slot on the back for inserting it. An external OLED display, RFID reader and rotary encoder will still be needed to provide the rest of the functions. The board provides voltage levels of 3.3 and 5.0 V. The board can be seen in the Fig. 3.

### IV. Wireless Communication

#### A. communication to the TTN public network

In the previous work, the initial effort was to establish a connection via the LoRa module to the public network. For this purpose, the environment from The Things Industries -

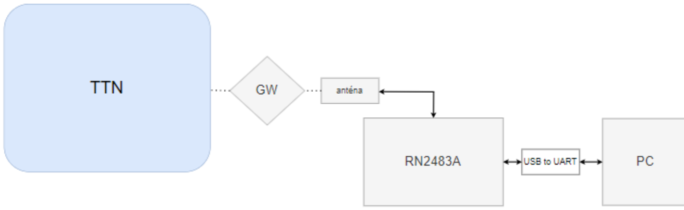



Fig. 4. Block diagram of wireless communication on TTN network

The Thing Stack (TTS) was chosen. It is the latest and most recent version of the LoRaWAN server of The Things Network (TTN). It provides all the functions of the LoRaWAN network. The hardware in the Fig. 4 was used for implementation.

For transmitting was used RN2483 module from Microchip, which contains radio transmitter SX1276 from Semtech. It has a UART interface through which ASCII commands can be sent. Using the user manual to control the module, the necessary parameters to establish a connection to the network (unique device identifiers and application key) were set up in sequence. Then the commands were used to connect to the TTN and send a simple ASCII message. The commands were sent to the module from the computer via the serial monitor. The connection was made via a CP2102 USB converter. An example of a recorded communication from this device to the TTN public network is shown in the Fig. 5.

First you need to create an account called The Things ID. Then you went to the console, which is used to manage the applications and gateways created in TTS. The option to manage applications was selected and a custom one was created. In settings, an end device was added. Next, a frequency plan had to be selected, which in this case was EU868. Next, the LoRaWAN network version. The important thing at the end was to enter the global device identifiers and the root key needed for OTAA (Over The Air Activation).

Connection and sending of data to the network with subsequent response back was successful through the gateway during testing in Brno and Zlín.

### B. Point-to-Point communication

The connection was established and text messages were sent in an understandable form between the RPi controlled Ra-01h module and the TTGO board. Both devices were able to work as single-channel transmitter and receiver. A publicly available library and code examples from an online sources were used for communication. The next step could be to modify the code or find a more suitable code containing time synchronization for the use case.

### C. Test scenarios

In order to test and verify the wireless communication parameters, it is necessary to define test scenarios in which specific parameters will be measured. Based on the results obtained, it will be possible to evaluate the characteristics of the system. Five different scenarios were proposed:

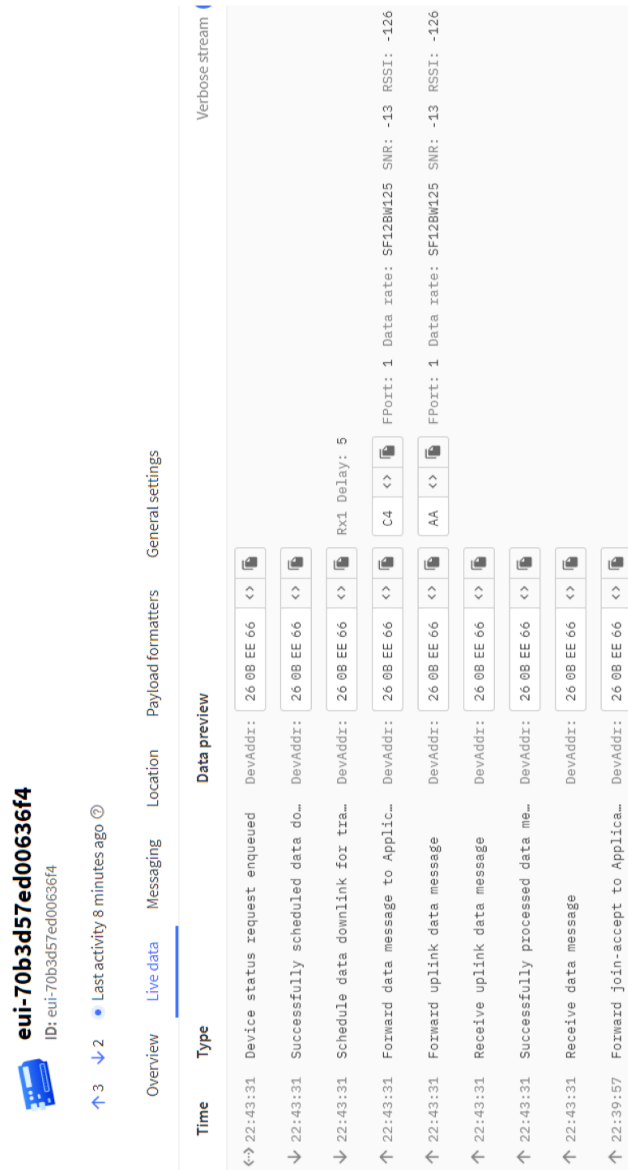



Fig. 5. Recorded communication in the TTS console

1) Communication range - important data: GPS, SNR, RSSI.
2) Reliability - define a counter variable that increments by 1 after each packet sent. The current packet number is sent along with its other contents. Thus, it is possible to evaluate how many messages never reached the destination.
3) Synchronization - Evaluate only after the connection between the master and the end device has been established. Then the end device will have the current time sent from the master device. The difference between the original time (at the master device) and the time received by end device will be evaluated. It is then possible to try to implement a PTP library to minimize this time difference.

4) Maximum size of the transmitted message - different sizes up to the maximum defined LoRa depending on the distance.
5) Message frequency - per unit time. Keep track of the transmission time and accordingly determine the maximum number of messages transmitted per day for a given LoRa configuration, so as not to violate the limitations imposed by the legislation.

### D. Networking

The created architecture comprises more devices which can transmitt a message at a time. If the LoRa device enabling only single-channel communication is used, the message passing can be corrupted. In such case, some methods of communication control at network level can be adopted such as TDMA (Time Division Multiple Access) or Token passing. The both techniques have limitations; TDMA needs to define the time slot according to the system topology (node ocunt). Token passing techniques require token management according to the device capabilities and demands.

If the multi-channel gateway is used, the concurrent messages are possible while channel settings on every device is necessary. Then, LoraWAN protocol manages the communication at the network level [5].

## V. Conclusion

According to the scrutiny, the LoRa technology is a suitable communication for WAN systems. It enables the establishment of a private network without an internet connection comprising of central and peripheral devices. Some countries also supports global networking over the public LoRaWAN TTN network. The both architectures were implemented in this work using LoRa devices Ra-01h and SX1276 (RN2483). Currently, a simple message containing a text string and the sequence number of the packet to be sent can be sent between two nodes. Moreover, test scenarios have been designed to verify the features of the system.


## Acknowledgment

The completion of this paper was made possible by the grant No. FEKT-S-23-8451 - “Research on advanced methods and technologies in cybernetics, robotics, artificial intelligence, automation and measurement” financially supported by the Internal science fund of Brno University of Technology.


## References


[1] *LoRaWAN.* Online. The Things Network. Available at: https://www.thethingsnetwork.org/docs/lorawan/ [cit. 2023-10-04].

[2] RAZA, U.; KULKARNI, P. a SOORIYABANDARA, M. *Low Power Wide Area Networks: An Overview: An Overview. IEEE Communications Surveys and Tutorials.* 2017, roč. 19, č. 2, s. 855-873. ISSN 1553-877X. Available at: https://doi.org/10.1109/COMST.2017.2652320

[3] *LoRa and LoRaWAN: A Technical Overview.* Online. In: SEMTECH CORPORATION. Semtech: Developer portal. December 2019. Available at: https://lora-developers.semtech.com/uploads/documents/files/LoRa_and_LoRaWAN-A_Tech_Overview-Downloadable.pdf [cit. 2023-11-08].

[4] ETSI. *System Reference document (SRdoc); Technical characteristics for Low Power Wide Area Networks Chirp Spread Spectrum (LPWAN-CSS) operating in the UHF spectrum below 1 GHz* Online. V1.1.1. 2018. Available at: https://www.etsi.org/deliver/etsi_tr/103500_103599/103526/01.01.01_60/tr_103526v010101p.pdf [cit. 2024-01-02].

[5] LoRa Alliance. *What are LoRa and LoRaWAN*. Online. In: Semtech developer portal. Available at: https://lora-developers.semtech.com/documentation/tech-papers-and-guides/lora-and-lorawan/

[6] *ESP32 LoRa Wireless Module.* Online. In: LILYGO. Available at: https://www.lilygo.cc/products/t-beam-v1-1-esp32-lora-module. [cit. 2024-03-08].

# Micromobility and their charging

Michal Papež
*Department of Electrical Power Engineering*
*Brno University of Technology*
Brno, Czech Republic
240783@vutbr.cz

***Abstract*—This paper is focused on electric micromobility, deals with the definition of micromobility, presents selected types of electric micromobiles and their parameters especially their battery capacity. The second part of the article discusses possibilities and main factors essential for the development of charging infrastructure for electric micromobility vehicles, primarily the infrastructure for electric bikes.**

***Keywords—micromobility, shared micromobility, battery, micromobile, charging infrastructure***

## I. Introduction

The paper introduces electric micromobility as a way of traveling by electric vehicle which is smaller and lighter than a common car. The author of the word „*Micromobility*“ is Horace Dediu, and he limited this group of smaller vehicles by the maximum weight of the vehicle, which should be 1000 kilograms [1]. Horace Dediu chose this number as a sufficient limit which cannot limit the future development in this section of vehicles. Among electric micromobiles belong electric scooters, electric bikes, electric longboards and skateboards, electric unicycles and electric wheelchairs. The number of sold electric micromobiles is increasing (for example electric bikes in Fig. 1) and this fact raises the question about the state of the actual charging infrastructure. For a better charging infrastructure, the actual state needs to be mapped and, if necessary, add charging points in places without any charging infrastructure to improve the traveling comfort of an e-biker.

## II. Types of electric micromobiles

Types of electric micromobility vehicles are mainly based on non-motorized or internal combustion engines variants of vehicles. An electric micromobile in a form of an electric bike offers a good ride quality even after the battery has been discharged. The way of riding an electric bike is more familiar to the population than riding other types of electric micromobiles, which contributes to their popularity.

Electric scooters are also widely used, as we can see in cities, but they do not offer the rider as many of the health benefits as riding an electric bike. Riding an electric bike requires at least a minimum of physical activity.

Among the most popular electric micromobiles belong electric unicycles. The speed of the electric unicycle is affected by shifting weight forward or backward. Rolling is achieved by tilting sideways. The unicycle allows the rider to be balanced by a mechanism that uses accelerometers, gyroscopes, and a magnetometer. [2]

Fig. 2 shows battery capacities of selected types and models of electric micromobiles, which are commercially available. The data are based on [3]-[23].

## III. Shared micromobility

Shared micromobiles (without engine) and electric micromobiles are very common especially in cities. The main advantage is smaller size of the vehicle which can help in traffic jams and while parking, another positive is that micromobiles are powered by human power or by electric engine and that can lead to cleaner air, another advantage is that user does not need to buy own vehicle. Shared micromobility vehicles are primarily used for travel distances shorter than 5 kilometers. Such distances account for 50-60% of the total distance traveled in China, Europe and the USA. [24]

Well-known companies which share micromobility vehicles are for example Bolt, Lime and many others. Company Bolt offers electric scooter named Bolt Model 5 [25] and Bolt Model 6 [26]. The Bolt Model 5 has a battery capacity of 676.8 Wh, a nominal voltage of battery is 47 V, a nominal power of 350 W and a peak power of 700 W. [25]

The Bolt Model 6 has a battery capacity of 1100 Wh, and maximum range of the Model 6 is about 90 kilometers. [26]

## IV. Development of charging infrastructure

Following steps illustrate the motivation and options for expanding charging infrastructure for electric micromobiles.

### *A. Motivation*

The motivation for finding a way how to improve the charging infrastructure can be the increasing number of sold units of electric bikes in the European Union from 2006 to 2022, which is shown in Fig. 1.

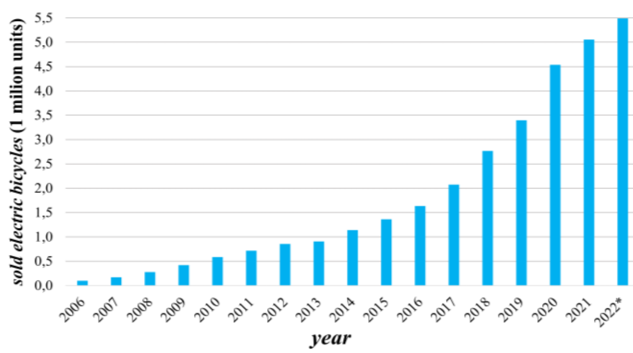



Fig. 1. Sales of electric bikes in the European Union 2006 – 2022 [27][28]

*the number of units in 2022 includes the United Kingdom (Fig. 1)

Another factor that can motivate to build and to improve charging infrastructure comes from a political decision. For example, a new upcoming European standard that specifies how many bike parking spaces are available at building. Among these buildings belong residental buildings and non-residental buildings which are new or buildings which are going through major renovation. [29]

According to the new European standard, TABLE I. shows minimal number of parking spots for bikes in buildings.

TABLE I. NUMBER OF NEEDED SPOTS FOR BIKES [29]

| **Type and condition of the building** | **Number of parking spots for cars** | **Number of parking spots for bikes** |
|---|---|---|
| New or reconstructered residental building | More than 3 | Min. 2 for every apartment |
| New or reconstructered non residental building | More than 5 | Min. 15% of the average people capacity of the building or at least 10% of the total capacity of the building |
| Existing non-residental building | More than 20 | Min. 15% of the average people capacity of the building or at least 10% of the total capacity of the building |

## B. *Interval of battery charging*

The time interval required for charging can limit the user when traveling. The time for charging can also be limited by the type of location. The time the user spends at the location almost equals the time available for charging.

Each location can offer a certain level of potential for charging, the potential being caused by the attractiveness or relevance of the location. The biggest potential for charging can be for example at workplace or schools etc.

The interval of charging is affected by capacity size of the battery and chosen charger. Battery capacity is limited by the size of the battery, battery size is also affected by the location within the electric bike design. The battery can be fully integrated into the down tube. Battery can be also mounted on the downtube, on the rear carrier or behind the seat tube.

TABLE II. shows selected batteries from BOSCH, BAFANG and SHIMANO, the table includes charging times for different types of chargers compatible with the selected battery.

TABLE II. SELECTED BATERRIES OF ELECTRIC BIKES AND THEIR CHARGING TIME

| **Battery** | **Charger** | **Time of charging (hours)** |
|---|---|---|
| BOSCH PowerPack 400 (400 Wh) [30] | 2A charger | 6 (100%), 2.8 (50%) |
| | 4A charger | 3.5 (100%), 1.5 (50%) |
| BOSCH PowerPack 625 (625 Wh) [30] | 2A charger | 9.3 (100%), 4.6 (50%) |
| | 4A charger | 5.4 (100%), 2.1 (50%) |
| BAFANG BT C01.750.UC (750 Wh) [31] | 2A charger | 10 (100%) |
| | 3A charger | 6.5 (100%) |
| BAFANG BT F014.410.C (410 Wh) [32] | 2A charger | 6 (100%) |
| | 3A charger | 4.5 (100%) |
| Shimano BT-EN805 (504 Wh) [33] | EC-E8004 | 4 (100%), 2.5 (80%) |
| Shimano BT-EN606 (630 Wh) [33] | EC-E8004 | 4.8 (100%), 3.2 (80%) |

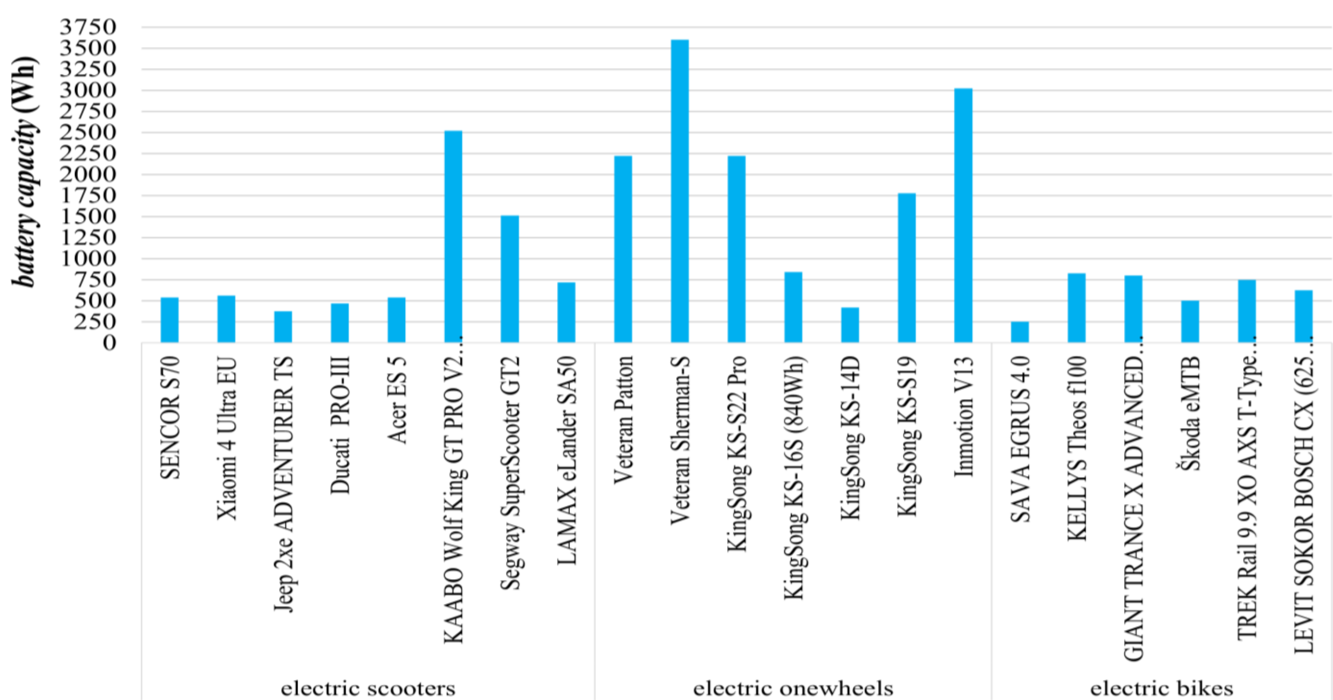



Fig. 2. Battery capacities of selected micromobiles [3]-[23]

## V. Development opportunities

The new charging stations can be outdoor (uncovered) type, these stations are mainly used for short or medium charging times. The second option is a covered variant, which provides protection against weather and theft. An example of a covered variant is shown in the Fig. 3. For both options, it is important to ensure a safe charging process in terms of fire protection.

Fig. 3. Spot for electric bikes at Brno University of Technology [34]

Subchapters A, B and C represent the possibilities of expanding the infrastructure.

### *A. Synergy with electric cars charging stations*

This synergy was made for example in the state of Oregon in the USA [35]. Adding an outlet to charging station is an opportunity for e-bikers and the impact of the innovation is minimal.

### *B. Synergy with public lightning infrastructure*

Lightning infrastructure can be used to charge electric cars, the charging station is on the lamp. The biggest problem is that cars need more space for parking and it's a big limitation for this strategy of improving the charging infrastructure for electric cars. This variant of development has bigger potential for charging electric micromobiles because micromobiles require a smaller parking space.

### *C. Station with own energy source*

Some companies offer charging stations with their own energy source. For example, the Czech company LEDEOS [36] offers stations with photovoltaic panels. This way is suitable for locations without local power lines, such as areas in nature.

## VI. Place selection

After battery capacity analysis and construction options, it is worth mentioning the place selection. Place selection is influenced by the state of the current charging infrastructure. The current charging stations may be overused, adequate for the need or the place may have no charging stations at all.

Possible locations for building are at the user´s place of residence, at the user´s place of work, and at popular tourist locations such as rest areas, cultural and nature monuments, etc.

A possible method of finding locations where charging infrastructure needs to be expanded is to focus on specific bike routes. In Europe, these are for example EuroVelo bike routes. Four EuroVelo routes pass through the Czech Republic (Fig. 5). The bike EuroVelo routes through the Czech Republic are EV4, EV7, EV9 and EV13.

Fig. 4. EuroVelo bike routes through the Europe [37]

Fig. 5 shows the EV4 route in the Czech Republic with charging points along the route. The locations of the charging points are taken from Mapy.cz [38], Dobiju.cz [39], Kdenabiju.cz [40] and LEDEOS [41]. The disadvantage is that the information about the location is taken from the internet, for accuracy the actual state of the selected location must be personally checked.

From the Fig. 5 is obvious that some parts of the route are without infrastructure. The longest distance between 2 charging stations is about 90 kilometers (on the marked EV4 route). In this part of route are charging points needed. Good charging infrastructure offers charging point every 5-10 km of the route (a distance of 5-10 km represents a distance that an e-biker should be probably able to ride without the help of an engine).


Fig. 5. EV4 in the Czech Republic and charging places along the route [38][39][40][41]

## VII. Conclusion

The points mentioned in the article were intended to highlight the elements that cannot be overlooked in the design process of expanding charging infrastructure. An essential element is the analysis of battery capacities (Fig. 2), because the capacity affects range and recharge time.

For a complete design, it is necessary to select specific place of interest for development. These may be locations that are near cycle routes (Fig. 5), residential buildings, sports facilities, places of work, etc. Depending on the location, a covered or uncovered charging station design will be selected. It is also necessary to determine the expected number of bikes at the place and the associated number of chargers from specific manufacturers and some outlets for chargers of different types than those available at the place.

The information presented in this article will soon be used in a theoretical proposal to expand the charging infrastructure in a selected location in the Czech Republic.

## ACKNOWLEDGMENT

I would like to express my appreciation to my thesis supervisor Ing. Martin Paar, Ph.D., for his expert suggestions, support and substantive stylistic advice that helped in the preparation of this article.

## REFERENCES

[1] H. Dediu, “The Micromobility Definition”, *Micromobility industries*, 2019. [Online]. Available at: https://micromobility.io/news/the-micromobility-definition.

[2] “Electric unicycle”, *Wikipedia: the free encyclopedia*. [Online]. Available at: https://en.wikipedia.org/wiki/Electric_unicycle.

[3] “SCOOTER S70 KOLOBĚŽKA”, *SENCOR*. [Online]. Available at: https://www.sencor.cz/kolobezka/scooter-s70.

[4] “Xiaomi Electric Scooter 4 Ultra EU”, *Xiaomi Česko*, c2010-2023. [Online]. Available at: https://www.xiaomicesko.cz/xiaomi-electric-scooter-4-ultra-eu.html.

[5] “Jeep 2xe ADVENTURER TS”, *Jeep urban e-mobility*, c2023. [Online]. Available at: https://www.ejeep.cz/product/jeep-2xe-adventurer-TS/ .

[6] “ELECTRIC SCOOTER PRO-III” , *Ducati Urban e-Mobility*. c2022. [Online]. Available at: https://www.ducatiurbanemobility.com/electric-mobility/pro-iii-2/.

[7] “Řada Acer ES 5”, *Acer*, c2023. [Online]. Available at: https://www.acer.com/cz-cs/emobility/escooters/acer-es-series-5.

[8] “KAABO Wolf King GT PRO V2 CZ EDITION”, *Kaabo Czech*, 0c2023. [Online]. Available: https://eshop.kaabo-official.cz/cs/emobilita/1295-wolf-king-gt-pro-2022-cz-edition-8594176662527.html.

[9] “Segway SuperScooter GT2”, *Segway-Ninebot*, c2023. [Online]. Available at: https://www.segway-store.cz/elektrokolobezky/segway-superscooter-gt2/.

[10] “LAMAX eLander SA50”, *LAMAX*, c2023. [Online]. Available at: https://lamaxshop.cz/cs/emobilita/1694-lamax-elander-sa50-8594175358803.html.

[11] “Veteran Patton, 2,220WH Battery, 3,000W Motor, Suspension”, *EWheels*, c2023. [Online]. Available at: https://www.ewheels.com/product/veteran-patton/.

[12] “Veteran Sherman-S, 3,600WH Battery, 3,000W Motor, Suspension”, *EWheels*, c2023. [Online]. Available at: https://www.ewheels.com/product/veteran-sherman-s/ .

[13] “KS-S22 Pro”, *Kingsong*, August 12, 2022. [Online]. Available at: https://www.kingsong.com/products/ks-s22.

[14] “KS-16S”, *Kingsong*, August 12, 2022. [Online]. Available at: https://www.kingsong.com/products/ks-16s-16-inch-series.

[15] “KS-14D”, *Kingsong*, August 12, 2022. [Online]. Available at: https://www.kingsong.com/products/ks-14d-14-inch-series.

[16] “KS-S19”, *Kingsong*, September 5, 2023. [Online]. Available at: https://www.kingsong.com/products/ks-s19.

[17] “Inmotion V13”, *INMOTIONSTORE*. [Online]. Available at: https://inmotion-store.com/product/inmotion-v13/.

[18] “EGRUS 4.0”, *Sava bicycles*, c2023. [Online]. Available at: https://www.sava-bicycles.cz/products/egravel-4-0?variant=40543100108883.

[19] “THEOS F100 SH 29"/27.5" 825Wh”, *Kellysbike*, c2016. [Online]. Available at: https://kellysbike.com/cs/horska-c989/theos-f100-sh-29-275-825wh-p75239.

[20] “TRANCE X ADVANCED E+ 1”, *Giant Bicycles*, c2023. [Online]. Available at: https://www.giant-bicycles.com/cz/trance-x-advanced-eplus-1.

[21] “Jízdní kolo Škoda eMTB”, *Škoda auto*, c2023. [Online]. Available at: https://eshop.skoda-auto.cz/cs_CZ/jizdni-kolo-skoda-emtb/p/000050212CE.

[22] “Rail 9.9 X0 AXS T-Type Gen 4”, *Trek Bikes*, c2023. [Online]. Available at: https://www.trekbikes.com/cz/cs_CZ/kola/horsk%C3%A1-kola/horsk%C3%A1-elektrokola/rail/rail-9-9-x0-axs-t-type-gen-4/p/41579/?colorCode=orange.

[23] “SOKOR BOSCH CX (625 Wh Teal white pearl)”, *Levit*, c2023. [Online]. Available at: https://www.levit.com/sokor-bosch-cx.

[24] R. G. Ö. Karlı and S. Çelikyay, “Current Trends in Smart Cities: Shared Micromobility”, *Innovations in Smart Cities Applications*, pp. 187-198, March 2022. [Online]. ISSN 2367-3389. Available at: https://www.researchgate.net/publication/359092923_Current_Trends_in_Smart_Cities_Shared_Micromobility.

[25] “Introducing the Bolt 5, our fifth generation scooter”, *Bolt*, MAR 30, 2022. [Online]. Available at: https://bolt.eu/en/blog/bolt-introduces-the-bolt-5-its-fifth-generation-scooter-model/.

[26] “Safer. Smarter. Stronger. Meet the new Bolt 6 scooter”, *Bolt*, JUL 13, 2023. [Online]. Available at: https://bolt.eu/en/blog/introducing-bolt-6-scooter/.

[27] “Number of electric bicycles sold in the European Union (EU-28) from 2006 to 2021”, *Statista*, Dec 18, 2023. [Online]. Available at: https://www.statista.com/statistics/397765/electric-bicycle-sales-in-the-european-union-eu/.

[28] “*Sales volume of electric bicycles in Europe in 2022, by country*”, *Statista*, Dec 18, 2023. [Online]. Available at: https://www.statista.com/statistics/1353317/electric-bike-sales-european-countries/.

[29] F. Küster, “Two bicycle parking spaces per apartment set to become new European norm”, *European Cyclists´ Federation*, 15 Dec, 2023. [Online]. Available at: https://ecf.com/news-and-events/news/two-bicycle-parking-spaces-apartment-set-become-new-european-norm.

[30] “The Bosch eBike Charger: Charge batteries efficiently”, *BOSCH*. [Online]. Available at: https://www.bosch-ebike.com/en/products/charger/.

[31] “BT C01.750.UC”, *BAFANG*. [Online]. Available at: https://www.bafang-e.com/en/oem-area/components/component/battery/bt-c01750uc.

[32] “BT F014.410.C”, *BAFANG*. [Online]. Available at: https://www.bafang-e.com/en/oem-area/components/component/battery/bt-f014410c.

[33] “E-BIKE / Battery Charger”, *SHIMANO*. [Online]. Available at: https://productinfo.shimano.com/#/spec/E-BIKE/Battery.

[34] J. Prokopius, “[Nová kolárna na VUT]”, *VUT*, 02-Jun.-2023. [Online]. Available: https://www.vut.cz/vut/aktuality-f19528/vut-prichystalo-cyklistum-darek-ke-dni-jizdniho-kola-d242159.

[35] “Planning for Micromobility”, *U.S. Department of Transportation*, 0Wednesday, July 5, 2023. [Online]. Available at: https://www.transportation.gov/rural/electric-vehicles/ev-toolkit/planning-micromobility.

[36] “Solární multifunkční dobíjecí stanice TIP / TOP SPOT (mobiliář)”, *LEDEOS*, c2022. [Online]. Available at: https://www.ledeos.cz/solarni-multifunkcni-dobijeci-stanice/.

[37] “EuroVelo Schematic Diagram”, *EuroVelo*. [Online]. Available at: https://en.eurovelo.com/about-us.

[38] “Chargebike”, *Mapy.cz*. [Online]. Available at: https://mapy.cz/turisticka?q=chargebike&x=14.5768567&y=50.0789358&z=7.

[39] “Jezděte, objevujte, dobíjejte”, *Dobiju*, c2021. [Online]. Available at: https://dobiju.cz/.

[40] “Dlouhé výlety na elektrovozítku už nejsou scifi”, *KdeNabiju.cz*, c2024. [Online]. Available at: https://kdenabiju.cz/ .

[41] “Reference”, *LEDEOS*, c2022. [Online]. Available at: https://www.ledeos.cz/reference/.

# Parking house as energy source

Zdeněk Novotný
*Department of Electrical Power Engineering*
*Brno University of Technology*
Brno, Czech Republic
236939@vut.cz

***Abstract*— The article deals with the design of a parking house (PH) as a source of electrical energy with the concept of V2G (Vehicle-to-grid) and it investigates the energy potential of parking houses (PH) with integrated charging stations for electric vehicles (EVs). It focuses on four selected PHs in Brno and analyses different scenarios with respect to charger types, number of parking spaces and available capacity.**

***Keywords—V1G, V2G, electric vehicle, parking house, available capacity***

## I. Introduction

The increasing growth in the number of electric vehicles brings with it new opportunities to harness battery capacity that could provide valuable electricity storage services to the grid. This work focuses on the aforementioned V2G concept, and its associated use as a parking house for energy recovery. V1G technology is described, which could be one of the alternatives to V2G, as V2G requires special types of vehicles. For example, the Nissan Leaf, which was one of the first car models to allow bi-directional charging, has the largest representation.

Mass adoption of V2G will require extensive changes and further developments such as strong acceptance, behavioral changes from consumers, continuous development, digitization of the electric grid, as well as increased collaboration between ecosystem participants such as government. The deployment of this large-scale technology will still take several years, but it is a solution that can offer many benefits to consumers in the future, especially in times of electricity shortages, rising prices and ever-increasing sales of electric vehicles.

## II. V2G

The Vehicle-to-grid (V2G) concept allows energy from the electric vehicle (EV) battery to be fed back into the grid while the car is parked. The V2G concept was first introduced at the beginning of the 21st century to take advantage of the assumption that EVs were becoming increasingly common in society, and therefore large amounts of electrical capacity would be available to provide valuable electricity storage services to the grid. [2] An EV with V2G technology would no longer be a mere means of transportation but could function as a mobile power plant that delivers electricity to the grid whenever it is needed.

Due to the bi-directional transmission of power flow, the necessary operations with the power electronic circuits must be performed to match the type of power supply. A step-down AC/DC converter is used to convert the grid voltage to the appropriate level. In order to transfer the DC power back to the grid, there is a DC/AC step-up converter. A control and monitoring unit is located between the charging station and the transformer to ensure accurate input to the inverter by comparing the reference signal and the inverter output signal. A general block diagram of the V2G structure is shown in Fig. 1. The wiring diagram for charging an electric car shows the basic principles and components needed to charge an electric car.

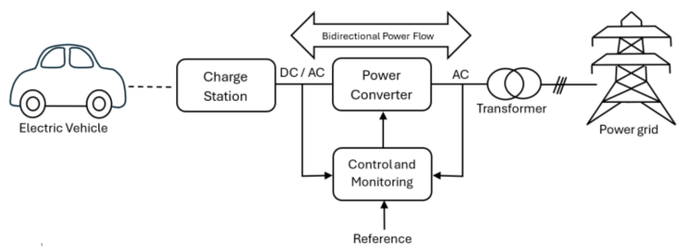



Fig. 1. Block diagram of the V2G structure

Vehicles with V2G technology typically charge when electricity production is higher or when the price of electricity is low (so-called coordinated charging). They sell electricity (discharge) when the demand for energy is high or, conversely, when the price of electricity is high during peak load (namely peak shaving), as shown in Fig. 2. [2] The current scenario in the electrical system of the country is characterized by considerable fluctuations. In the future, efforts can be expected to balance and stabilize the given waveform in order to optimize the operation of the network.

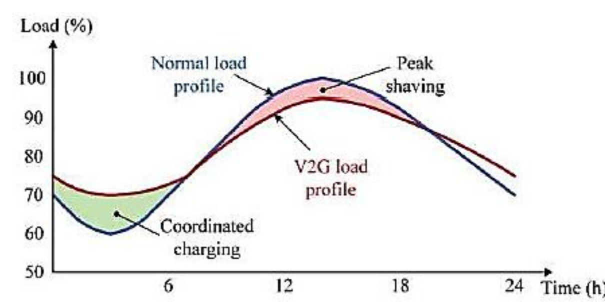



Fig. 2. Coordinated charging and energy peak balancing [2]

Nowadays, we are more often seeing the term V1G or G2V (Grid-to-vehicle). This can be understood as simply charging an electric vehicle from a charging station. [3] It follows that if the car and the charging station have V2G technology, it is possible to send electricity from the battery back to the grid in addition to the normal charging of the car from the grid. This bi-directional communication between EVs and the power grid opens up new possibilities to optimize energy use and increase the reliability of the power system.

V2G technology has the ability to balance calendar (temperature and SoC dependent) and cyclic (depth of discharge and power load dependent) battery aging. Both of these factors affect the rate of battery degradation. V2G can optimize battery health and improve battery condition by 8.6-12.3% in one year compared to convective charging alone. This corresponds to one extra year of operation. [4]

## III. V2X

The V2X (Vehicle-to-everything) concept is a collective term for all concepts that have the capability of bi-directional energy flow. V2X is a generic term where X represents a variable. In addition to charging electric vehicles (V2G), a charged battery can supply energy to the grid (V2G), supply households (V2H) and buildings (V2B) and power appliances (V2L). In this case, the battery of the electric vehicle is used to store locally produced energy or as a backup energy source in case of a possible power failure. The amount of energy transferred from the vehicle depends on the number of grids connected e.g. V2H requires 1 to 3 vehicles, V2B requires 1 to 30 vehicles and V2G requires at least 5 to 50 vehicles. The minimum vehicle power requirements for transmission to each system are shown in TABLE I but may vary based on the type and size of the system. These values are very important to ensure energy efficiency, continuity, reliability, and safety. The power transfer must be smooth, and must not change the voltage, power factor, or frequency of the grid. [1]

TABLE I. VEHICLE-TO-X ENERGY TRANSFER RANGE [1]

| Power Flow | kW | Vehicle-to-X |
|---|---|---|
| Bi-directional | 5–10 | Vehicle-to-home (V2H) |
| Bi-directional | 10–15 | Vehicle-to-building (V2B) |
| Bi-directional | 15–30 | Vehicle-to-grid (V2G) |

## IV. CONCEPTS FOR PARKING HOUSES WITH CHARGING

The studies that have addressed this issue have differed primarily in the virtual environment and the assumed constraints. In addition to the electricity grid, some studies include photovoltaics, power generators (e.g. wind power), or local battery energy storage system (BESS) in their simulations. Different sources also deal with different environmental, technical, and financial issues. The following can be used to control EV charging:

*1) Uncontrolled charging:* EVs immediately start charging at full power when parked. However, this poses technical (frequency and voltage fluctuations and grid congestion) and economic (sub-optimal generation) problems for the electrical system.

*2) Optimised (controlled) charging:* Based on a computer algorithm and an energy flow controller. It tries to charge the battery at the optimal time and rate based on certain criteria (objective function) [6].

Analytical models are used for charge control to perform and optimize the process. These models include, for example:

- Genetic algorithms (GA) - imitating the process of natural selection in biological evolution.
- Mixed integer linear programming (MILP) - a method that combines linear programming with integer programming.
- Mamdani-type fuzzy logic – a method of creating a control system by synthesizing a set of linguistic control rules obtained from experienced human operators. The output of each rule is a fuzzy set. [5]

The studies also differ in the objectives of their analyses, which include - maximizing the use of renewable energy sources (especially photovoltaics), maximizing profits for parking operators and EV users, or minimizing operating costs. Several studies focus on V1G (G2V), and it is common to see V2G, but there are also studies that focus on both concepts.

### B. *V2G estimation method*

Real-time estimation of V2G capacity is essential to ensure effective management and optimal performance of V2G traffic. The estimation is based on many factors and constraints, where the most important parameters are [8]:

*1) Maximum and continuous V2G power:* The ratings of the power converters used in charging and discharging the battery determine the maximum and continuous V2G power. Thus, the nominal battery power can be neglected. To give an idea, the Mitsubishi MiEV has a battery capacity of 80 kWh and a continuous power converter capacity of 2.5 kW. These values are key in determining the total power that an EV group can deliver in a given interval.

*2) Available power V2G:* The primary function of an EV is to satisfy the user's travel needs. Providing charging capability while also providing the vehicle as a means of transportation before the auxiliary V2G function. The available V2G energy depends on the current state of charge, the required SoC, and the minimum required SoC before departure.

### C. *Selection of PH sites in Brno*

Brno has several parking houses (PH) that would be suitable for the analysis of the use of the selected parking house for energy recovery. Only public PHs and those with more than 80 parking spaces are considered. If the V2G concept were to be installed in these PHs, it would prove unrealistic as they are not designed for this type of energy and power load. Therefore, only PHs that will be built in Brno in the future can be considered.

According to the source [9], several buildings are planned: the PH at the Children's Hospital for 400 cars, the PH Dubová with 41 spaces, the PH Juliánov for 160 cars, the PH Park & Ride Bohunice near the University Hospital, the PH Vlkova with 200 parking spaces, the PH Vodova for 441 cars, and the parking house Voříškova for 220 cars. In the next few years, it is also planned to build a PH on Žlutý Kopec near the Masaryk Cancer Institute, which would be used mainly by patients and visitors to the institute and for resident parking in the evening. At the same time, a four-story PH is being built on Šumavská Street, which is expected to be completed in April 2024. It will provide 441 parking spaces and another 222 on the adjacent area under the resident parking regime.

## V. Proposed Scenarios

Individual scenarios will be proposed for selected parking structures in Brno, specifically PH Pinki Park, PH River Park, PH Domini Park and PH Rooseveltova - JD Park. For each PH, the structure of vehicles that are most frequently found in them is selected. The vehicle structure is based on the division of EVs by battery capacity into three categories:

- A - Electric cars with small battery capacity (up to 35 kWh). They are suitable for shorter ranges and urban use, while being more affordable and cheaper to purchase.
- B - Mid-range EVs (around 60 kWh). They are suitable for commuting and longer trips. They have higher comfort and range.
- C - Electric cars with a large battery capacity (over 90 kWh). Suitable for comfortable travel and long distances.

As all of the proposed parking structures are located in the city center, there will be a predominance of medium battery vehicles. In order to adequately simulate future parking scenarios, it was necessary to define the representation of each vehicle category.

During the period, the selected parking houses were visited repeatedly at different times of the day and days of the week. During these visits, data was collected on the representation of vehicles with small, medium and large battery capacities. The data collected was statistically analyzed to determine whether the representation of each vehicle category varied by time of day, day of the week and other factors. The data remains consistent, with only minimal variation, and medium battery capacity strongly outweighs the others, which may change in the future.

The composition of the vehicles was based on realistic parameters, taking into account the forecast of future development, which envisages a gradual decline of internal combustion engines.

Considering the current number of approximately 25,000 electric vehicles in the Czech Republic and considering the socio-economic factors of Brno as the second largest city with growing purchasing power, an increase in the number of electric vehicles in the area is expected. Thus, the dominance of electric vehicles is expected in 2035, reducing the relevance of alternative energy sources such as photovoltaic panels, wind turbines or local battery systems. However, for a more comprehensive analysis and assessment of the energy balance of the car park, RES should still be included in the considerations.

The proposed scenarios will vary depending on the number of fast chargers, the number of AC chargers of different outputs (11 and 22 kW) and the allocation of parking spaces without charging infrastructure, e.g. for vintage cars and hybrids. In the remaining time, options with wireless charging could also be considered. Each PH will have its own scenarios as they vary in total potential occupancy as well as average parking time. Based on the specific characteristics of the Pinki Park PH, with low average standing time, it appears that installing fast charging stations may not be the optimal solution. In this case, only an increased number of AC charging stations with lower power will be addressed, allowing efficient charging during parking time. A tailored approach to scenario development will ensure efficient and meaningful use of available resources and technologies. The output of each scenario will be:

- Available PH capacity,
- Number of one-way and two-way charging stations,
- Transformer size,
- Energy demand response capacity,
- Economic evaluation.

The available charging (V1G) and discharging (V2G) capacity will vary dynamically depending on the hours and days of the week, emergencies (e.g. cultural events in the area) as they contribute to an increase in charging demand and thus affect the available capacity, and on temperature. Temperature has a significant impact on the charging process and EV range. In general, lower temperatures slow down the charging process as chemical reactions are slower. At the same time, the battery temporarily loses some of its capacity, reducing the available range. At higher temperatures, charging is faster and the battery capacity does not decrease as much with temperature.

Compared to other concepts, no other source of electricity such as PV, wind turbine, or local battery energy storage (BESS) is considered. To extend the analysis, RES could be considered. A schematic of the considered EVPL is shown in Fig. 3.

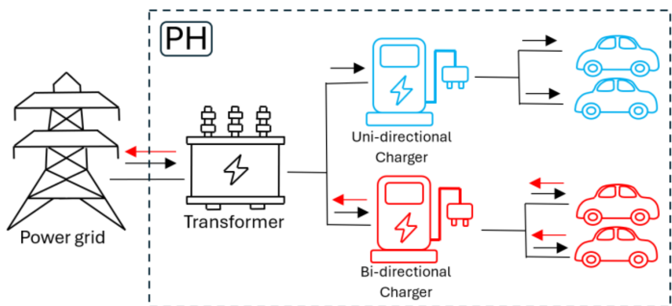



Fig. 3 Schematic of the proposed EVPL with both UD and BD chargers

### A. *V1G instead of V2G*

Besides the V2G technology, one could only consider the V1G variant, which is already widespread. In this case, the electric vehicle could only draw energy from the grid, which would simplify the technical aspects and costs considerably. Instead of building bi-directional (BD) chargers, only uni-directional (UD) chargers would suffice. Given the low number of EVs compatible with V2G technology, V1G in particular represents a solution that is not dependent on the distribution of a specific vehicle type.

Although V1G would not offer all the benefits of V2G, such as grid stabilization and possibility of feeding energy back into the grid, it could still contribute to the development of electromobility. In fact, it would allow to expand the infrastructure of charging stations and thus potentially reduce the cost of acquiring and operating EVs.

### B. *Available capacity*

To find a given capacity per year, one needs to determine the ranges, or simulations, under which conditions vehicles will drive into PH discharged and provide almost no energy to a second marginal condition where they arrive charged and provide their capacity. For a V1G-only concept, the reverse would be true in that if a vehicle arrived discharged it can provide discharge capacity and a charged car will provide almost no capacity. Utilizing the capacity of vehicles parked for longer periods of time represents another potential energy source but requires the implementation of appropriate tariff and incentive mechanisms, which would be one possible scenario. Successful implementation of this strategy could lead to a reduction of car traffic on the streets and the associated positive impacts on the environment and quality of life in cities.

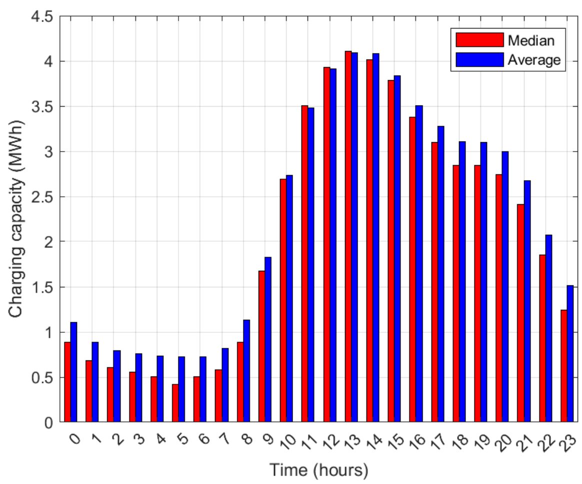



Fig. 4 Average and median charging capacity during the day

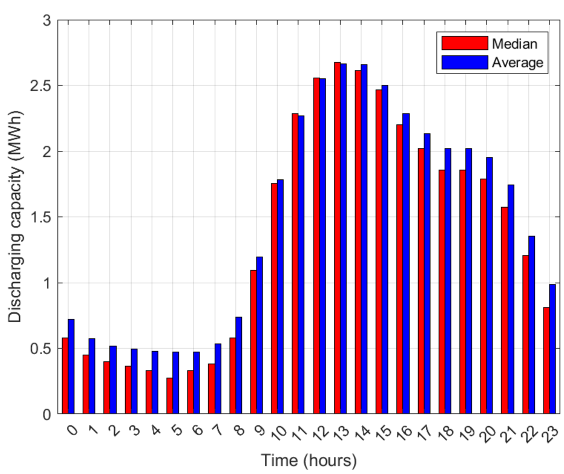



Fig. 5 Average and median discharge capacity during the day

Figures 4 and 5 show the average charging and discharging capacity that could be provided by a PH Domini Park over the course of a year. Marginal conditions have also been set for the 12.15 MWh (V2G) and 17.01 MWh (V1G) of capacity that PH could theoretically provide. The minimum values vary at different hours and reach values of 30-150 kWh at minimum occupancy. Extreme cases are also defined, e.g. when there is no vehicle in the PH. The boundary conditions and average capacities have been determined for all four PHs.

## VI. Conclusion

It should be noted that while V2G represents a promising future for electric vehicles and the electricity grid, the technology is still under development and not yet widely available. V1G is the current standard for EV charging and offers a simple and familiar method. V2G is a more advanced technology with the potential to innovate the way EVs interact with the electricity grid, bringing benefits to both grid operators and EV owners. EVs with V2G can also be used to store energy from renewable sources such as solar and wind power.

Scenarios are outlined for four selected parking houses in Brno. The boundary conditions and average capacity values that could be expected during the day were determined. The specific form of the scenarios varies depending on the specific characteristics of the car park in question, such as total capacity, average parking time and representation of electric vehicle types. Parking structures with integrated EV charging stations represent a significant potential source of energy for the energy grid. The impact on the energy balance of a car park depends on a number of factors, including the type of car park, the type of charging stations, the number of parking spaces and the available capacity. To reach its full potential, V2G technologies need to be further developed and appropriate tariff and incentive mechanisms implemented.

## Acknowledgment

I would like to thank the supervisor Ing. Martin Paar, Ph.D for his professional guidance, consultations, patience and his unique incentives to motivate work.

## References

[1] S. Han and M. A. Acquah, *Grid-to-Vehicle (G2V) and Vehicle-to-Grid (V2G) Technologies*. Basel, Switzerland: MDPI, 2021, pp. 75-79.

[2] K. T. Chau, ELECTRIC VEHICLE MACHINES AND DRIVES. Singapore: John Wiley & Sons, 2015, pp. 14-16.

[3] M. Komárek, "Technologie V2G a její možnost implementace v podmínkách Česka", Diplomová práce, Praha, 2021.

[4] A. Peachey, "V2G charge points can improve the battery life of EVs, project finds", transportandenergy, June 2022. [Online]. Available: https://transportandenergy.com/2022/06/08/v2g-charge-points-can-improve-the-battery-life-of-evs-project-finds/.

[5] MathWorks®, "Mamdani and Sugeno Fuzzy Inference Systems", Accessible on: https://www.mathworks.com/help/index.html.

[6] T. Ma and O. A. Mohammed, "Optimal Charging of Plug-in Electric Vehicles for a Car-Park Infrastructure, " in *IEEE Transactions on Industry Applications*, vol. 50, no. 4, pp. 2323-2330, July-Aug. 2014, doi: 10.1109/TIA.2013.2296620.

[7] A. A. Al-obaidi and H. E. Z. Farag, "Optimal Design of V2G Incentives and V2G-Capable Electric Vehicles Parking Lots Considering Cost-Benefit Financial Analysis and User Participation," in *IEEE Transactions on Sustainable Energy*, vol. 15, no. 1, pp. 454-465, Jan. 2024, doi: 10.1109/TSTE.2023.3307633.

[8] K. N. Kumar, B. Sivaneasan, P. H. Cheah, P. L. So and D. Z. W. Wang, "V2G Capacity Estimation Using Dynamic EV Scheduling," in *IEEE Transactions on Smart Grid*, vol. 5, no. 2, pp. 1051-1060, March 2014, doi: 10.1109/TSG.2013.2279681.

[9] Stavby Brno, "Stavby-plánované-ostatní", Accessible on: https://stavby-brno.webz.cz/stavby-planovane-ostatni/.

# Automatic production line efficiency evaluation and optimization

1st Zdeněk Štěrba
*Department of Control and Instrumentation*
*Brno University of Technology*
Brno, Czech Republic
240451@vutbr.cz

2nd Václav Kaczmarczyk
*Department of Control and Instrumentation*
*Brno University of Technology*
Brno, Czech Republic
kaczmarczyk@vut.cz

***Abstract*—The article deals with the analysis of the production line in terms of overall efficiency. The aim is to define the areas that play a role in overall efficiency and then devise appropriate solutions to improve them. Emphasis is placed on the analytical method Basic MOST designed for work norming, according to which the normalized time of operators is determined. The next part of the paper is devoted to the analysis in order to find the weak points of the line that could be optimized in order to increase efficiency or financial savings. The production line and its analysis is also processed as a discrete event system in Siemens Plant Simulation software, focusing on the data flow and variability of the whole system.**

***Index Terms*—Basic MOST, Yamazumi chart, optimization, Siemens Tecnomatix Plant Simulation, OEE**

## I. Introduction

The work is processed in the AVX company in Bzenec. It is a company focused on the production of connectors. Connectors are mainly installed in control units and active and passive safety elements.

The goal was to select a specific production line and perform analysis and optimization on it. In order to achieve this goal, four problems were identified to focus on. All four problems play a key role in determining overall efficiency. The first task is the norming of operators' work, which is used to identify, assess and evaluate the time consumption of a given operation and process. The next task is the analysis of the line condition, according to which line balancing can then be performed, which is a tool for reducing production time. The last point is to suggest possible improvements based on the previous data obtained to help with overall efficiency and optimization. The discrete event system in Siemens Tecnomatix Plant Simulation is then a chapter in itself. With this tool, it is possible to perform analysis with dynamically changing parameters over time and the whole system is also used as a visual demonstration of a real line.

## II. Basic MOST method

The Basic MOST method is one of the most widely used methods in the MOST family. Every company probably has activities for which Basic Most is a more reasonable and practical tool for measuring work. Other methods in the MOST family, such as Mini MOST or Maxi Most, serve more as supplements to Basic MOST.

This method is structured into two basic sequential models representing the activities necessary for measuring work: General move and Controlled move. The remaining two models were added to simplify the use of hand tools and activities associated with the hand crane. The sequential hand crane model is only used where heavy object handling is performed.

### *A. Basic Properties of General Move*

General move deals with the spatial movement of one or more objects. If an object comes into contact with another object while moving, it is not possible to use the General move model, but rather the Controlled move model.

The General move sequential model consists of fixed activities identified by the following steps:

- Action Distance (A)
- Body Motion (B)
- Gain Control (G)
- Placement (P)

Further, the sequential General move model is divided into three phases, which are shown in the following table:

TABLE I
Phase Of The General Move [1]

| Get | Put | Return |
|---|---|---|
| A B G | A B P | A |

A special time unit TMU (*Time Measurement Units*) is used to determine the predefined times of the partial parts of the operation. The time value in the TMU for each sequential model is calculated by summing the values of the indices and multiplying the sum by 10. [1]

TABLE II
Conversion Ratios Between TMU And Standard Time Units [2]

| | |
|---|---|
| 1 TMU | 0,00001 hours |
| 1 TMU | 0,0006 minutes |
| 1 TMU | 0,036 seconds |
| 1 hour | 100 000 TMU |
| 1 minute | 1 667 TMU |
| 1 second | 27,8 TMU |

Each parameter is determined by observing or visualizing the operator's actions. Accordingly, an index is then assigned to the parameter based on the data card (Fig.1). [1]

| Index x 10 | A Action Distance | B Body Motion | G Gain Control | P Placement |
|---|---|---|---|---|
| 0 | ≤ 5 cm | | | Pickup or Toss |
| 1 | Within Reach | | Light Object<br>Light Objects Simo | Lay Aside<br>Loose Fit |
| 3 | 1-2 Steps | Sit or Stand | Light Objects Non-Simo<br>Heavy or Bulky<br>Blind or Obstructed<br>Disengage<br>Interlocked<br>Collect | Loose Fit Blind or Obstructed<br>Adjustments<br>Light Pressure<br>Double Placement |
| 6 | 3-4 Steps | Bend and Arise | | Care or Precision<br>Heavy Pressure<br>Blind or Obstructed<br>Itermediate Moves |
| 10 | 5-7 Steps | Sit or Stand with Adjustments | | |
| 16 | 8-10 Steps | Stand and Bend<br>Bend and Sit<br>Through Door | | |

Fig. 1. Data Card Basic MOST [3]

### B. Standardisation Of Operators' Work

There are two operators on the line. The role of the first operator is to take the finished piece from the trolley and after a short inspection place it in the prepared blister. This task is carried out for each part separately, i.e. it is carried out three times in one cycle, because there are three pieces in one trolley. The task of the second operator is identical with a different sequence. After the first operator removes the made pieces, the second operator waits for an empty trolley to reach him, in which he places the three pieces, which then travel through all the stations of the line. The second operator's job is then to grab a piece from the prepared blister pack he has on hand and place it on the trolley. As with the first operator, this sequence is performed three times for the three pieces.

The determination of the parameter indexing is the same for both operators. For the determination of time, a general move model was used, which falls under the operators of the activity performed.

$$\text{Operator } 1/2 = A_1 \quad B_0 \quad G_1 \quad A_1 \quad B_0 \quad P_6 \quad A_0$$

$$[(1+0+1+1+0+6)\cdot 3]\cdot 10 = 270\ TMU = 9.72\ s$$

We can see that the *Get* phase contains the index $A_1$. This represents the distance of the movement *Within Reach*. When grasping the fabricated piece, there is no movement of the body in the torso area, so the index $B_0$ is chosen. Since this is a grasp of a light object, the index $G_1$ is used here. In the next phase *Put* the place of placing is at the distance *Within Reach* ($A_1$) and again there is no body movement ($B_0$). In addition, the index $P_6$ is used here to represent the action *Intermediate Moves*. This index is chosen because of the necessary final inspection of the part when the operator wraps the held piece and then places it in the blister. The total time of the operation came out to 270 TMU and when converted to seconds, it is 9.72 seconds.

For the second operator, the parameter indexing is the same as for the first operator. The difference is that the second operator grabs the pieces from the blister and then places them on the cart. The $P_6$ parameter is used again here because the operator must recheck the grabbed piece for gross errors.

## III. Analysis And Optimization Of The Selected Production Line

### A. Description Of The Production Line

The construction of the line is modular, the individual modules are connected into a single unit. The machine operator inserts the unassembled connector bodies into the trolleys, these trolleys are moved along the conveyor belts to the inlet of the line. Inside the line, the trolleys are moved by pneumatic cylinders to the individual working positions. After fitting and testing, the trolleys with the connectors are moved onto the conveyor belt and the operator removes the completed connectors and places them in pallets - blisters. The line is designed to work in automatic mode.

### B. Analysis of the Current State of the Line

This section summarises the key aspects of the overall efficiency of the current line status. These aspects serve to evaluate the status and further serve for possible optimization. The data are shown in the following table (TABLE III).

TABLE III
Key Data For The Line Analysis

| | | |
|---|---|---|
| | trolleys | pcs |
| Number of trolleys/pcs in the trolley | 25 | 3 |
| | t [s] | % |
| Planned production time | 3300 | 100 |
| Total downtime | 240 | 7 |
| Real production time | 3060 | 93 |
| | pcs | |
| Made OK | 319 | |
| Made NOK | 35 | |
| | t [s] | |
| C/T actual | 25,4 | |
| C/T reduction possibilities | 19,5 | |
| C/T indirect calculation | 25,9 | |
| | % | |
| Availability | 92,73 | |
| Performence | 88,30 | |
| Quality | 89,03 | |
| OEE | 72,90 | |

The first information that is mentioned in the table is the number of carts that are circulating on the line along with the pieces (products) on each cart. So we can see that there are 25 carts in total and there are 3 pieces on each cart. This information is crucial for the subsequent evaluation of the OEE because the calculation must take into account that the line cycle is based on trolleys and not on pieces.

The time used is shown next in the table (blue part). We can see that the actual production time here was 93 % of the total measurement and therefore the downtime was 7 %.

The line in the long run produces ± 90 % correct pieces (OK) and therefore 10 % rejects (NOK).

The current cycle time of the line was set to 25.4 seconds (purple part). This information was obtained by direct measurement, where the cycle time for all trolleys on one section was measured. This information is considered to be conventionally correct and is used for further calculations.

The last aspect is the efficiency of the line, which is divided into four parts (red part), namely the Availability, Performance, Quality and then the Overall equipment effectiveness (OEE).

$$Availability = \frac{Real\ Production\ Time}{Planned\ Production\ Time} \tag{1}$$

$$Performance = \frac{Ideal\ Cycle\ Time \cdot Total\ Count}{Real\ Production\ Time} \tag{2}$$

$$Quality = \frac{Good\ Count}{Total\ Count} \tag{3}$$

$$OEE = Availability \cdot Performance \cdot Quality \tag{4}$$

Compared to other production lines, all three subparts of the OEE (Availability, Performance, Quality) came out relatively positive and therefore the line is efficient. If we wanted to focus on the weakest component of these data, it would be to focus on the reduction of downtime and after consultation with the designers, possible improvements could be drawn.

### C. Balancing Analysed Lines

The cycle time of the line is equal to the slowest (longest) part of the line. Therefore, the goal is to determine the times of individual machines. For a more detailed understanding, it is useful to divide the machine operation into individual sub-operations. Because the machine, unlike the operator, is not subject to time variations, it is sufficient to break down the provided image only once. The individual operation times are shown in the Yamazumi chart (Fig.2).

$$Cycle\ time = \frac{Real\ Production\ Time}{Total\ Count} \tag{5}$$

In addition, the graph shows the C/T current (black line) and C/T reduction (red line). It can be noticed that the line speed is uniquely determined by the robot arm. The other parts complete their work in almost half the time as the robotic arm. If there were no robotic arm on the line, the line cycle would be given by the speed of the setup operation (OH1-OH4) and in this case it would be less than fifteen seconds, which is more than ten seconds less than the current state. Furthermore, it can be observed that the operators form the fastest part of the line.

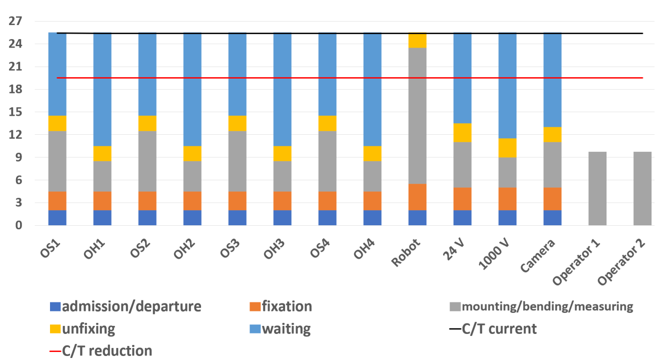



Fig. 2. Yamazumi Chart Of Individual Operations

### D. Possible Improvements

The graph (Fig.2) clearly shows that reducing the robot operation time would have a clear effect on the line cycle time. This design modification would result in a noticeable financial saving in the long run.

Further improvements could be made in better use of operators. The graph (Fig.2) shows that both operators create the shortest time cycle. In this case, the work of the operators is underutilized and the total work converted to the time aspect of one cycle is less than half of the machine cycle time. One possible solution could be to have one operator do the work for both of them. In this case, the operator would have to be experienced enough to do the work fast enough to avoid possible cycle slowdowns. Another solution could be to replace one operator with another robotic manipulator. This solution seems to be the most optimal and the second operator can be reassigned to another station where his time pool will be better utilized.

## IV. Simulating A Production Line Using A Discrete System

The Siemens program, specifically the Tecnomatix Plant Simulation, was used to simulate production lines with a focus on time analysis.

Other possible tools include Arena Simulation Software, which is software from Rockwell Automation. It is based on elements called modules, which are based on the SIMAN language. To show the flow of information, the modules are graphically linked. Another competing tool can be considered as WITNESS, which can assist in predicting bottleneck-related problems in production. Unlike Arena Simulation Software, which links modules graphically, in WITNESS the linking can be done in the form of code. [4]

Efficient logistics is crucial to the success of companies in the context of rising production costs. This tool provides objective criteria for comparing alternative approaches and effectively supports production planning. [5]

This tool is used to simulate and optimize the passage of goods through any logistics system. It allows you to model production plants, lines, transport processes and visualize the flow through complex systems. From the model it is possible to know the behaviour of a system better than in real time. By testing variants, the behaviour of the system can be better understood and conclusions can be drawn based on the information obtained to make various optimizations. [6]

### A. System Description

The whole system represents the real form of the production line (Fig.3). Each station of the system corresponds to a real speed. The total time flow is therefore identical. Workers are assigned work that is equal in time to the final value calculated by the Basic MOST method.

The system is processed dynamically. This means that the user can choose how many trolleys will be in circulation before the simulation starts. The user can also set the number of

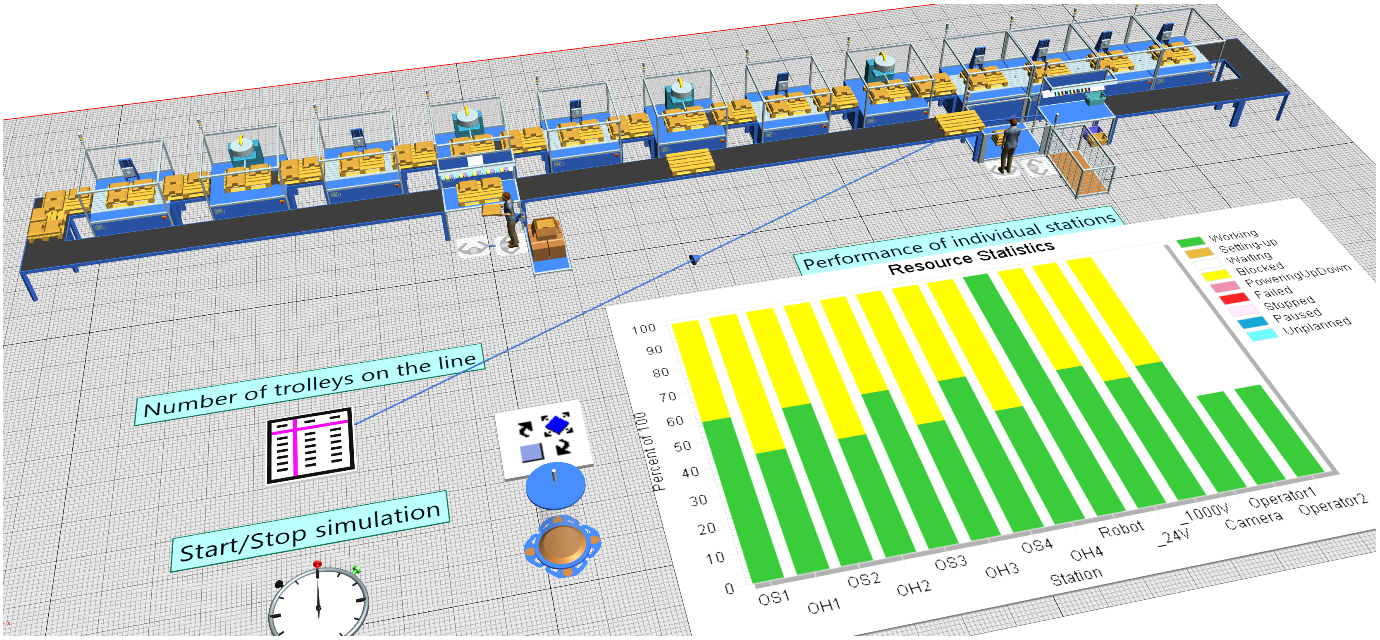



Fig. 3. 3D Model Of The Production Line System

pieces that are on one cart. Thus, this information will also dynamically change the time pool of operators.

Another key feature of this system is a dynamically changing graph that corresponds to the workload of each station, including the work of the operators. Thus, after a long simulation time, this graph becomes identical to the manually created graph (Fig.2). However, the graph in the program has many advantages. One advantage that has already been mentioned is the dynamically changing structure. Since it is also possible to set up faults at individual stations that cause downtime, these are written into the graph over time. We can see that already now the program distinguishes between when a station is stationary because it is waiting for the slowest component of the line (yellow part), and when a station (operator) is waiting because there is no trolley (grey part).

The program can be switched to 2D view (Fig.4), which can be useful if we are interested in the data flow.

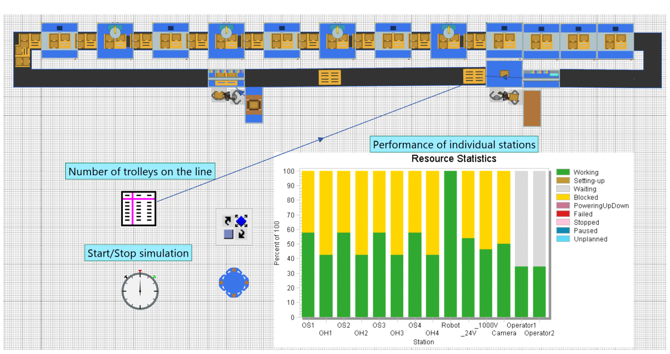



Fig. 4. 2D Model Of The Production Line System

## V. Conclusion

An analysis of the line was performed and it was found that the cycle time of the line is equal to the slowest part, which is the robotic manipulator, which has a noticeably longer cycle time compared to other parts of the line. Furthermore, the overall efficiency of the equipment was determined through the analysis, which came out very positively, and there was no significant problem with any of the sub-aspects related to the overall efficiency. The determination of cycle time was verified by indirect calculation where these values were correlated. The normalized operator time for the activities performed was determined, after which it was found that the operator time pool relative to the line cycle was underutilized. Some solutions were proposed to either reduce the line cycle time or to make better use of operators. The primary idea for better utilization of operators would be to replace one worker with a robotic manipulator to perform its current function. Finally, a discrete event system was designed using the Tecnomatix Plant Simulation tool, which has wide applicability and can be applied to other lines of this type with slight modifications.

## References


[1] K. B. Zandin, in MOST Work Measurement Systems, Third Edition., CRC Press, 2003, p. 548.

[2] L. Stanislav, “Průmyslové inženýrství”, KVS. .

[3] P. López Díaz, “BasicMOST Data Card”, Paulina López Díaz - Academia.edu. .

[4] A. Nikakhtar, K. Yew Wong, M. Hossein Zarei, and A. Memari, “Comparison of Two Simulation Software for Modeling a Construction Process”, IEEE Xplore. .

[5] “Use plant simulation and throughout optimization to improve manufacturing performance”, Siemens Resource Center. .

[6] V. Kaczmarczyk, Průmysl 4.0 - Virtuální výroba. VUT, 2023.

# Digital Twin: Education And Training Purpose

Vojtech Hollý
*dept. of Control and Instrumentation*
*Brno University of Technology*
Brno, Czech Republic
xholly02@vutbr.cz

Ing. Michal Husák
*dept. of Control and Instrumentation*
***Brno University of Technology***
Brno, Czech Republic
michal.husak@vut.cz

***Abstract*—This paper presents a digital twin simulation model with focus on education and testing purpose.For this reason, we have adopted a chain of digitization software tools from Siemens Digital Industries: Product Lifecycle Management (PLM). Firstly, it will be introduced the simulation tools used to digital twin build. Secondly, the software interface that's enabling communication between individual tools. Lastly the contribution concludes with a description of the education model of SCARA manipulator driven by virtual control environment.**

***Index Terms*—Software in the loop, Digital Twin, NX Mechatronic Concept Designer, TIA Portal, SIMIT Simulation Platform**

## I. Introduction

This article focuses on a digital twin created using software from Siemens. The reason of this article is introduce the options and limitations of the simulation programs and create a digital twin of virtual education model. This model involves taking a cup from a conveyor belt and placing it on a table using a SCARA robotic arm with proprietary gripper. Subsequently, the model will be expand with automated task. There will be added elements from university live project called "bartender". Data from digital twin could be connected to project SQL database, and comparing the differences in simulation between the simulation and physical part. Utilizing remote operation and maintenance functions through the digital twin. The program chain forming the digital twin consists of NX Mechatronic Concept Design 1980, TIA Portal V17, S7-PLCSIM Advanced V4.0 SP1, and SIMIT SP V11.0. The main role of simulation tools is to increase efficiency, prevent malfunctions or risks, accelerate product launch and training of professional personal.

NX Mechatronic Concept Designer (MCD) is a software package for CAD/CAM/CAE. It's used to create virtual models of kinematics's and check the design of a machine. Validation of behaviours by the simulation before the manufacturing process. If the MCD is connected to the TIA portal, it is possible to check the functionality of the code.

TIA Portal is for automating industrial processes. It integrates all the important development software's for automation. It means allows us to programmed and maintain of all equipment and systems.

SIMATIC S7-PLCSIM Advanced is a programmable logic controller (PLC) simulator. It is not a part of TIA Portal, but it allows working at the project only with your programmer station without hardware PLC.

The simulation platform SIMIT is used for comprehensive testing of automation applications and for running systems, machines, and processes on a virtual level. Within the SIMIT platform, complex digital twin can be created to uncover design and functional errors and optimize them in a virtual environment. SIMIT SP is used for simulating signals, devices, and device responses. [1]

## II. Digital twin

Since the beginning of Industry 4.0, digital twins have been increasingly utilized and have gained more importance in the industry. A digital twin is a data replica of something physical, such as a machine or a production line. Their advantage lies in their transparency; through this technology, it's possible to monitor, control, maintain, and predict the performance of manufacturing machines, production flows, and cost structures. Digital twin gathers current and historical data, which enables it to detect faults, alert to fluctuations and unexpected situations. Additionally, it allows for adjustments and refinements to be made for improvements. [2]

The main advantages are:

- Engineering - Improving products using CAD programs
- Monitoring and Diagnostics - Current states of physical systems are used for operation monitoring
- Maintenance and Repairs - Carrying out repairs and maintenance via remote access leads to faster problem resolution
- Control - Remote control of parameters in the digital model
- Predictive Maintenance - Predicting potential failures using predictive algorithms and planning repairs before the occurrence of failures
- Training - Training operators and technicians without production limitations
- Development - Implementing modifications and debugging
- Safety - Operation without the need for physical presence

## III. Software in the loop

Software in the Loop (SWL/SIL) is a testing method used to test and verify code in a simulated environment for a specific system. This provides a faster and more cost-effective way to evaluate new features and discover errors compared to live testing. The main advantages of SWL simulation include the

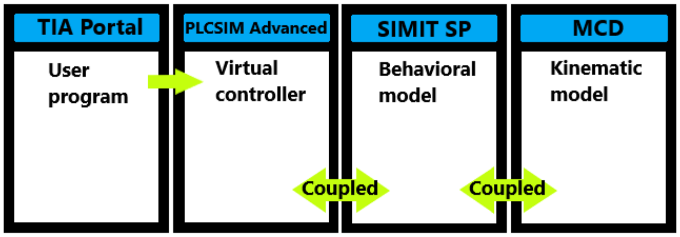



Fig. 1. Representation of the connection of the SWL chain

ability to run on a standard computer without the dedicated real task hardware. SWL does not require real-time simulation because it verifies algorithms. The simulation process can be faster or slower than real-time depending on the complexity of the simulation. The term ”in-the-loop” refers to parts of the software environment, specifically the controlled system, being simulated. While the simulation of a closed control loop isn’t always necessary, as some systems being tested, especially in module testing, don’t require closed control loops. An illustration of SWL for an automated task chain is shown in Fig. 1. [2]

## IV. Couplings

Communication between SIMIT SP and other programs is facilitated by built-in connectors imported into the program. SIMIT SP contains three different types of connectors: connection to a real, emulated SIMATIC controller, or connection to an external partner. If existing connectors are insufficient, it is possible to program additional connectors and integrate them into SIMIT SP. Expansion is created using the SIMIT API interface for external connectors. External connectors cannot be added to the SIMIT SP Demo version. When creating a connection, it is recommended that the programs between which the connection is being established are run in administrator mode.

In the section *A. - E.* are examples of built-in connectors and in the section *F. - K.* are examples of additional connectors.

### *A. PLCSIIM Advanced*

The communication between PLCSIM Advanced and SIMIT SP is facilitated by the PLCSIM Advanced connector. This connector ensures the cyclic exchange of input and output data with the CPU into which the program from TIA Portal is loaded. To establish this communication, the following programs need to be installed: SIMATIC S7-PLCSIM Advanced, TIA Portal, and the TIA Portal Openness program. The PLCSIM Advanced connector is suitable for simulating isochronous mode. Otherwise, it is recommended to use PLCSIM connectors, virtual controllers, or SIMIT Unit. [3]

### *B. Virtual Controller*

The Virtual Controller simulates the response of a SIMATIC controller type S7-300 or S7-400. A user PLC program is loaded into the virtual controller. The simulation of the process is directly connected through input and output process images. Up to 32 virtual controllers can be used in the simulation, and they can be distributed across multiple computers if the connections are based on IPv4 addressing space. When automation programs are run on MS Windows with a virtual controller, precise timing response cannot be guaranteed because Windows is not a real-time operating system. [3]

### *C. PLCSIM*

PLCSIM allows you dynamic data exchange between SIMIT SP and PLCSIM. A prerequisite for this exchange is the installation of software on the same computer. PLCSIM have to be launched before starting the simulation in SIMIT. If PLCSIM is terminated during the simulation, the connection will be severed and cannot be restored online. To establish a PLCSIM connection, PLCSIM software version 5.4 SP5 or higher is required. In this connection, only input and output signals defined by configured hardware can be exchanged. [3]

### *D. Shared memory*

SIMIT SP utilizes Shared Memory (SHM) connections for communication with other applications, supporting up to 32 connectors. This connection is universal and powerful. Input signals from SIMIT are written into the memory area, while output signals from SIMIT SP are read from it. The SHM connector shares configuration limits with the Plant Simulation connector. For instance, if 12 SHM connections are created, it’s possible to create 20 Plant Simulation connections. Using the SHM connection, software like SIMIT can be linked with Matlab, for example. [3]

### *E. Mechatronics Concept Desiner*

This connection is used to debug the 3D kinematic model simulation in NX Mechatronic Concept Design (MCD). Communication between SIMIT SP and MCD is done using SHM via the built-in connector. SIMIT SP sending and receiving signals from NX MCD and application must allow the connection by pressing the ”Allow connection” button in MCD. This button maintains its state after being pressed once. If the button is selected in multiple instances of NX MCD, the connection will be established with any of them. The exchange of simulation data occurs only for signals whose name and type match in SIMIT SP and NX MCD. The ”Isochronous” mode synchronizes the simulation step in MCD application with the SIMIT SP. The ”Time slice” specifies the cycle during which calculations are performed and data is exchanged. [4]

### *F. PLCSIM (TIA Portal)*

PLCSIM establishes the connection between SIMIT SP and the emulated SIMATIC S7-1200 controller. The integration of the emulated controller into the simulation environment is done using the Software-in-the-Loop approach. The emulation of SIMATIC S7-1200 is managed by the S7-PLCSIM software integrated into the TIA Portal. [4]

### G. SIMOTION – SIMOTION Advanced

The SIMOTION connector facilitates communication between SIMIT and SIMOTION Advanced. SIMOTION Advanced simulates one or more SIMOTION D CPUs in combination with the behavior model in SIMIT. SIMOTION Advanced is focused on executing integration scenarios and system testing. It tests the PLC program, verifies machine behavior, and tests failure scenarios. When combined with SIMIT and NX Mechatronic Concept Designer, SIMOTION Advanced includes components for commissioning the virtual machine into operation. [4]

### H. SINUMERIK

The project is located in SIMIT, and the SINUMERIK simulation runs in Create MyVirtual Machine / Run MyVirtual Machine. The connector to SIMIT is called "Create MyVirtual Machine," which serves as an external connector, acting as a linking tool for visualizing the mechanical model and computer numerical controller (CNC) like: milling machines and lathes. It can be replaced with "Run MyVirtual Machine." [4]

### I. ABB

Connection to ABB Robot Studio is achieved through the Shared Memory. It's an embedded feature within ABB Robot Studio for connecting internal signals via SHM. Data exchange can only occur if SIMIT and ABB Robot Studio are located on the same PC. [4]

### J. Fanuc

The connection between SIMIT and Fanuc ROBOGUIDE is established through the Fanuc robot interface (FRRJIF.DLL). To use ROBOGUIDE together with SIMIT, it is necessary to install the FANUC Robot Interface (Runtime). [4]

### K. TwinCat ADS

The external connector enables the integration of PLC Beckhoff TwinCAT 3 into SIMIT simulations. The ADS-Interface exchanges data between SIMIT and TwinCAT PLC during simulation. The application requires TwinCAT 3 PLC as a communication partner and cannot be used with older TwinCAT 2. TwinCAT 3 installation on the same computer as SIMIT SP is a prerequisite. [4]

## V. AUTOMATED TASK

The digital twin consists of programs such as NX Mechatronic Concept Designer 1980, used for kinematic, physics, and colision validation; TIA Portal V17 for PLC control logic validation; and SIMIT SP V11.0 for sensors and drive simulation. Within the SIMIT project, couplings are established: Mechatronics Concept Designer coupling facilitates communication between SIMIT SP V11.0 and NX Mechatronic Concept Designer 1980, while the PLCSIIM Advanced coupling ensures communication between SIMIT SP V11.0 and SIMATIC S7-PLCSIM Advanced V4.0 SP1, where the program from TIA Portal V17 is loaded. Upon initiating the simulation in SIMIT SP V 11.0, a CPU is automatically created in the SIMATIC S7-PLCSIM Advanced V4.0 SP1 program.

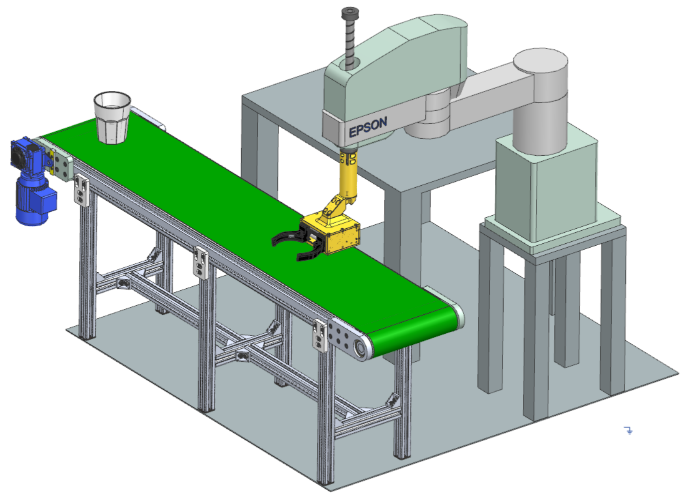



Fig. 2. The assembly in NX Mechatronic Concept Designer 1980

In NX Mechatronic Concept Designer, we control the axes of the SCARA robot and the movement of the conveyor using speed control. The opening and closing of the gripper are controlled by signals generated by the gripper function, as well as the necessary movement to grip and release the object being moved. The virtual twin consists of components that are set up according to need for either rigid or collision bodies. The program automatically sets the weights of the elements. Properties of the elements are configured among themselves using joints. The representation of the virtual twin can be found in Fig. 2.

In the SIMIT SP program, connectors are configured to ensure proper data exchange. In the chart section, connections are established between signals from TIA Portal and NX Mechatronic Concept Designer using macros, which are simpler to use like templates. Examples of macros are shown in Fig. 3. Another macro is used to simulate of an asynchronous motor controlled by a frequency converter for conveyor. Fig. 3 depicts a simulation of a servo motor with a rotary encoder and conversion of input and output values to values used by the programs.

In TIA Portal, in the device configuration section, the CPU SIMATIC S7-1500T and the control unit SINAMICS S120 CONTROL UNIT CU320-2 PN are selected. Additionally, in the hardware catalog, we define the module and telegrams through which communication occurs. The control of automated components is performed using technological objects for motion control. The Speed axis technological object is designated for controlling the conveyor motor, while the Positioning axis is used for controlling the SCARA robot axis. In the kinematics technology object, the robot's SCARA kinematics axes are configured, as well as its geometry.

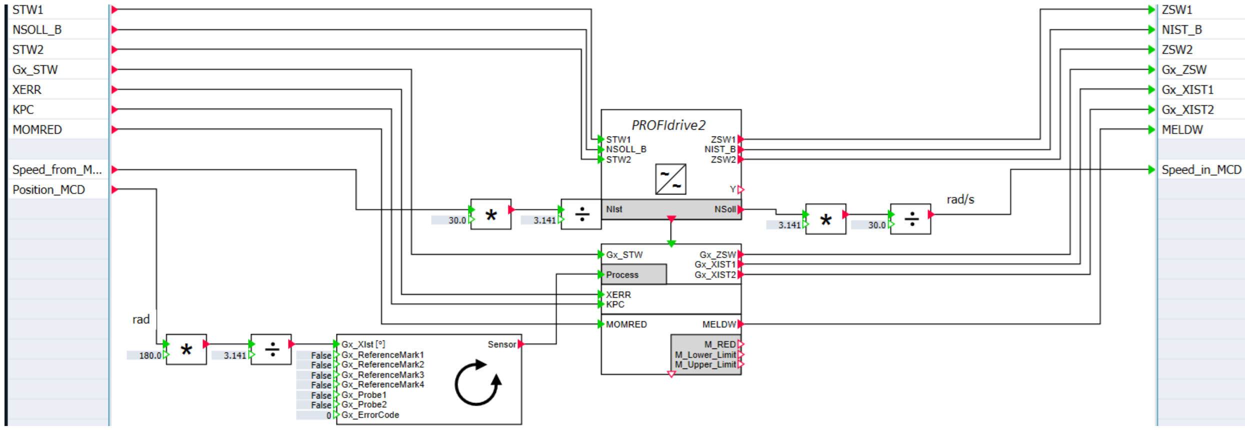



Fig. 3. SIMIT SP servo motor of SCARA robot

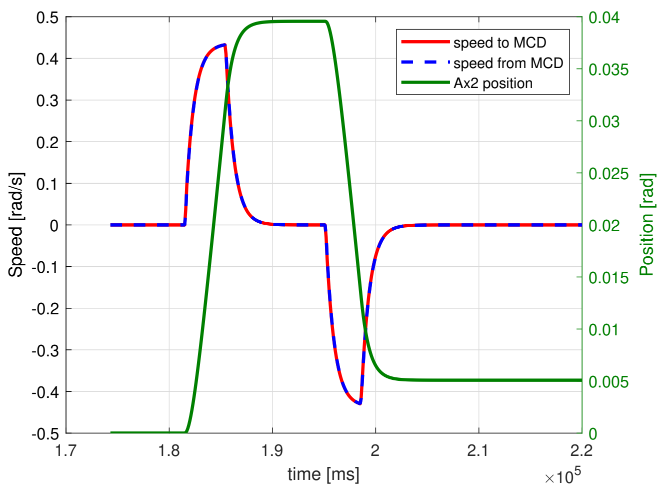



Fig. 4. Simulation process of the SCARA robot's rotary axis.

## VI. Automated Task - Monitoring

In the SIMIT SP V11.0 monitoring section, it is possible to create a trend in which we can monitor the progress of the signal. SIMIT does not allow editing the created graph, therefore the values are exported and subsequently processed in Matlab. Fig. 4 shows the behavior of the second rotational axis of the SCARA robot. The green line represents the robot rotation position value in radians, the blue line represents the robot velocity in degrees per second from the NX MCD, and the red line represents the feedback velocity value in degrees per second from the NX MCD.

## VII. Conclusion

We have created an automated task in which we developed a digital twin using programs such as NX MCD 1980, TIA Portal V17, S7-PLCSIM Advanced V4.0 SP1, and SIMIT SP V11.0. Thanks to this integrated chain, we were able to test control mechanisms and debug errors in the automated assembly line without the need to create and test on a physical model. Presented simulated model could be adopted by digitalization courses at Brno University of Technology.

## VIII. Acknowledgment

The completion of this paper was made possible by the grant No. FEKT-S-23-8451 - "Research on advanced methods and technologies in cybernetics, robotics, artificial intelligence, automation and measurement" financially supported by the Internal science fund of Brno University of Technology.

## References

[1] Virtual commissioning and operator training with SIMIT. Online. SIEMENS. From: https://www.siemens.com/global/en/products/automation/industry-software/simit.html. [cit. 2023-12-23].

[2] HUSÁK, Michal. Virtual Twin For Testbed Industr. Online. In: Proceedings I of the 26st Conference STUDENT EEICT 2020: General papers. TECHNICAL UNIVERSITY BRNO, Faculty of Electrical Engineering and Communication Technologies, 2020, s. 234-237. ISBN 978-80-214-5867-3. From: http://hdl.handle.net/11012/200566. [cit. 2024-03-01].

[3] SIMIT Simulation Platform (V11.1). Online. 2023, s. 1024. From: https://cache.industry.siemens.com/dl/files/650/109823650/att 1152291/v1/SIMIT enUS en-US.pdf. [cit. 2023-12-24].

[4] Robotic Simulation with PLCSIM Advanced – SIMIT – NX-MCD Information. Online. 2021, s. 22. From: https://cache.industry.siemens.com/dl/files/246/109795246/att 1059269/v2/109795246 SIOS RobotVRCcouplings Info en.pdf. [cit. 2023-12-24].

[5] SIMIT TwinCat Coupling. Online. 2022, s. 17. From: https://cache.industry.siemens.com/dl/files/950/109780950/att 1140977/v1/SIMIT TwinCAT3 Coupling v2.3.0.0.pdf. [cit. 2023-12-24].

# AAS Interpreter for the Testbed Industry 4.0

1st Luboš Chmelař
*Department of Control and Instrumentation*
*Faculty of Electrical Engineering and Communication, BUT*
Brno, Czech Republic
240355@vutbr.cz

2nd Ondřej Baštán
*Department of Control and Instrumentation*
*Faculty of Electrical Engineering and Communication, BUT*
Brno, Czech Republic
xbasta02@vut.cz

***Abstract*—This paper describes an active element of the manufacturing process, the AAS interpreter, which implements elements of distributed control and horizontal integration to help achieve the goals of the Industry 4.0 initiative. For context, the text introduces the testbed for Industry 4.0 within which the AAS interpreter is integrated. Furthermore, the paper focuses mainly on the core functions and processes of the device itself.**

***Index Terms*—Smart glass, Industry 4.0, NFC, OPC UA, Asset administration shell**

## I. Introduction

The Industry 4.0 initiative and digitalization continues to resonate in the world of industrial automation and offers modern alternative ways to control production. One of the key principles of this initiative is the possibility of distributed control of production by the element that has the greatest interest in production, i.e. the product. For educational and research purposes, the Institute of Industrial Automation of BUT has created a physical model of a production facility, Testbed I4.0, which is shown in Fig.1, which attempts to exploit and present various principles of this initiative. One of the areas that the testbed deals with is a distributed product management. For a product to be able to produce, two conditions must be met: 1. The product must be able to store information about itself and about the production process; 2. The product must be able to interact with its environment. This paper discusses in more detail a variant where an NFC chip with memory, placed on a container for the product, is used to store production data and monitor production progress. A custom embedded device with an NFC reader, the interpreter, which is used for interaction between the product and the production systems, is placed at appropriate locations within the production. It is able to virtualize the production instance of the product and act as an agent on its behalf within the current production. The specific product connected to the interpreter device is then implemented by the so-called Asset administration shell (AAS) which uses the OPC-UA standard as a communication protocol.

### *A. Motivation to create an AAS interpreter*

In order to move from Industry 3.0 to Industry 4.0, it is necessary to decentralize, which in our case means decentralizing the management of the production process from a traditional MES solution to its decentralized instances ensuring the production of one specific product. For this purpose, we will use the AAS interpreter which uses the OPC-UA protocol to interact with the production environment.

Ideally, each product manages its own production, but this is very costly as each product would have to be equipped with expensive computing equipment. Alternatively, the product may carry the data of its production instance, with a limited number of interpreters at appropriate locations managing the production.

As a way to carry the data of the production instance, an NFC tag placed on the bottom of the product container was chosen, in the memory of which the information about the product and the steps of its recipe will be stored.


Fig. 1. Testbed Barman I4.0 [2]

## II. Testbed Barman I4.0 on DCI

The I4.0 testbed project consists of a workbench in a closed cage. Underneath this space there is an equally large area that is dedicated to the supporting elements of IT technology. On the workbench there are several autonomous units(cells). These consist of an external skeleton of aluminium profiles and an internal system. The cells can be divided into two groups - production cells, where there is a specific category of storage, and transport cells. At the same time, at the entrance of each cell there is an embedded device with an NFC reader to serve as an AAS interpreter. [1] The goal of this paper is to introduce the basic functions and processes that are handled by this interpreter.

### *A. Interpreter hardware*

The interpreter is consists of an embedded device that uses the Espressif ESP32 S3 SoC micro controller as its core. Communication with the network is provided by an Ethernet interface that uses a physical layer with the W5500 circuit. The entire device is powered via an Ethernet connector using the POE standard. For interaction with the product, an NXP PN532 integrated circuit is used, which together with the antenna creates the NFC reader.

The device is housed in 3D printed plastic box and includes a surface for placing the product container, under which the NFC transciever antenna is located. The device is placed at the entrance of each autonomous cell, where the product is allowed to activate an instance of its control and interact with its environment.

### *B. Asset administration shell(AAS)*

This is a digital shell based on the concept of the Industy 4.0 component. This model is used to better describe cyber-physical elements, and it describes the communication between virtual and cyber-physical objects and processes, all in real time. [5]

The shell describes the object in certain sub-models (e.g. identification, communication, security, energy efficiency, etc.). This shell practically includes the product instance, which has to be interpreted at the moment when the product has to interact with its environment. [6]

### *C. OPC Initiative(Openness, Productivity, Collaboration)*

To simplify and link communication protocols in the industry, an OPC working group was formed to create a standard that would collect data from devices and pass it to a higher-level system. [7]

Because of the different OPC architectures, the OPC-UA protocol was created to unify the different architectures and increase security. In the early days, security was greatly underestimated. [8]

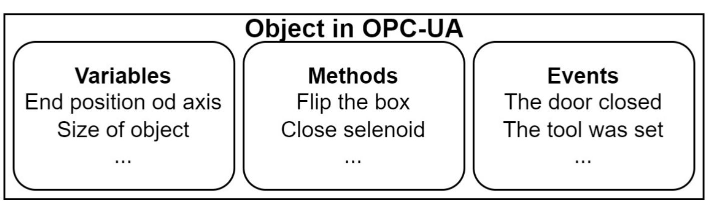



Fig. 2. Object in OPC-UA [7]

## III. Implementation of Industry 4.0 principles using the AAS interpreter

The deployment of the Industry 4.0 initiative is crucial for the functioning of the modern production process. The transition is accompanied by a few basic ideas that need to be followed in order for the testbed to be considered an I4.0 demonstrator.

### *A. The fundamental ideas of I4.0 required for process transformation*

Source [3] defines 6 basic ideas for the inclusion of a system in I4.0. They are:

1) Interoperability - This is the ability of I4.0 elements to communicate and cooperate with each other. The AAS interpreter allows products to be integrated into the communication process.
2) Virtualization - A virtual model of a smart factory can be created and real data can be applied to it. The AAS interpreter virtualizes an instance of the product when it is occupied by the product.
3) Decentralization - Decision-making capability moves from the master element to the level of individual machines. Thanks to the interpreter, control is performed directly by the product at its location.
4) Real-time operations - There is a need to respond to requests immediately with zero delay system response. The producer can assert their interests in real time through the AAS interpreter.
5) Service orientation - In I4.0, the enterprise must focus on service provision (e.g. service) as well as service purchasing (e.g. cloud services). The AAS interpreter allows the product to request services from autonomous cells.
6) Modularity - The overall system in I4.0 must be able to adapt to new requirements and changes. Modularity serves us to easily add new technologies. The AAS interpreter is a generic device that allows to connect the different participants in a standard way, thus ensuring modularity of the overall system.

At the same time [4] describes the basic characteristics of the difference between Industry 3.0 and I4.0. One of these differences is horizontal interconnection, which can be understood as the interconnection of business chains. It controls how the material gets from the supplier, through the production process, to the distribution of the product to the customer. In I4.0, the horizontal integration extends beyond the boundaries of the company, compared to I3.0. For example, it is possible to link all bussiness chain links (from suppliers, where outages can be responded to, to customers, where after sale service can be ensured). This change in horizontal integration makes it possible to respond to customer requirements in real time.

In our case, the AAS interpreter transmits information from the customer who orders the intended drink. All the while, various interpreters monitor the status of the recipe and perform the actions necessary to complete the recipe in sequential steps. At the same time, the interpreter is able to react to production problems and suggest solutions (occupancy of production cells, unavailability of production cells).

## IV. NFC Tag as glass memory

In order for the product to carry the data that represents its instance, it was necessary to choose the appropriate technology. NFC technology is a special category of RFID (Radio frequency identification) where the frequency used is 13.56 Mhz(high frequency band). At the same time, the emphasis is on short distance usage. [9]

A special category of NFC cards that we will use is the NFC tag. This is a sticker that contains a micro processor and an appropriate antenna. We use this solution as a carrier for an AAS instance.

As a suitable NFC tag we chose the MIFARE® Classic product from NXP. It is an ideal carrier of information, thanks to its large memory capacity of 1 kB. The minimum number of writes from the manufacturer is 100,000. This can also be considered sufficient as far as data integrity is concerned. It can be expected that problem of a different kind will occur with the glass before the write count limit is reached. [10]

## V. Data representation in the product data store

The data structure on the NFC tag is shown in Fig. 3. The diagram is shown in general, i.e. it does not contain all information specifically. It can be divided into two parts, the glass information and the recipe part.

The glass information contains basic information about the glass identification number, the number of drinks produced, and the current recipe step.

Recipe steps work basically like a linked list of items, i.e. each step of the recipe refers to the next. The above mentioned method is chosen because of the possibility to add a recipe step without having to scroll through all the following steps, therefore saving the number of memory entries.

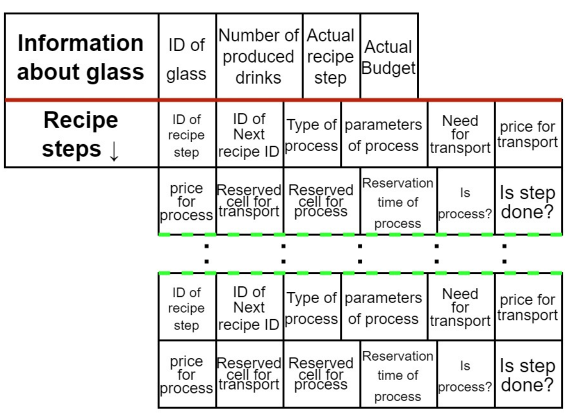



Fig. 3. Basic data structure on the NFC tag

The recipe step defines the type of production process and its parameters. This information is already present from the moment the recipe is uploaded to the smart glass. Other parameters are the properties of the currently executed recipe step. The selection of data to be written to the NFC tag must be carefully selected with respect to the maximum number of writes to the NFC tag, i.e. it is better to store only actual production data instead of analytical data (How long did the production take, etc.).

### A. *Ensuring data integrity*

Maintaining data integrity between the NFC tag and the micro processor is a very important condition for ensuring proper operation of the interpreter. Integrity in general is maintained in the same way as in Fig. 4. It is necessary to ensure integrity so that the glass can be removed from the reader at any time, and after reattaching it, it is possible to find out again in what state the glass is.

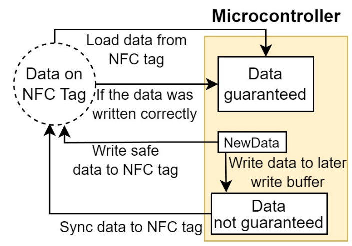



Fig. 4. Data integrity assurance representation

Integrity assurance is provided by a group of commands:

1) LoadData - Load data from the NFC tag into the micro controller's memory and mark this data as integrity assured.
2) Write - Data written this way is written to the internal write data buffer; this data is without guaranteed integrity now.
3) Sync - Buffer of data without guaranteed integrity is written to the NFC tag and then read again. If the written and read data match each other, the micro processor also writes them to the memory with data with guaranteed integrity.
4) WriteSafe - If we like to write to the NFC tag right away, we will use this command. The data will be written and verified immediately. If they match each other, the corresponding data with guaranteed integrity will be replaced in memory.

## VI. The process of interpretation

The process of performing activities within the recipe life cycle is simply described (without crisis situations) in Fig. 5. Throughout the recipe, we assume that the glass has a budget for the entire recipe, according to which we choose which cell to select as the winning cell for production execution.

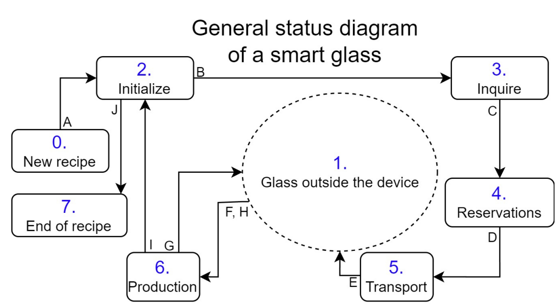



Fig. 5. Interpretation state diagram

The individual states of the state diagram are described below:

*0. New recipe:* This is the state when a new recipe is loaded into the glass. This step will most often be performed in a storage cell.

*1. Glass outside the device:* We get to this state if the glass disappears from the interpreter. At the same time, this state is used to decide the next state the state machine will enter.

*2. Initialize:* This state is used to move to the next step of the recipe and get all the necessary addresses to the cells that we will need, or think we might need, in the current step of the recipe. The Local Discovery Server will provide us with the necessary addresses. At the same time, crisis situations will be handled here (e.g. no transport cell found). If such a situation occurs, the AAS interpreter can intervene and create a sequence of recipe steps to restore the glass to its original state (Empty, clean, etc. ).

*3. Inquire:* In the third state, we evaluate all price and time offers by cells offering the processes needed for the current step, i.e. at what price and at what time they can offer production. If we are in a hurry to produce, we will prefer more expensive cells, but ones that can do it in a closer time. All negotiation is done via OPC-UA protocol methods.

*4. Reservations:* To start a reservation, we must first verify with all inquired winning cells whether their offer is still valid. If any of the cells reply that it is no longer valid, we need to inquire again. If all members answer in the affirmative, they are notified via OPC-UA methods that we are reserving them for the parameters of their bid.

*5. Transport:* In the current step, we check if it's already time to book. If so, we still need to check if someone has moved the reservation time for the winning cell. If the answer is yes, we need to save all the data on the NFC tag. Next, we need to notify the current cell that we are releasing it and the glass will be transported. In the last step, we call the method of the transport cell that the transport can take place.

*6. Production:* We can reach the production state in two cases:

1) We have just arrived at the production cell after transport.
2) The production process in the production cell has just taken place.

In case 1, we need to notify the current production cell that we have appeared in it and therefore it cannot serve anyone else. We change the state to "the transport has taken place", store the data in the NFC tag and notify the production cell that it can execute the production process. In some cases, the glass may disappear from the NFC reader.

If production has just taken place, we ask the production cell if the production went well. If the cell replies that everything went as planned, we set the recipe step to "completed".

*7. End of recipe:* There is a finite state in which we wait until we get a new recipe again.

## VII. Conclusion

The comprehensive AAS interpreter solution described in the article allows us to change the view of the entire production system as a whole. Compared to the traditional approach, it is no longer necessary to centralize all control processes within a single MES system, which then becomes the bottleneck of the entire solution. The proposed AAS interpreter solution allows the MES system tasks to be delegated to individual products that can then act autonomously.

In practice, this type of control can help to optimise production, increase productivity, flexibility, and ultimately resilience of production systems. Algorithms enabling distributed control are being further developed in this project, especially the part of the algorithm providing service negotiations, the so-called bidding.

## References


[1] KLEMENT, Petr, 2023. *Algoritmy řízení výroby pro demonstrátor Průmyslu 4.0*. Online, Bakalářské práce, vedoucí Ing. Václav Kaczmarczyk, Ph.D. Brno: VUT v Brně. Available at: https://www.vut.cz/www_base/zav_prace_soubor_verejne.php?file_id=253820.

[2] BOCH, Jan, 2022. *Autonomní sklad kapalin s chladicím systémem*. Online, Bakalářské práce, vedoucí Ing. Michal Husák. Brno: VUT v Brně. Available at: https://www.vut.cz/www_base/zav_prace_soubor_verejne.php?file_id=241380.

[3] TECHNOLOGICKÉ CENTRUM AV ČR, 2017. *Technology brief: Technologie pro Průmysl 4.0*. Online. Available at: https://vedavyzkum.cz/z-domova/technologicke-centrum-av-cr/technology-brief-technologie-pro-prumysl-4-0.

[4] PÁSEK, Jan. Digitální transformace průmyslu. Online. S. 21-24. Available at: https://www.vutbr.cz/www_base/priloha_fs.php?dpid=183201&skupina=dokument_priloha.

[5] ZEZULKA, František; MARCOŇ, Petr; VESELÝ, Ivo and SAJDL, Ondřej, 2016. Industry 4.0 — An Introduction in the phenomenon. Online. *IFAC-PapersOnLine*. Roč. 2016, č. 49, article 25, s. 8-12. Available at: https://doi.org/10.1016/j.ifacol.2016.12.002.

[6] DE LEEUW, Valentijn and BARTOŠÍK, Petr, 2019. Industrie 4.0 Asset Adminstration Shell: koncepce a využití v praxi. Online. *Automa*. Roč. 2019, č. 8-9, s. 73-75. Available at: https://www.automa.cz/cz/casopis-clanky/industrie-4-0-asset-adminstration-shell-koncepce-a-vyuziti-v-praxi-2019_08_0_12444/.

[7] SCHWARZ, Michael and BÖRCSÖK, Josef, 2013. A survey on OPC and OPC-UA: About the standard, developments and investigations. Online. In: *XXII International Symposium on Information, Communication and Automation Technologies, ICAT*. Sarajevo, Bosnia and Herzegovina: IEEE, p. 1-7. Available at: https://doi.org/10.1109/ICAT.2013.6684065.

[8] BRUCKNER, Dietmar; STĂNICĂ, Marius-Petru; BLAIR, Richard; SCHRIEGEL, Sebastian; KEHRER, Stephan et al., 2019. An Introduction to OPC UA TSN for Industrial Communication Systems. Online. *Proceedings of the IEEE*. Roč. 2019, č. 107, s. 1121-1131. Available at: https://doi.org/10.1109/JPROC.2018.2888703.

[9] MADLMAYE, Gerald; LANGER, Josef; KANTNER, Christian; SCHARINGER, Josef and SCHAUMULLER-BICHL, Ingrid, 2009. Risk Analysis of Over-the-Air Transactions in an NFC Ecosystem. Online. In: *International Workshop on Near Field Communication, NFC*. Hagenberg, Austria: IEEE, p. 1-6. Available at: https://doi.org/10.1109/NFC.2009.17.

[10] *MF1S50YYX_V1: MIFARE Classic EV1 1K - Mainstream contactless smart card IC for fast and easy solution development*, 2018. Online. In: NXP B.V. NXP. Available at: https://www.nxp.com/docs/en/data-sheet/MF1S50YYX_V1.pdf.


Dáváme smysl věcem, na kterých záleží.*

# Comparison of vowel formants with string and wind instrument formants

1st Dominik Steiner
*Faculty of Electrotechnics and Communications*
*Brno University of Technology*
Brno, Czech republic
238316@vut.cz

2nd Ondřej Jirásek
*Faculty of Electrotechnics and Communications*
*Brno University of Technology*
Brno, Czech republic
jirasekondrej@vut.cz

***Abstract*—This article deals with the analysis and subsequent comparison of the formant areas of sung vowels (soprano voice) with the formants of string (violin) and wind (flute) instruments. The article is primarily for musicians and arrangers for compositional and musical artistic purposes. However, it will also help linguists regarding the analysis of formants and resonant cavities for phonetic and phonological researches.**

***Index Terms*—formant, vowel, frequency, FFT, LPC, DFT, pitch, tone**

## I. Origin of timbre in human speech

Each musical instrument consists of an exciter, an oscillator, a resonator and a radiator. In the vocal tract, the excitatory function is performed by the air stream, the resonator function of the vocal tract and the excitatory function of the resonant cavity system of the vocal tract. The basic resonant space is the vocal tract tube, which can be represented as a semi-closed tube closed at one end by the vocal cords and opened at the other end by the mouth. The intrinsic resonance of the semi-closed tube is determined by the formula:

$$f_n = \frac{(2n+1)c}{4l}, \quad n = 0,\, 1,\, 2,\, ..., \tag{1}$$

where $f_n$ corresponds to the natural resonance frequency of the semi-closed tube, $c$ to the velocity of sound (*344* m/s at room temperature) and $l$ is the length of the tube (the average length of he vocal tract is 17 cm). Frequencies of self-resonances of the vocal tract are at odd multiples of the fundamental frequency. [1]

## II. Vowels

The Czech language, like every other language, consists of vowels and consonants. In written Czech there are simple vowels *a, e, i, o , u* (and their long variants). Vowels are tonal timbre, i.e. they have a certain tone as a fundamental frequency. Their basic quality is determined by the position of the tongue in the oral cavity together with the shape of the lip cleft. All Czech vowels are pronounced with the participation of vocal cord vibrations, so they are voiced vowels. According to the horizontal shift of the tongue, vowels are distinguished as front (*e, i*), middle (*a*) or back (*o, u*). For the back vowels, i.e. *o* and *u*, the lips are rounded. According to the vertical position of the tongue, they are divided into high (*i, u*), middle (*e, o*) and low (*a*). [2]

## III. Timbre

In harmonic analysis, the periodic signal is decomposed into sub-harmonic components. Their frequencies are given by an integer multiple of the fundamental frequency $f_1 = 1/T_1$. That means the frequencies of the fundamental and higher harmonic components are $1f_1,\, 2f_1,\, 3f_1,\, 4f_1,\, \ldots$ [3]
The amplitudes of the partials are given by their amplitude by the module spectrum and their initial phases by the phase spectrum. In harmonic sound, the fundament is very important. Higher harmonic frequencies are layered into a natural harmonic series. The number of functional higher harmonic frequencies can be distinguished in the range from 1. to 64. This series is divided into four logical bands according to the principles of acoustics, mathematics and psychoacoustic research. [4]

- **First zone: intervals**
  Includes components from the 1st to the 8th partial. This band is perceived by man as a colour composed of individual intervals, where the first aliquot tone has the interval of the prime, the second is the interval of the octave, the third is the fifth the fourth is the fourth, the fifth is the major third, the sixth the minor third, the seventh the minor third and the eighth the major second interval. [1]
- **Second zone: clusters**
  The second band is in the range from the 9th to the 16th note. The human brain recognizes them already as chords, and mostly as clusters (clusters composed of diatonics, thus still harmonically). The intervals between these tones are around a major second.
- **Third zone: microintervals**
  The third band contains components from the 17th to the 32nd higher harmonic frequency. The higher in pitch the notes progress, the more they recede into the microintervals. A microinterval is an interval that is even smaller than a conventional semitone. It is therefore a quarter-tone as well as smaller ratios. The human auditory system perceives this layer as a continuous spectrum, resembling the timbre of clinking glass or metal.
- **Fourth zone: quasi noise**

The microintervals in this band are so small and their number so high that the brain perceives them only as noise. [4]

## IV. FFT

FFT (Fast Fourier Transform) analysis is one of many methods used in signal analysis.

### A. DFT

Mathematically described Fourier transform can be applied only to continuous signals. When converting an analog (continuous) signal into a digital signal, the signal is sampled and therefore becomes a discrete signal. To further work with such a signal, it is necessary to introduce the DFT (Discrete Fourier Transform):

$$X_k = \sum_{n=0}^{N-1} x_n e^{-jk\frac{2\pi}{N}n}, \tag{2}$$

where $X_k$ represents the DFT image, $x_n$ the sequence of processed samples and *N* the final length of the $x_n$ sequence.

### B. FFT in MATLAB

The FFT differs from the DFT algorithm in that the length of the *N* sequence $x_n$ takes the value $2^n$, where $n \in N$. This causes the computation to speed up many times due to the saving of the number of products needed for the computation. The Matlab environment offers the FFT option. [3]

## V. LPC

**LPC** (Linear Predictive Coding) is one of the most efficient methods of speech signal analysis and backward construction. The assumption is that each of the formants can only be represented by a transfer function of polar form in the form.

$$y(n) = x(n) + \sum_{i=1}^{P} a_i y(n-i), \tag{3}$$

$Y(n)$ is the discrete time response to the excitation input signal $x(n)$. The coefficients for the second term of the equation 3 are generally calculated as approximations to the original sequence $x(n)$. Thus, the prediction of the speech signal is the weighted sum of its previous values, where

$$\tilde{y}(n) = \sum_{i=1}^{P} a_i y(n-i) \tag{4}$$

is a linear predictor. *P* indicates the order of the predictor or the number of LPC coefficients and $a_i$ are the prediction coefficients. The residual signal is then designated as:

$$e(n) = y(n) - \tilde{y}(n). \tag{5}$$

In the LPC form traditionally incorporated into audio signal processing, the coefficients are chosen to minimize the quadratic deviation between the observed signals $y(n)$ and $\tilde{y}(n)$. [5] [6]

## VI. Analysis of Soprano Singing and Musical Instruments

### A. Soprano

The Fig. 1 shows the FFT spectrum of a soprano singing the vowel *e* with a pitch of G5, which corresponds to a frequency of approximately 784 Hz.

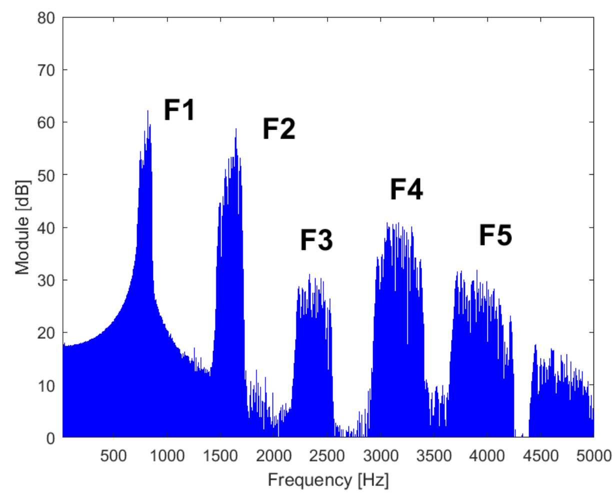



Fig. 1: FFT spectrum of soprano singing vowel *e* with pitch of G5

Each of the vowels has a different specific position of the formants F1 and F2. For example, the vowel *e* has formant F1 at 600-700 Hz and formant F2 at 1600-2400 Hz. To determine formants F1 and F2 correctly from Fig. 1, it's needed to subtract the F1 formant. F2 then corresponds to F1 (600-700 Hz) and F3 then corresponds to F2 (1600-2400 Hz).

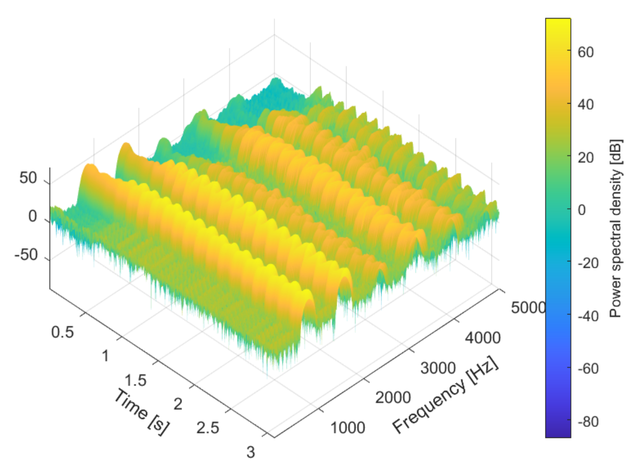



Fig. 2: Spectrogram of vowel *e* sung by soprano with vibrato

A complication in evaluating soprano results is the fact that the singer automatically resorts to vibrato when singing. It is not so easy to reach purely one frequency, but only to the band of frequencies around the desired frequency. The progression of vibrato over time is shown in Fig. 2.

### B. Flute

From the FFT spectrum in the figure, it can be seen that the flute has a full spectrum, as in the case of the human voice. From physical point of view, it is a tube open at both ends.

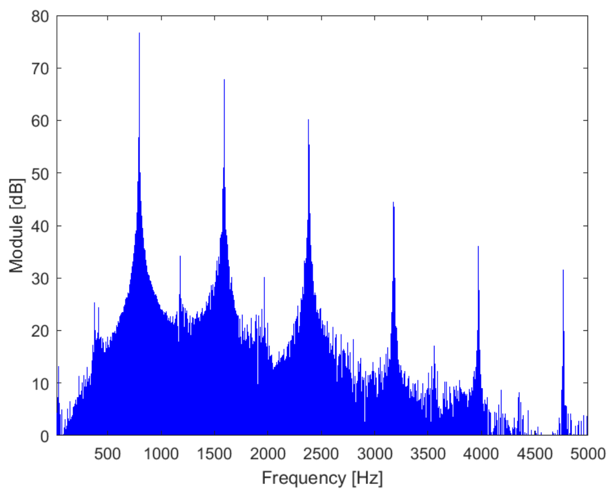



Fig. 3: FFT spectrum of flute playing tone G5

Two courses of the LPC analysis are shown in Fig. 4. The red curve shows the flute waveform and the blue waveform belongs to the soprano. The peaks are located at the same frequencies up to 5 kHz, where the waveforms start to diverge. As far as the analysis of the human voice is concerned, the region up to 5 kHz is the most important for analysis.

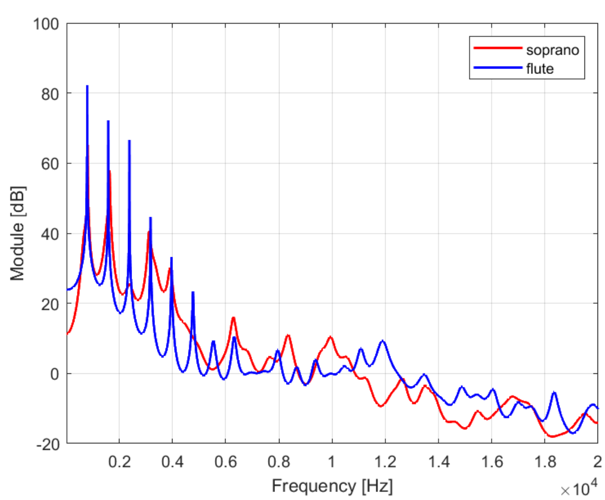
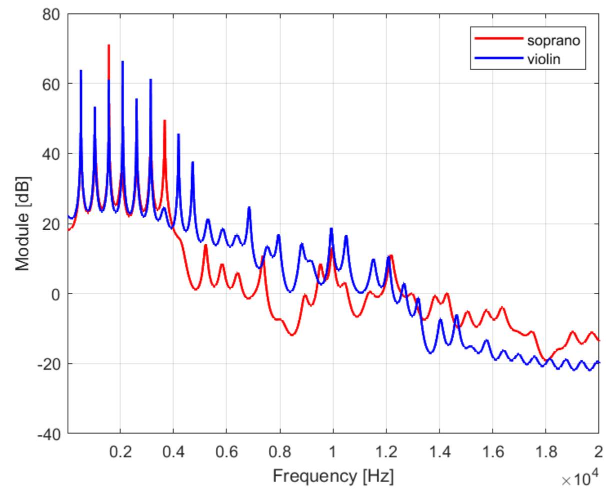


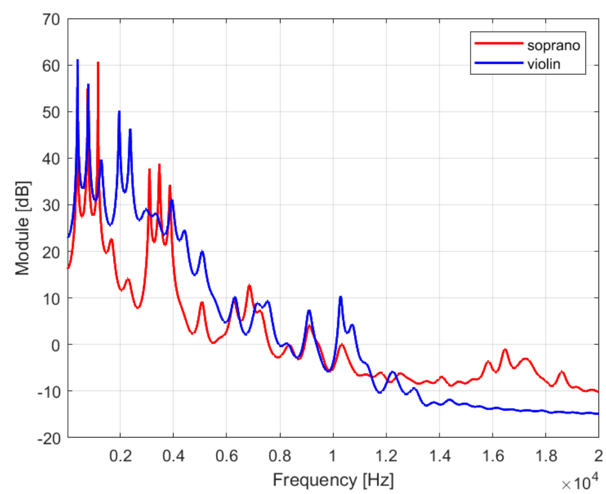

Fig. 4: LPC analysis of flute playing tone G5 and soprano singing vowel *e* with pitch of G5

### C. Violin

The violin has a full frequency spectrum, just like the flute, which is a prerequisite for the possibility of resemblance to sung vocals. Inconsistency was achieved for different heights using LPC analysis. For the note C5 the greatest similarity was found for the sung vowel *e*. The LPCs for pitches C5 and G4 are plotted in Fig. 5.




(a) LPC for tone C5




(b) LPC for tone G4

Fig. 5: LPC analysis of soprano and violin

For the pitch C5 (5a) there is a strong similarity up to around 4 kHz. This is followed by the band up to 8 kHz, where the frequency components differ only in intensity but have a similar pattern. For the pitch of the note G4 (5b), all the vowels have similar progressions and do not coincide as much with the violin's progression. This is due to the fact that the soprano's voice is no longer as strong in this register. The G4 tone is settled low in her vocal range, and thus can be expected to no longer have as full a timbre as the higher notes. Even so, it was concluded that the greatest similarity was achieved by the sung vocal *a*. The graph shows that there is a significant decrease in the representation of the soprano's frequency components in the 2 - 3 kHz frequency range. In the 3 kHz region, there is again an increase in intensity. Around this frequency region there is a vocal formant which is produced in trained opera singers.

## VII. Conclusion

Both the flute and the violin have been found to have similarities to certain sung vocals. The timbre of flute is the most similar to the soprano singing vowel *e* with pitch of G5. Another important finding was that the *i* vocal was the second most similar vocal. For the violin, it was also the *e* vocal when playing the note C5. The similarity in the case of the flute is due to the air column, which is also the cause of sound in the human voice. In the case of the violin, the soundboard plays the main role. The findings can be further developed and compared with more instruments and voice types. The dependencies between the mechanical parameters of oscillators with frequency response can also be further explored.

For the purpose of this paper, code was written for the Matlab environment in which FFT and LPC calculations and analyses of recorded audio samples are performed. LPC is one of the most widely used methods for the analysis and simulation of the human voice. It is more suitable for human voice analysis. FFT, on the other hand, is more accurate for the analysis of musical instruments where natural vibrato does not play such a role. The research can be further enriched by investigating other parameters. For example dynamics, different techniques of playing a musical instrument, different techniques of singing, psychological effects on the listener or the quality of a musical instrument are some of the possible parameters.

## References


[1] V. Syrový, Hudební akustika, 3rd, suppl. 2nd ed. Prague: Academy of Performing Arts, 2013.

[2] "Tvoření českých samohlásek", Internet language guide, 2008-2023.

[3] S. Zdeněk, Analýza signálů a soustav - BASS (scriptum), 2012 ed. VUT, 2012.

[4] O. Jirásek, "Harmonické i neharmonické barvy a šum: Jejich využití při instrumentaci", in Opus musicum, 2022, pp. 50-65.

[5] B. Gold and N. Morgan, Speech and audio signal processing: processing and perception of speech and music, Second Edition. New Jersey: John Wiley, 2011.

[6] I. I. Ibrahim and M. I. El-Adawy, "Determination of the LPC coefficients using neural networks", Proceedings of ICNN'95 - International Conference on Neural Networks, 1995.

[7] V. Mišun, Tajemství lidského hlasu. In Brno: VUTIUM, 2010.

# Spectral analysis of Streicher and Bösendorfer pianos in the works of J. Brahms and F. Liszt


1st Marek Pokorný
*Faculty of Electrical Engineering and Communication*
*Brno University of Technology*
Brno, Czech Republic
240195@vut.cz

2nd Ondřej Jirásek
*Faculty of Electrical Engineering and Communication*
*Brno University of Technology*
Brno, Czech Republic
jirasekondrej@vut.cz



***Abstract*—This paper demonstrates the use of signal analysis methods, spectral analysis in particular, to illustrate the differences in musical timbre using two historical pianos from different eras and one contemporary piano as examples. The historical pianos in question are two 19th century grand pianos similar to those used by world-famous Romantic composers such as Franz Liszt and Johannes Brahms. The spectral analysis was performed using the MATLAB computing environment software in conjunction with the Steinberg Wavelab digital audio editor, and the grand piano samples were taken from the Pianoteq instrument VST plug-in developed by Modartt, which efficiently samples the real instruments for a near-identical piano sound and experience. The methods described in this article are suitable for analysing isolated tones as well as complex harmonies; depending on the specific application, the results can be interpreted in a number of ways.**



***Index Terms*—musical timbre, signal analysis, spectral analysis, spectrum**


## I. Introduction

A musical tone can be described with four parameters:

- absolute frequency in Hz or musical pitch/notation,
- amplitude or loudness,
- duration,
- timbre.

Timbre is undoubtedly the most difficult parameter to analyse, as it depends on many factors, such as playing technique, the materials from which the instrument is made and the acoustics of the listening environment. Spectral analysis of musical instruments is useful for quantifying audible differences in timbre and for comparing these differences with other instruments in the same family. Due to the complex nature of the piano making process, this can help us in identifying an instrument based on its tonal characteristics alone. Spectral analysis plays an important role in the categorisation of musical genres. It can also help in the development of improved designs or quality control, among other applications.

The timbre of a piano is dominated by transient sounds. Not only do the partials build up and decay at different rates, but the attack sound includes some rather prominent mechanical noises (including key ”thump”) and sounds generated by longitudinal string vibrations. [1]

## II. Methods

The following methods were used for spectral analysis of the pianos:

### A. Spectrogram

The spectrogram can be defined as a visual representation of the Short Time Fourier Transform (STFT) - or a sequence of consecutive FFT windows that is used to analyze the fast changes in an audio signal. [2] In this case, the spectrogram parameters were optimized for each audio sample, resulting in an ideal resolution in the frequency range where the base frequency and its partials appear. The main parameters include the FFT window type (Hann, rectangular) and the FFT resolution (either 2048 or 4096 samples).

### B. Spectral Centroid, Flatness and Skewness

Spectral Centroid is an visualization of the weighted mean of the frequencies present in the signal, i.e. where the most amount of energy is concentrated in a particular spectrum. The centroid reflects the effect of partials on the timbre over time. [3]

Also called the tonal coefficient, spectral flatness is a ratio between the geometric and arithmetic mean of the signal, in other words, the ratio of noise to the tonal component. It expresses the amount of peaks in the signal. For tones, the flatness is close to 0%, for noise waveforms it is close to 100% (or 0 to 1 in mathematical terms). [4]

Spectral skewness characterizes the degree of asymmetry of spectral power in a signal's frequency spectrum, relative to the centroid. Completely symmetric distribution is equal to zero. Spectral skewness specifically considers the distribution of power across different frequency components, meaning a positive skew signals more energy below the spectral centroid and a negative skew indicates greater energy content above the centroid. [3]

## III. The pianos

The piano samples come from the J. B. Streicher model 1852 and the Bösendorfer model 1829, owned by Kremsegg Schloss, an Austrian museum. [5] Johannes Brahms was gifted a slightly newer model 1868, on which he composed. [6] The article also mentions a modern Steinway & Sons Model D grand piano; today these pianos are widely used for studio recordings as well as live performances by world-class pianists.

## IV. Analyzing single tones

Differences in timbre can be observed in isolated tones played on the different pianos. This method can highlight and show the behavior of partials at a single glance.

### *A. Spectrograms - G5 note*

Using the G5 (769.736 Hz) note as an example, the behavior of the individual pianos was analyzed and compared using the spectrogram.

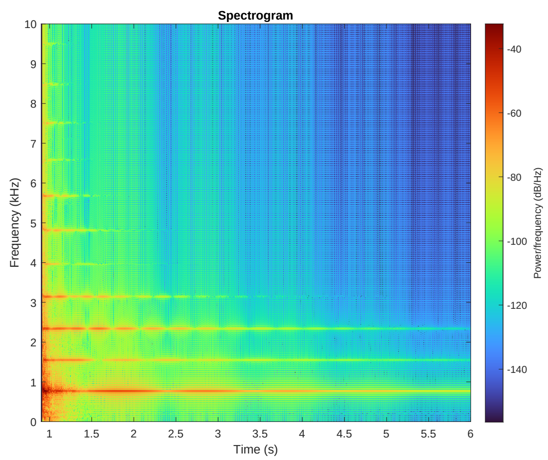



Fig. 1: J. B. Streicher piano, G5 tone

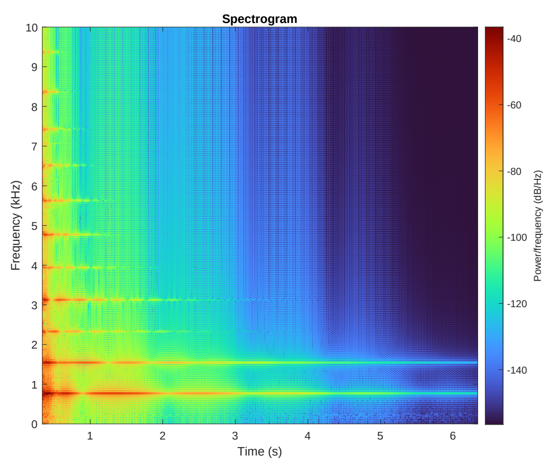



Fig. 2: Bösendorfer piano, G5 tone

Due to the behaviour of thinner piano strings, they naturally produce more partials in the higher octaves. At the same time, as we move up the octave scale, the limits of human hearing render the extreme partials (10ths and above) inaudible, and therefore do not contribute to the richness of perceived timbre. Pianos usually don't have such a wide spectrum, so the focus is on the first 8 partials, divided into odd and even. [1] The third, fifth and seventh odd partials add hollowness and dullness, nasality and roughness to the timbre. Even partials generally provide additional weight and clarity to a tone. [7] The spectrograms show 9 visible partials in the Streicher sample, while the Bösendorfer has 7. Here the Streicher's third partial (2307 Hz) has a greater amplitude than the Bösendorfer. On the contrary, the second partial of this piano is more present. In addition, each partial behaves differently in time, rising and falling in amplitude and decaying at a certain point in the long duration of the tone. The Bösendorfer is more penetrating, with a sound more akin to an upright piano, while the Streicher is closer to the qualities of some of today's pianos, such as the softer sound at higher dynamics. In general, both of these historical pianos are suitable for performance of both romantic and classical pieces. Musically, the Bösendorfer is a bit similar to the pianoforte, the predecessor of the piano.

### *B. Spectral parameters - G5 note*

Looking at the spectral parameters, we can observe the influence of partials on the spectral centroid, as the tone deviates from its fundamental frequency, especially on the Bösendorfer. There is very little peakiness in either signal, only the Bösendorfer exhibits some peaks in the second half of the tone duration between 600 and 1000 milliseconds.

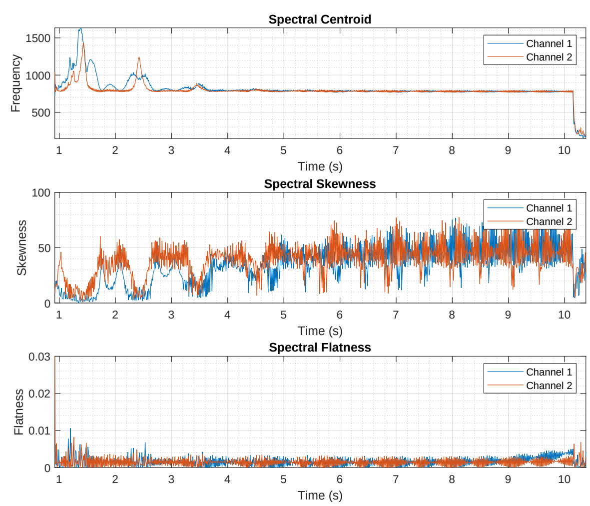



Fig. 3: J. B. Streicher piano - spectral parameters, G5 tone

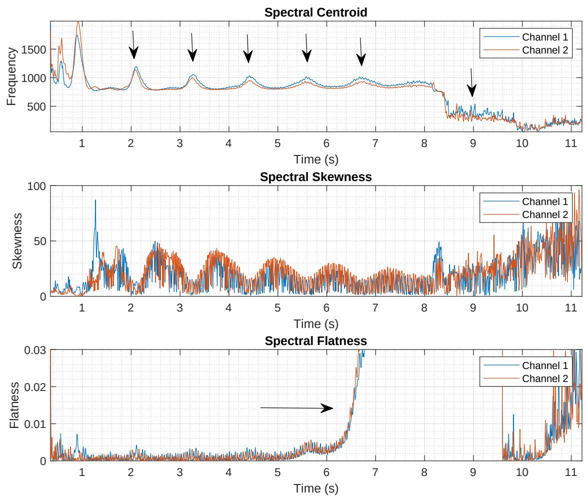



Fig. 4: Bösendorfer piano - spectral parameters, G5 tone

## V. Analyzing passages of musical compositions

Music rarely focuses on individual tones. Therefore, analysing compositions that are historically relevant to the creation of our instrument gives us a better idea of how timbre changes and/or differs. Let's choose Franz Liszt's Concerto No. 1 in E flat (1830, transcribed for solo piano in 1849), which has been converted into MIDI and played by the virtual instruments. [7]


Fig. 5: Liszt Concerto excerpt - page one


Fig. 6: Liszt Concerto excerpt - page two

### A. *Spectrograms - excerpt*

In the context of this excerpt, differences in the partials' distribution can be easily spotted. The Bösendorfer piano displays higher amount of peaks in the spectrogram, producing partials at up to 7 kHz with greater power (-50 to -70 dB), in comparison with the Streicher instrument, which sounds closer to modern piano designs such as the Steinway & Sons Model D line. For the player, this means that the Streicher sounds more nasal in the lower mid-range harmonies (largely due to the prominent third partial), but shows signs of modern piano sound in the higher register, whereas the Bösendorfer timbre has much less bass while demonstrating clarity throughout its range.

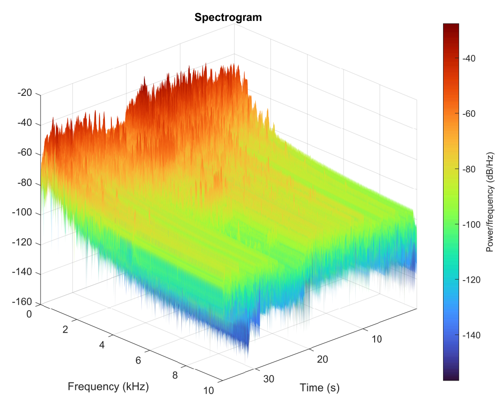



Fig. 7: J. B. Streicher piano - 3D view of spectrogram, excerpt

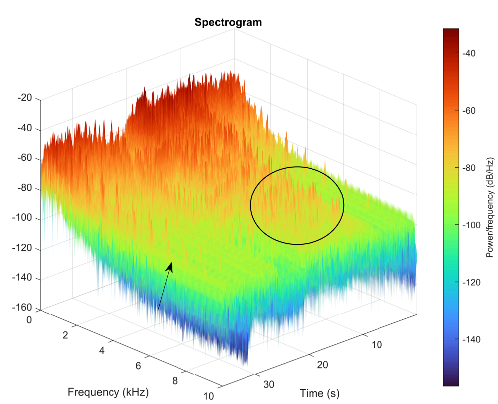



Fig. 8: Bösendorfer piano - 3D view of spectrogram, excerpt

### B. Spectral parameters - excerpt

The behaviour of both pianos is further confirmed by calculating the spectral parameters, with the Bösendorfer having more extreme peaks and troughs in the spectral centroid graph. The average skewness value is higher on the Streicher due to the lower average centroid value, but the lower frequency components are more dominant, thus forcing the skewness values into positive integers. This suggests that the timbre of each piano is more influenced by the first few partials than the higher frequency multiples. Spectral flatness is very close to zero, indicating the use of soft hammer materials and thus little to no inharmonic content. [8]

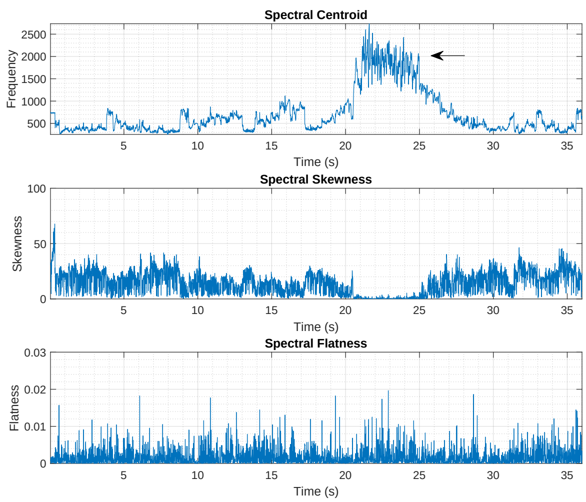



Fig. 9: J. B. Streicher piano - spectral parameters, excerpt

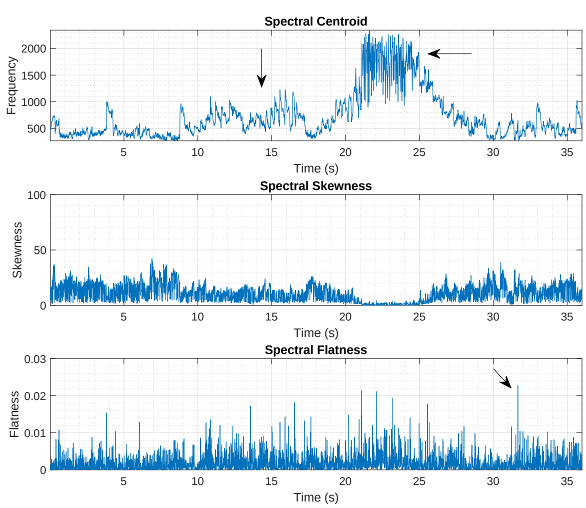



Fig. 10: Bösendorfer piano - spectral parameters, excerpt

The average parameters of each spectrum can be compared in the following table.

TABLE I: Arithmetic mean of spectral parameters, excerpt

| | Centroid [Hz] | Skewness | Flatness |
|---|---|---|---|
| **J. B. Streicher** | 698.72 | 14.46 | 0.0017 |
| **Bösendorfer** | 725.03 | 10.99 | 0.0019 |

The average values of the parameters do not differ from each other to a high degree, but can play a large role for the composer or player depending on the stylistic, historical and timbral intent of the composition.

### C. Loudness

Finally, let's look at the differences in the loudness of each piano using Steinberg's Wavelab analyser, according to the EBU loudness standard R-128. Loudness Units (LU) or Loudness Units Full Scale (LUFS) is a metric used to logarithmically measure the level of an audio signal, similar to decibels, but with additional consideration given to human hearing. A K-weighting filter is applied to the signal, simulating the roll-off at the low and high ends of the audible frequency range. LUFS are equivalent to dBFS (in digital audio), but differ when the weighting of frequencies changes, e.g. if the signal contains a higher proportion of high frequencies, a standard level value in dB may increase more than the loudness value in LU. [9] The integrated loudness represents the average volume of the entire clip, or the RMS value of the signal with the K-weighting filter added. Loudness range represents the dynamic range of the audio signal. Momentary and short-term loudness are calculated every 100 milliseconds and 1 second respectively. [10]

A Steinway & Sons Model D piano is included for comparison with a modern piano design. It is consistently the loudest, with a maximum peak of -13.0 LUFS of momentary loudness. Of the two historic pianos, the Bösendorfer achieves lower loudness values than the Streicher (around 1-2 LUFS lower), while also having the lowest recorded momentary loudness value of the three pianos at -53.5 LUFS. The J.B. Streicher has the best dynamic range of the three, at 10.8 LU, beating even the Steinway (note that this could have been caused by the sampling process - it would require verification.).

Loudness and dynamic range are important to composers and performers alike because they allow the piano to be heard in a concert hall (either solo or with an orchestra), and a wider dynamic range allows the composer to change the overall feel of a piece by introducing quiet, calm passages or strikingly loud sections, which in turn make the composition more interesting and contrasting. Different dynamics also change the timbre of the instrument significantly, as a piano hammer striking a string with greater velocity introduces more partials. Despite this, pianos generally differ in timbre mainly at higher dynamics (playing *ff* emphasizes the higher partials). [1]

TABLE II: Loudness overview, Liszt excerpt

| | J. B. Streicher | Bösendorfer | Steinway |
|---|---|---|---|
| Integrated Loudness | -24.0 LUFS | -25.1 LUFS | -20.5 LUFS |
| Loudness Range | 10.8 LU | 10.2 LU | 10.5 LU |
| Short-Term Loudness: Maximum | -19.3 LUFS | -20.9 LUFS | -15.9 LUFS |
| Short-Term Loudness: Minimum | -33.2 LUFS | -37.8 LUFS | -28.0 LUFS |
| Momentary Loudness: Maximum | -15.2 LUFS | -17.0 LUFS | -13.0 LUFS |
| Momentary Loudness: Minimum | -47.0 LUFS | -53.5 LUFS | -44.8 LUFS |

## VI. Conclusion

The results of the analysis show the trend of technological progress in the production of pianos. At the same time, each manufacturer retains its characteristic sound and timbre. It was found that the number of partials, their behaviour over time and the dynamics at different pitch ranges have the greatest influence on the overall tonal profile of each model and, as a result, will have the greatest influence on the subjective choice of instrument for composition and performance. It should be noted that the newer piano models are generally superior to the historical ones in terms of today's playing standards.

Spectral analysis of musical instruments and their timbres relies heavily on consistent measurements and an equally consistent playing style. The advantages of analyzing sampled instruments include the perfect reproducibility of the same notes with identical dynamics, due to the absence of inaccuracies introduced by a human performer (in this case a pianist), and the elimination of measurement errors due to imperfect frequency characteristics of the measuring microphones or the frequency response of our recording room. The Liszt piece can be played identically on both sampled pianos using MIDI. On the other hand, recording real instruments makes it easier to identify individual causes of variations in timbre, such as the use of different hammer materials, strings, soundboards, etc. All in all, the methods employed may give musicians, sound engineers and instrument makers a fundamental idea of the objective, measurable differences in piano sound, whether historical or contemporary.

## References


[1] N. H. Fletcher, T. D. Rossing, The physics of musical instruments, 2nd ed. New York: Springer, 1998. ISBN 978-0-387-21603-4.

[2] J. O Smith III, Mathematics of the Discrete Fourier Transform (DFT), Second Edition, online, 2007, [cit. 2024-02-08]. Available: http://ccrma.stanford.edu/ jos/mdft/.

[3] Š. Zemánková, Rozpoznávání hudebního žánru za pomoci technik Music Information Retrieval [online]. Brno, 2019 [cit. 2024-03-01]. Diplomová práce. Vysoké učení technické v Brně. Vedoucí práce Ing. Tomáš Kiska.

[4] A. Bernier-Robert, Spectral Envelope, Timbre and Orchestration Resource, 2023, Available: https://timbreandorchestration.org/writings/timbre-lingo/spectral-envelope (accessed Feb. 28, 2024).

[5] [J.B.Streicher, I. Besendorfer] In: Cp.E G. Joly, *modartt.com* [online]. [cit. 2024-02-28]. Available: www.modartt.com/kremsegg1.

[6] C. Cai, Brahms's Pianos and the Performance of His Late Works, Performance Practice Review: Vol. 2: No. 1, Article 3. 1989. [cit. 2024-03-01]. Dostupné z: https://scholarship.claremont.edu/cgi/viewcontent.cgi?article=1015 &context=ppr.

[7] Musescore, Liszt Piano Concerto No. 1 (1st Mvmt) Arr. for 2 pianos [online]. 2019 [cit. 2024-03-27]. Available: www.musescore.com/hmscomp/liszt_piano_concerto_no_1_1st

[8] O. Jirásek, Opus Musicum hudební revue [online]. 2022(01), 50-65 [cit. 2024-03-09]. Available: https://viewer.joomag.com/om-0122-web/0441553001649704937?short&

[9] Audio Mixing Mastering. What is LUFS? Everything about LUFS [online]. 2023 [cit. 2024-03-04]. Available: https://www.audiomixingmastering.com/post/what-is-lufs

[10] Steinberg, EBU Loudness Standard R-128 [online]. [cit. 2024-03-04]. Available: https://steinberg.help/wavelab_yellowtec_edition/v1/en/wavelab/topics/wavelab_concepts/ebu_loudness_recommendation_r128__c.html#ebu_loudness_recommendation_r128__c

# The Innovation of Oscillators in Rhodes pianos

Šimon Prokop
*Department of Telecommunications*
*Brno University of Technology, Technicka 12*
616 00 Brno, Czech Republic
221488@vut.cz

***Abstract*—For decades, the enchanting sounds of mechanical-electric keyboards have been a source of inspiration. However, there has been a lack of significant innovation in their mechanical components for many years. This paper deals with development of oscillators for mechanical-electric keyboard instruments, specifically focusing on innovations in a frequency spectrum due to modifications in the geometry of mechanical parts, variations in materials, sensing position and hammer strike position. Drawing from the original oscillators of Rhodes Mk7 instrument, new virtual models were created using software SolidWorks and subsequently analyzed using software Ansys. These models were then physically fabricated and subjected to examinating. The final assessment involved comparing the frequency spectra of the upgraded assemblies with the original assemblies and analyzing resulting data.**

***Index Terms*—Tines, oscillators, hammer, Rhodes, piano, spectral analysis**

## I. Introduction

In mechanical-electric instruments, there are several essential components crucial for sound generation, including the excitation system which contains hammer, oscillation system – tines or tips supported by tone bar (in Rhodes pianos) – and sensor that detects changes in the electromagnetic field engendered by oscillation of the tine [1]. First pioneers of this kind of musical instruments came from war times early in 1940's, when Harold Rhodes made his first acoustic prototype of *Xylette*, also known as *Army Air Corps Piano* which evolved in electrified version of today mechanical-electric pianos known as Rhodes pianos with the first mass-produced model *Fender Rhodes Piano Bass*. Over the years, numerous new models have been produced and sold worldwide. Worth mentioning is iconic model Mk I which had been produced from 1976 to 1979 in two variantions (Suitcase and Stage, one with and one without 80W amplifier from the Peterson FR7054 series, which later in 1977 was replaced by a newer FR7710 series amplifier with 100W stereo output). This model introduced new twisted tone bars that created a much more compact system, stable and longer tone with a supporting second harmonic. However, there was no significant mechanical change in the tines. [2] [3] [4]

This paper studies inovations in frequency spectrum compared to three keys from Rhodes piano model Mk7 (F1, F3 and F6) and simulations simulated in software Ansys dependent on changes in mechanical parts, especially geometry of the tines base and their material, position from the sensor and hammer shapes with their strike position. For technical evaluation commonly use methods in data and signal proccesing were chosen, namely Short-time Fourier Transform (STFT) and Fast Fourier Transform (FFT). From more musical perspective were used two laws, Weber-Fechner psychoacoustic law and Ohm-Helmholz psychoacoustic law (assembled from two different laws, Theory of hearing from H. von Helmholz and Ohm's acoustic law) [5] [6].

Different base geometries were considered, simulated and finally ended on three new shapes for tines that were manufactured – rectangular, triangular and square. Simulations of these shapes reported significant differences in their frequency spectrum compared to the original assemblies. Every tine was tested and manufactured in two materials configuration – stainless steel and structural steel (basic properties of used materials are listed in Table I) [7]. Manufactured assemblies were tested on a simple wooden system with an adjustable position for smooth setup of the different hammer strike positions.

The results indicate distinctions in energy mostly in lower pitch notes, keys F1 and F3, which also have the greatest potential for generating higher number of partials than higher notes. On the other hand, higher notes have ability to generate much more inharmonic sound and clusters of partials and add more noise to overall sound. The most interesting and unique assembly, whether in terms of simulation or real measurements, seems to be F1 key assembly with square base, made from structural steel, that generates stable partials in first and second band according to Ohm-Helmholz psychoacoustic law. Other assemblies will be described in the following results section.

The remainder of the paper is structured in the following manner. Section II. briefly describes the assembly and wooden system modelling procedure, their individual parts and setup for virtual assembly simulations in Ansys. Section III. explains the actual tests of the assemblies and describes in detail the testing of the hammer strike positions including their designed shapes and the different tine sensing positions. At the end, Section IV. describes the best assemblies in terms of innovativeness of the frequency spectrum followed by discussion and conclusions in Section V.

## II. Creation and simulation of assemblies

### A. Parts creation

Firstly, it was necessary to measure sizes of the individual parts in assemblies. The individual proportions of the oscillator

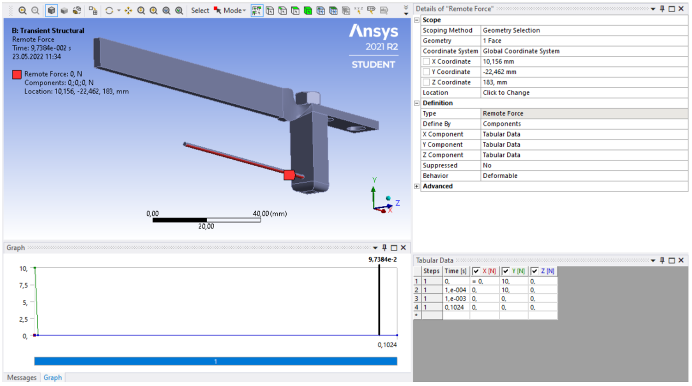



Fig. 1. F3 key oscillation system with 10 N *Remote Force* applied on Y axis.

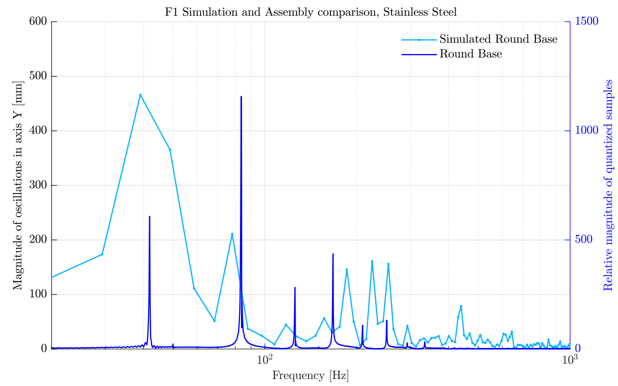



Fig. 2. Comparison of simulation and physical assembly.

systems were transformed into virtual assemblies using SolidWorks, which were then modified to create the experimental tine base shapes. Each oscillation system consists of a tine mounted (most often welded) on a block that connects tine with a support component – tone bar, through bolt. Specific shape of tone bar, usually used for lower notes, ensures that the oscillating tine will not interfere with a tone bar. This assembly creates look similar to a tuning fork and provides more stable tone. Example of virtual assembly tested with 10 N *Remote Force* in Ansys is shown in Figure 1.

All hammer parts (hammer tips, hammer, holder and key pedestal) were printed on 3D printer from PLA material. Three basic hammer tip cross-section shapes have been proposed with different curvature at the point of impact. To better accommodate the energy transfer of the tip strike, one of the designed tips was wrapped with a damping material.

Main wooden construction of the testing system consists of three basic parts, two gates and one sliding axis (shown in Figure 3). One of the gates is used for mounting the electromagnetic sensor and connecting this sensor to the TS connector. The other gate is used to mount the oscillator assembly itself. On the sliding axis there is a small wooden block for screwing the striker of the hammer tip and the key itself.

### B. *Ansys simulations*

The created virtual assemblies were then exported from SolidWorks and imported into the Ansys environment. At first, a contacts of the individual contact parts were defined (these were bolt-to-tone bar, tone bar-to-block and bolt-to-block contacts) which ensure stability of simulation. Without defining these constraints, the simulation itself would collapse and the individual parts would interfere. Subsequently, the mesh with finite number of elements for the overall assembly was generated. To calculate the frequency spectrum from the simulation, the sequence of analyses was set in the order of *Modal Analysis* and *Transient Analysis*. Prior to the actual process of calculating the transient analysis, the modal analysis that will subsequently be an input for the transient analysis needs to be set up. Thus, the 100 modes of the assembly were set to be captured over the audible spectrum range, i.e. 20 Hz to 22 kHz. A very important point of the simulation was the choice of initial and boundary conditions, which are vital for the resulting values of the simulation. By definition, boundary conditions are those that a function must satisfy at certain points. These points often lie on the edge of the domain that the function solves [8].

Main condition for the *Modal Analysis* was a solid support (referred to as *Fixed Support*) defined on the tone bar, specifically on the walls of the bolt fixing holes. For *Transient Analysis*, an external force (referred to as *Remote Force*) of various magnitudes ranging from 5 N to 100 N was applied to points at $\frac{1}{4}$, $\frac{1}{6}$, $\frac{1}{8}$, and $\frac{1}{16}$ of the length of the tine (measured from the fixed end). The forces were oriented upwards, perpendicular to the bottom part of the tine. Overall time of simulation was set to 102.4 milliseconds with step 0.1 millisecond. This setup ensures that a number of output data or samples will corresponds to a power of two and will fit into FFT window correctly. Important result data were obtained from *Directional Deformation*. The results table contains data of the average, maximum and minimum deflection of the system in millimeters relative to a certain axis of the system. For the purpose of FFT analysis, the data for the maximum deflection was stored and further processed in Excel and Matlab. A comparison of final FFT analysis from F1 key simulation and associated physical assembly is shown in Figure 2.

## III. Physical Testing

After all simulations have been performed, the real assemblies were manufactured and tested on previously introduced wooden system. The designed PLA tips were tested first on a testing assembly with rectangular base from structural steel and fixed strike position (approx. $\frac{1}{4}$ of the length of the tine, also measured from the fixed end). These tips were in three cross-section configurations – semicircular, triangular wrapped with damping material and beveled triangular.

Semicircular and beveled triangular tips generated more clusters in second band and less carrier harmonics from first band. On the other hand, tip with damping material amplified

TABLE I
BASIC PROPERTIES OF USED MATERIALS.

| Material | Properties | Value |
|---|---|---|
| Structural steel | Modulus of Elasticity in Lift $E$ | 190 GPa |
| | Specific Electrical Conductivity $\sigma$ | $0.6 \cdot 10^7$ S/m |
| Stainless steel | Modulus of Elasticity in Lift $E$ | 180 GPa |
| | Specific Electrical Conductivity $\sigma$ | $0.2 \cdot 10^7$ S/m |
| PLA | Modulus of Elasticity in Lift $E$ | 0.098 GPa |
| | Density $\rho$ | 1.25 g/cm$^3$ |

TABLE II
BANDS OF PARTIALS IN THE SPECTRUM AND THEIR PROPERTIES ACCORDING TO OHM-HELMHOLZ LAW.

| Band | Range and Properties |
|---|---|
| 1st Band | 1st to 8th Harmonic Partials |
| Properties | Fundamental color composed of intervals |
| 2nd Band | 9th to 16th Harmonic Partials |
| Properties | Cluster band composed diatonically |
| 3rd Band | 17th to 32nd Harmonic Partials |
| Properties | Continuous, quasi-harmonic spectrum |
| 4th Band | 33rd to 64th Harmonic Partials |
| Properties | Noisy, non-harmonic spectrum |

more harmonics from 1st band and brings more stability to a overall sound with less inharmonic partials and noise.

A reference tip (with triangular cross-section wrapped with damping material) was selected for subsequent testing of strike positions, with regard to its performance in generating four harmonic overtones in the first band on the tested assembly (basic characteristics of individual bands are described in Table II). This adds clarity and stability to generated tone.

Thereafter, on testing assembly with rectangular base the strike positions were examined. These positions were taken depending on length of the current tested tine in later experiments, but for assessment general behaviours of the system were performed tests for F1 key only. The strike positions were divided into three areas on the tine – from attachment (or fixed end) to $\frac{1}{3}$ length of tine, from $\frac{1}{3}$ to $\frac{2}{3}$ and from $\frac{2}{3}$ to end of the tine. Area from attachment to $\frac{1}{3}$ generates weak fundamental frequency which is newly represents by stronger subharmonic partial. This area also amplifies rather high frequencies around 1000 Hz and higher, which makes unstable and inharmonic sound. Next area, from $\frac{1}{3}$ to $\frac{2}{3}$, generates more fuller spectrum in 1st band and solves problem with sustainability of the overall sound. The third section was determined at a distance of the remaining third of the total length of the tine. This section has the characteristic of simple bending, so the tine cannot be played too vigorously (forte or fortissimo), because they might not vibrate at all and the tone would not be produced. This would cause a serious problem for players, which is why this area was discarded after first few attempts.

After selecting preferred hammer shape and the location of the strike, an effect of distance between sensor and tine were tested and additional experiment with strumming technique similar to excitation system in cembalo. In distance experiments positions were tested on and off-axis of the sensor, specifically 5 mm, 20 mm and 40 mm distances on-axis and 0 mm, 10 mm and 20 mm off-axis. The on-axis tests ended with preffered position of tine remoted 20 mm from the sensor. The condition of these tests was to find a signal transmission without unwanted effects. In position 5 mm from the sensor, signal transmits sounds with significant attenuation. The signal was the loudest due to the greater influence of the magnetic field compared to the other tests, but the problem was also caused by the large area of the magnetic field, which caused a unidirectional vibrato effect in the steady state part of the tine oscillation, in other words, caused the tine to stop oscillating faster. On the other hand, the sensed signal at a distance of 40 mm away was significantly quieter and unusable due to coil noise from sensor. The signal at a distance of 20 mm was clearer than in previous tests, with no loss of volume and no unwanted effects, so 20 mm was chosen as the ideal distance for further tests. The off-axis tests were examined also in three states – 0 mm, 10 mm and 20 mm. The distance of the tine from the sensor axis largely affected only the loudness of the overall transmission but not its timbre, so a point at 0 mm, i.e. on the axis, was chosen as the reference point.

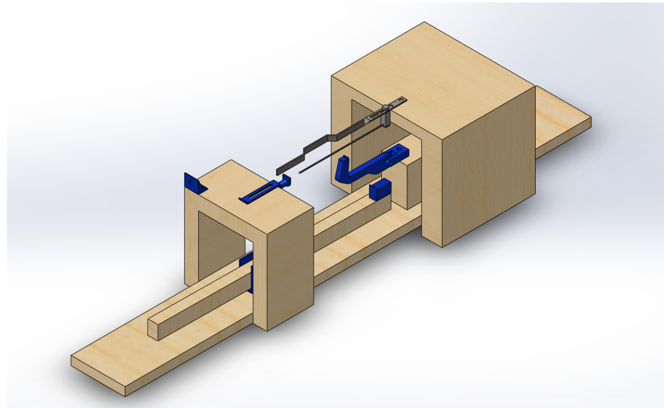


Fig. 3. Virtual model of full testing system with F1 key.

Experimental excitation tests were based on the method of playing an atypical keyboard instrument – cembalo. The exciter for this test was a classic acoustic guitar pick. The plucking was done for both tests in the center of the tine. In the first test, tip of the tine was in the same plane as the tip of the sensor magnet. In the second experimental measurement, the tip moved deeper from the sensor plane and thus affected the magnetic field from the side. The resulting signals from both measurements had a small amount of generated partials (mostly only fundamental frequency) and more audible clusters of inharmonic partials.

## IV. RESULTS

Compared to the original instrument, the redesigned tines reach mainly into the first band of partials. The most different timbre of the F1 assemblies was the one with the square base made from structural steel. This was the most prominent in both the simulated and physical versions in generating partials

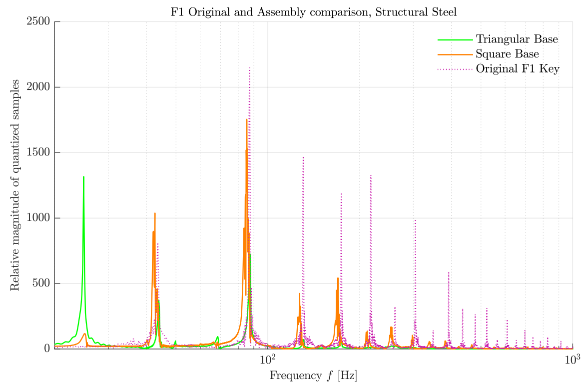



Fig. 4. Comparison of the F1 assemblies with original.

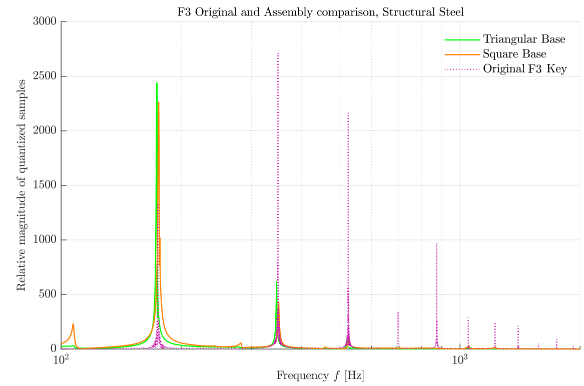



Fig. 5. Comparison of the F3 assemblies with original.

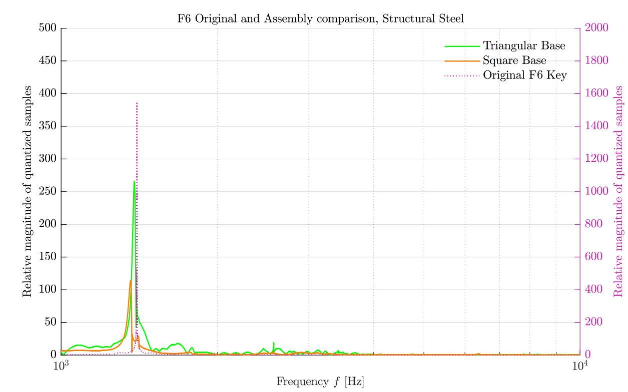



Fig. 6. Comparison of the F6 assemblies with original.

in both the 1st and 2nd harmonic bands (shown in Figure 4), and was also notable for the stable retention of these overtones over a relatively long reverberation time. The most significant F3 assembly, due to the stability and representation of the partials in the spectrum, was the structural steel assembly with the triangular base, which generated 7 of the total 8 partials from the first band. This assembly also had significant, more stable in time and energy, overtones than the previous variants (see Figure 5). The most new, in terms of timbre, setup of all the F6 assemblies tested was the one with the triangular base, which in both material variants produced interesting and unusual partials (especially inharmonic clusters that produce a metallic and tinkling sound) that are distinctly different from the original pure tone (shown in Figure 6). More detailed results and graphs are available in my bachelor thesis [9].

## V. Discussion and Conclusions

In this paper several different mechanical parts of the Rhodes piano were redesigned with a focus on spectrum innovations, simulated and measured. New designs were highly focus on base of the tine (main oscillator of the assembly) and hammer tips. These designs were modelled in software SolidWorks and then simulated in software Ansys. Simulations were performed in several different configurations of the hammer positions and different impact forces to better understanding the effect of a given force and position on the generation of partials by the assembly. Then the physical assemblies were tested and compared with simulations and original instrument. Finally, newly designed assemblies that generated a unique sound and had the most stable partials in different ratios compared to the original were introduced.

## Acknowledgment

I would like to thank MgA. Mgr. Ondřej Jirásek, Ph.D. for his expert guidance and helpful thoughts during the writing of this paper and bachelor thesis. This work has been structurally and materially supported by Nástrojárna Hrubý s.r.o., Horní Branná, the joiner Mr. Peter Lebedinsky. A great debt of gratitude also goes to Mr. Igor Hianik, who provided a metrics from his Rhodes piano collection.

## References

[1] RHODES Keyboard Instruments U.S.A., “Service Manual”, https://www.fenderrhodes.com/org/manual/toc.htm, (accessed on 4. December 2021)
[2] J. Garfield, F. Adlers, “Pre-CBS Models”, https://www.fenderrhodes.com/pianos/early.html, (accessed on 3. December 2021)
[3] J. Garfield, F. Adlers, “Rhodes Mark I (1975-1979)”, https://www.fenderrhodes.com/pianos/mark1b.html, (accessed on 5. December 2021)
[4] M. Guštar, “Elektrická piana Fender-Rhodes – historie a vývoj legendárního piana”, https://www.casopismuzikus.cz/clanky/elektricka-piana-fender-rhodes-historie-vyvoj-legendarniho-piana, (accessed on 6. December 2021)
[5] O. Jirásek, “Harmonické i neharmonické barvy a šum”, Opus musicum, Brno, pp. 51–63, January 2022, ISSN 0862-8505
[6] J. Schimmel, “Studiová technika BPC-STE, Zvuk a zvukové signály”, Brno University of Technology, Department of Telecommunications, Academic year 2020/2021, (accessed on 20. May 2022)
[7] D. Vojtěch, “Kovové materiály”, University of chemistry and technology Prague, vol. 1 Prague, 2006, ISBN 80-7080-600-1, http://147.33.74.135/knihy/uid_isbn-80-7080-600-1/pages-pdf/001.html, (accessed on 22. May 2022)
[8] J. Tippner, V. Sebera, M. Trcala, E. Troppová, “Fyzikální model (fyzikální podstata problémů, počáteční a okrajové podmínky, materiálové modely)”, Mendel University in Brno, 2021
[9] Š. Prokop, “Inovace oscilátorů u mechanicko-elektrických klávesových nástrojů”, Brno University of Technology, Department of Telecommunications, Brno, 2022, https://www.vut.cz/studenti/zav-prace/detail/141306, (accessed on 1. March 2024)

# Influence of the Reference Signals on the McRuer Models

1st Martin Malatinec
*Department of Control and Instrumentation*
*Brno University of Technology*
Brno, Czech Republic
martin.malatinec@vut.cz

2nd Ondrej Mihálik
*Department of Control and Instrumentation*
*Brno University of Technology*
Brno, Czech Republic
ondrej.mihalik1@vut.cz

***Abstract*—The article is concerned with modelling human-machine systems using McRuer's models, specifically modelling a control loop where a human acts as a controller. The main goal of this work was to investigate the influence of input signals on the response of the human operator. The article describes the functionality of the simulation scheme and the principle of measuring human responses. Finally, the measured data are processed and the observed results are evaluated.**

***Index Terms*—McRuer's models, human-machine system, input signals, human operator**

## I. INTRODUCTION

The main goal of modeling a human is to determine the "laws" that can be applied to estimating or predicting the dynamic behavior of the operator.

There are many tasks a human is capable of doing and their range is wide, this means that a person can control a large amount of systems due to the human characteristic of adapting to the situation. It follows that the mathematical description of a human as a whole would be extremely complex. When we investigating one specific task, the mathematical description of a human will be greatly simplified. [1], [2]

The human-machine system can be found in countless areas of human life. Some of the best researched types of human-machine systems are tasks dealing with driving cars [3], planes [4], ships and trains. In these systems the control objective is stabilization, steering, navigation and control of the vehicle [5]. McRuer models [2] constitute a popular modeling approach applicable in aforementioned control tasks. However, it was reported that these model depend on the bandwidth of the reference path (reference function) [6]. The effect of reference functions on the parameters of the human model is an interesting phenomenon. Therefore, our study investigates this effect in a simple feedback control task.

## II. MCRUER'S MODELS

Human as an operator doesn't need to be described as just one block in the control scheme. Instead, it's observed that a person acting as an operator is divided into multiple blocks, resembling individual PID components of the controller [2]. Therefore, for our purposes, we will simplify the model of the human operator to a linear time-invariant system.

In the referenced literature [2], the basic diagram of the control scheme for the human-machine system is presented. In our scenario, a steering wheel with its own dynamics is incorporated into the control system loop. Consequently, an update to the control scheme is required [1]. The updated control scheme is shown in the Fig. 1.

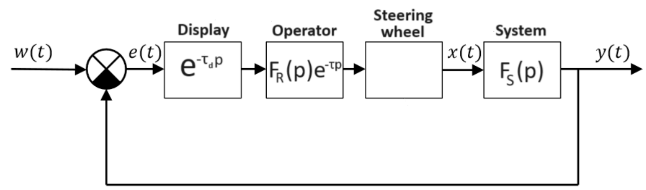



Fig. 1. The updated control scheme

In the referenced literature [2], specific systems underwent testing, leading to the experimental determination of transfer functions of human behavior.

The conclusion drawn was that as operators, humans consistently adjust their characteristics to achieve the same open loop

$$F_{\mathrm{O}}(p) = \frac{K_{\mathrm{R}}K_{\mathrm{S}}}{p}\mathrm{e}^{-\tau p}. \tag{1}$$

Here, $K_{\mathrm{R}}$ and $K_{\mathrm{S}}$ are the static gain of the controller and controlled plant, respectively. The $p$ is the Laplace operator and $\tau$ is the open-loop transport delay.

## III. SIMULATION SCHEME AND MEASURING HUMAN RESPONSES

### A. Simulation scheme

The simulation scheme shown in the Fig. 2, created using MATLAB Simulink, serves the purpose of recording and saving all significant signals. These signals include the input signal, output signal, control error, and human action.

In Fig. 2, there is an apparently missing connection in the simulation scheme between the block "joystick input" and the summation block, which calculates the control error. However, the missing feedback signals is intentional. At this place, a human operator intervenes in the simulation scheme, bridging these blocks by monitoring and adjusting the steering wheel based on the control error.

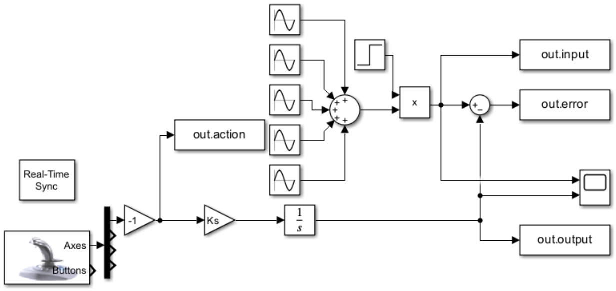



Fig. 2. A simulation scheme for measuring human responses

### B. Perception of control error by humans

The human operator responds to the control error by turning the steering wheel. The control error is perceived visually, which is the most common way humans perceive control error. This control error is displayed to the human operator using a simple radial display, which is virtually connected to the control error signal. [2]

### C. Determination of transfer function describing human operator

The mathematical models of human operator were determined using a MATLAB script with the function `tfest` [7]. This function minimizes the error of approximation between the mathematical model of the system and the measured data. The data contains the input (control error) and output (human action) time series of this system (in our case human operator).

## IV. Reference signals

We investigated the impact of changing the bandwidth of the reference signal on human operators. To assess this effect, we selected a reference signal that the human operator could control while being unable to predict. Our choice was the sum of five sine signals. We maintained a consistent signal length of 60 seconds across different bandwidths.

The reference signal is multiplied by a step function to ensure that the first second from the reference signal is always zero. This delay gives the human operator sufficient time to prepare for regulation process.

### A. Reference signal with determined frequencies

The individual frequencies of the reference signal were inspired by the bachelor thesis [1], where the frequency characteristics of the operator's transfer function were analyzed based on the operator's response to the PRBS reference signal. The control loop reported in [1] transmitted frequencies up to about 1 rad/s. The frequencies of our reference signal were chosen from the vicinity of this bandwidth. The phase shifts were calculated to ensure that the reference signal starts at zero and with zero first derivative of the signal in the first second of the signal.

The concrete reference signal with the determined frequencies reads

$$\begin{aligned} w(t) = 0.1[&\sin(0.5t + 3.0003) + \sin(t + 4.6266) + \\ &+ \sin(1.1t + 0.74896) + \sin(1.3t + 0.29247) + \\ &+ \sin(1.5t + 3.2264)]\sigma(t-1), \end{aligned} \tag{2}$$

where the $\sigma(t)$ is the unit step function and $t \in \langle 0\ \mathrm{s}, 60\ \mathrm{s}\rangle$ denotes the continuous time. The resulting reference signal is shown in figure (6).

### B. Reference signals with various width of the frequency spectrum

In order to explore various bandwidths of a reference signal (2), we had to design additional reference signals that are identical to the reference signal but with different bandwidths; they are depicted in Fig. 6.

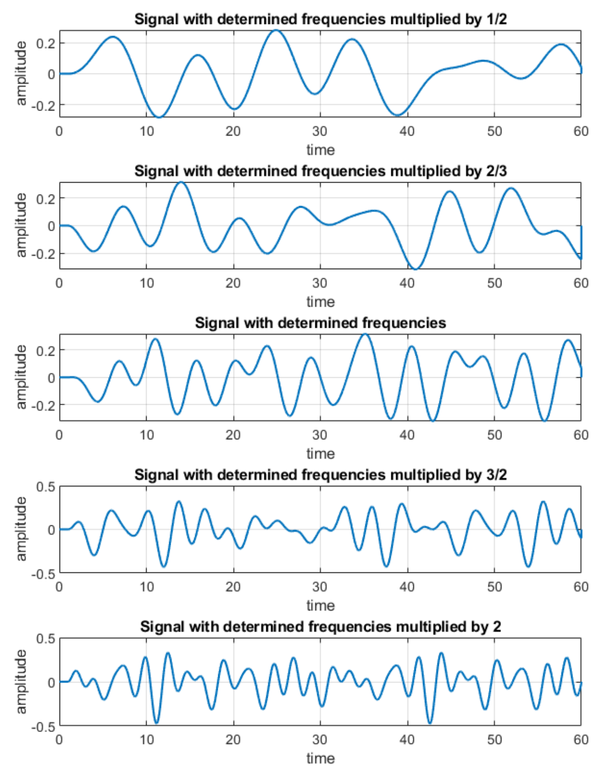



Fig. 3. Reference signals with with different bandwidths

## V. Determination of human operator transfer function

We determine the transfer functions of seven test subjects controlling the system $F_C = 1.5/p$ with the aforementioned reference signals (2).

When we tried to determine the transfer function of the human operator, we experimented with various combinations of zeros and poles of the transfer function. The requirements for the resulting transfer function were mainly that it should be as simple as possible but at the same time should approximate the measured action of the human operator as accurately as possible.

The resulting transfer function with no poles and no zeros (pure gain with delay) was able to approximate the measured action with similar quality as more complex transfer functions. We determine that transfer function of human operator controlling the system $1.5/p$ has same form for all reference signals in Fig. 3.

### A. Identification of transfer functions of one test subject

In this section, we present time series from the control loop for different bandwidths of the reference signal. Measurements presented in each plot Figs. 6–8 correspond to five different bandwidths. For each bandwidth there were 10 repeated measurements. Repeated measurement were used for identification of 10 transfer functions, whose parameters were averaged to obtain the mean parameters $\overline{K_\mathrm{R}}$ and $\overline{\tau}$.

The reference signals are derived from the default signal (2) by simply changing the frequencies using a multiplicative constant. These signals were labeled with numbers 1–5:

1) Signal with bandwidth of (2) multiplied by 1/2 yields

$$F_R(p) = \overline{K_R} \cdot e^{-\overline{\tau}p} = 1,1391 \cdot e^{-0,654p}. \qquad (3)$$

2) Signal with bandwidth of (2) multiplied by 2/3 yields

$$F_R(p) = \overline{K_R} \cdot e^{-\overline{\tau}p} = 0,8009 \cdot e^{-0,529p}. \qquad (4)$$

3) Signal (2) yields

$$F_\mathrm{R}(p) = \overline{K_\mathrm{R}}\mathrm{e}^{-\overline{\tau}p} = 0.8065\mathrm{e}^{-0.285p}. \qquad (5)$$

4) Signal with bandwidth of (2) multiplied by 3/2 yields

$$F_R(p) = \overline{K_R} \cdot e^{-\overline{\tau}p} = 0,5714 \cdot e^{-0,254p}. \qquad (6)$$

5) Signal with bandwidth of (2) multiplied by 2 yields

$$F_R(p) = \overline{K_R} \cdot e^{-\overline{\tau}p} = 0,5394 \cdot e^{-0,322p}. \qquad (7)$$

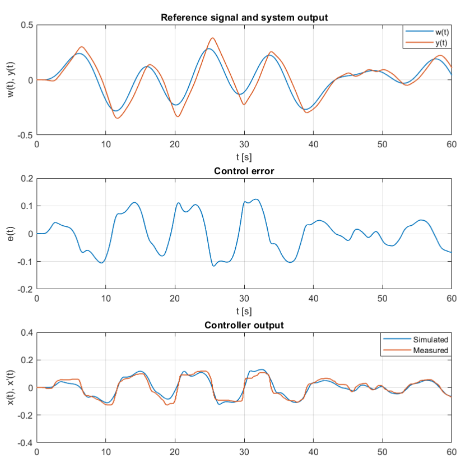



Fig. 4. Example of time series of measured and simulated signals with reference signal 1

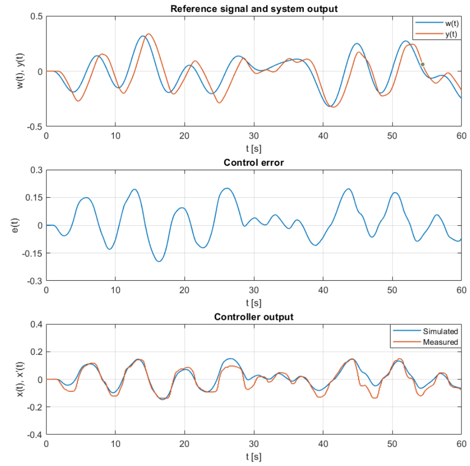



Fig. 5. Example of time series of measured and simulated signals with reference signal 2

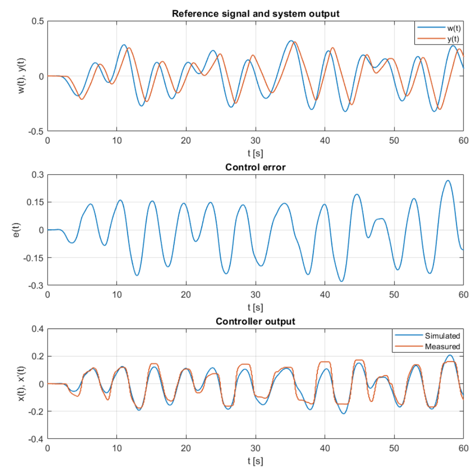



Fig. 6. Example of time series of measured and simulated signals with reference signal 3

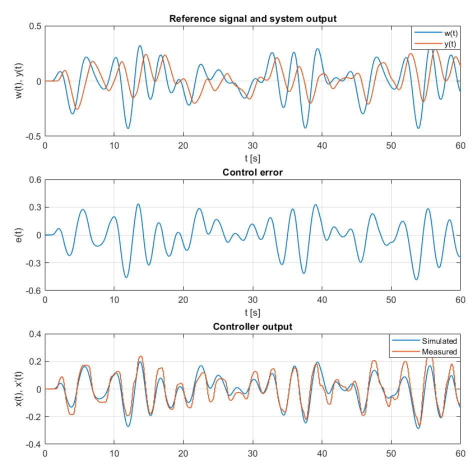



Fig. 7. Example of time series of measured and simulated signals with reference signal 4

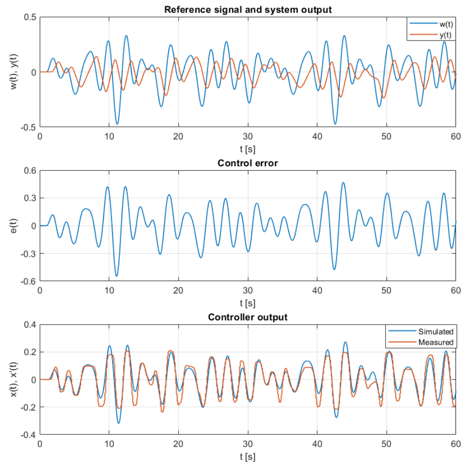



Fig. 8. Example of time series of measured and simulated signals with reference signal 5

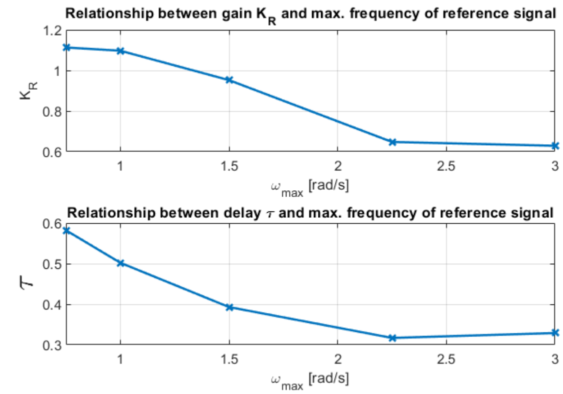



Fig. 9. Relationship between the average values of gain and delayand the maximum frequencies present in the reference signals

### B. Determined transfer functions of all test subjects

The transfer functions of the human operator controling system $1.5/p$, exhibits a consistent form (pure gain with delay) across all subjects for different reference signals. The specific values of gains and delays for all subjects are presented in Table V-B.

TABLE I
SPECIFIC VALUES OF GAINS AND DELAYS FOR ALL SUBJECTS

| | | Reference signals | | | | |
|---|---|---|---|---|---|---|
| | | Signal1 | Signal2 | Signal3 | signal4 | Signal5 |
| Subject 1 | $K_R$ | 1.1391 | 0.8009 | 0.8065 | 0.5714 | 0.5394 |
| | $\tau$ | 0.654 | 0.529 | 0.285 | 0.254 | 0.322 |
| Subject 2 | $K_R$ | 0.6644 | 0.5743 | 0.5035 | 0.4375 | 0.5584 |
| | $\tau$ | 0.266 | 0.106 | 0.112 | 0.229 | 0.395 |
| Subject 3 | $K_R$ | 0.9945 | 0.6733 | 0.5670 | 0.2904 | 0.2517 |
| | $\tau$ | 0.750 | 0.610 | 0.624 | 0.439 | 0.545 |
| Subject 4 | $K_R$ | 1.3837 | 1.6795 | 1.2803 | 0.9013 | 0.6855 |
| | $\tau$ | 0.651 | 0.641 | 0.494 | 0.375 | 0.320 |
| Subject 5 | $K_R$ | 0.6719 | 1.0838 | 1.0938 | 0.7885 | 0.6284 |
| | $\tau$ | 0.386 | 0.423 | 0.418 | 0.318 | 0.232 |
| Subject 6 | $K_R$ | 1.6925 | 1.6082 | 1.4285 | 0.9736 | 1.1561 |
| | $\tau$ | 0.594 | 0.520 | 0.405 | 0.279 | 0.254 |
| Subject 7 | $K_R$ | 1.2442 | 1.2581 | 0.9861 | 0.5718 | 0.5838 |
| | $\tau$ | 0.773 | 0.687 | 0.410 | 0.328 | 0.232 |
| Average values | $K_R$ | 1.1129 | 1.0970 | 0.9522 | 0.6478 | 0.6290 |
| | $\tau$ | 0.582 | 0.502 | 0.393 | 0.317 | 0.329 |

In the graphical representation in Fig. 9 is shown the relationship between the average values of gain, $K_{\mathrm{R}}$, and delay, $\tau$, in the transfer functions of the human operators and the maximum frequencies present in the reference signals.

## VI. CONCLUSION

It can be concluded that the identified transfer functions of human operators, controling periodic signals with various bandwidths, align with the McRuer models described in literature [2].

As was mentioned in subsection V-B the form of transfer function of the human operator controlling a simple integrator $1.5/p$ remains consistent across all test subjects and for all bandwidths of the reference signals.

Another shared characteristic among all subjects is the decrease in both gain and delay of the operator as the frequency of the reference signal increases. This phenomenon can be caused by the fact that with higher frequencies, the reference signal appears to change more rapidly, necessitating quicker responses from the operator, that may results in decrease of delay. However, even with faster steering wheel adjustments, the operator may still struggle to reach the peak of the wave before the signal changes direction, that may results in decrease of gain.

In the further research the changing transfer function may be adressed using non-linear models, such as neural controllers or fuzzy systems.

## REFERENCES


[1] J. Čelko, "Modelování systémů člověk-stroj pomocí McRuerových modelů." B.S. thesis, Dept. Control and Instrum., Brno Univ. Technol., Brno, Czech Republic, 2022.

[2] D. T. McRuer and E. S. Krendel, "Elementary concepts for the quantitative description of human operator dynamics in control systems" in *Mathematical Models of Human Pilot Behavior*, 1st ed. London, United Kingdom: AGARD, Jan. 1974, ch. 2, pp. 5–15.

[3] S. I. N. Tientcheu, S. Du, and K. Djouani, "Review on Haptic Assistive Driving Systems Based on Drivers' Steering-Wheel Operating Behaviour", *Electronics*, vol. 11, no. 13, Art. no. 2102, 2022. DOI: 10.3390/electronics11132102

[4] S. Xu, W. Tan, A. V. Efremov, L. Sun, and X. Qu, "Review of control models for human pilot behavior," *Annu. Rev. Control*, vol. 44, pp. 274–291, Oct. 2017. DOI: 10.1016/j.arcontrol.2017.09.009.

[5] M. Havlíková, "Diagnostika systémů s lidským operátorem." Ph.D. dissertation, Dept. Control and Instrum., Brno Univ. Technol., Brno, Czech Republic, 2009.

[6] K. van der El, D. M. Pool, M. M. van Paassen, and M. Mulder, "Effects of Target Trajectory Bandwidth on Manual Control Behavior in Pursuit and Preview Tracking," *IEEE trans. human-mach. syst.*, vol. 50, no. 1, pp. 68–78, 2020. DOI: 10.1109/THMS.2019.2947577.

[7] L. Ljung, "System Identification Toolbox: Reference." Matlab version R2023b, Natick, MA, USA. [Online] Available: https://www.mathworks.com/help/pdf_doc/ident/ident_ref.pdf Accessed on: Mar. 1, 2024.

# Design of 90 degree phase shift hybrid coupler

1st Adam Ondrejka
*Department of Theoretical and Experimental Electrical Engineering*
*Brno University of Technology*
Brno, Czech Republic
xondre13@vut.cz

2nd Tomáš Kříž
*Department of Theoretical and Experimental Electrical Engineering*
*Brno University of Technology*
Brno, Czech Republic
krizt@vut.cz

***Abstract*—This paper describes the design procedure for quadrature hybrid for UHF band. Circuits are based on distributed parameters with anticipated manufacturing on a PCB. Emphasis is put on a compact size, three methods are used for size reduction of transmission lines. All designs are simulated in electromagnetic simulation software. The best design using meandered lines is verified by measurement with no detectable deviation from target working frequency. Temperature simulation is conducted with the best design.**

***Index Terms*—quadrature hybrid, branchline, meander, microstrip, numerical model, simulation, PCB**

## I. Introduction

The goal of this paper is to present the design of a circuit which can handle the splitting of 100 W of input power in a ratio of 1:1 while operating at UHF, specifically at 433 MHz. Circuit also needs to shift its outputs 90 degrees out of phase. This setup is needed for quadrature driving in an experimental laboratory apparatus.

Quadrature hybrid is therefore needed, a 4 port circuit providing the above-mentioned properties while also maintaining impedance matching for all its ports with one port being isolated [1].

There are many ways to construct such a hybrid. In this case there was a push for the device to be cheap and simple in its construction, another requirement was the ability to easily adjust the central frequency of the circuit by simple design changes. Due to relatively high input power, losses were another concern and finally also the physical dimensions of the circuit.

## II. Topology and substrate selection

The previously mentioned criteria allow using PCB as a manufacturing process. Therefore the first step was to choose transmission line structure. Microstrip was chosen because it needs only a single layer of substrate and simple design equations exist, which allow quick and quite accurate calculation of dimensions for the desired microstrip [2].

Another crucial step was the choice of substrate, this was limited by the manufacturer of PCB to FR4, RO4350B and ZYF300CA-C. Because of the relatively high input power, a substrate with the lowest resulting losses was chosen. Based on volume losses from simulations, ZYF300CA-C turned out to be the best choice. The relative permittivity of this material is 2.94 and the substrate's thickness is 1.52 mm. Both copper layers were chosen to be 35 $\mu$m thick.

As for the choice of circuit topology, there are two options, which can be easily produced with microstrip, branchline and rat-race hybrid [2]. In addition, the operation of these two circuits is based solely on how long their transmission line segments are, which makes them very easy to design. Since rat-race hybrid does not have a 90-degree phase shift between its outputs, it needs an additional quarter wavelength line segment on one of the output ports [3].

After performing simulations of both branchline and rat-race hybrids, it was clear that branchline is better, because its phase shift was more stable around the central frequency, it also has a smaller footprint, because it uses only 4 quarter wavelength line segments, as displayed in Fig. 1.

50 Ω external characteristic impedance was considered in the design, therefore individual branchline segments have an impedance of either 35.36 Ω or 50 Ω.

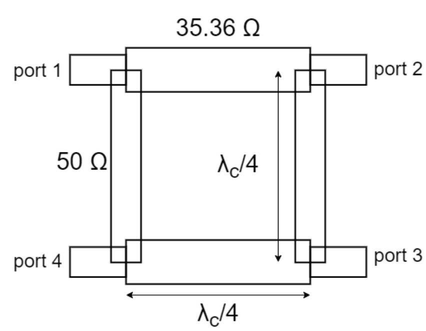



Fig. 1. Basic branchline hybrid layout.

## III. Size reduction of branchline hybrid

Despite using the central frequency of 433 MHz, the wavelength is still rather large. Therefore the branchline hybrid has to be modified to make it more compact. The circuit would also benefit from reduction of wasted space in the original square branchline hybrid. There are two ways to achieve this.

### *A. Line transformation with added capacitance*

The transmission line segment shown in Fig. 2(a). can be made physically shorter while maintaining its electrical length by increasing its characteristic impedance and adding compensating susceptance $B_{o1}$.

Calculation of value for the added susceptance for the Π-cell from Fig. 2(c). can be done using the following equations [6]:

$$B_{o1} = \frac{\cos\theta_S - \cos\theta}{Z\sin\theta}, \tag{1}$$

$$Z_S = \frac{Z\sin\theta}{\sin\theta_S}. \tag{2}$$

Fig. 2(b). shows the T-cell configuration described by [7]:

$$B_{o1} = \frac{2}{Z_S\tan(2\theta_S)}, \tag{3}$$

$$Z_S = \frac{Z}{\tan\theta_S}, \tag{4}$$

where $Z$, $\theta$ are characteristic impedance and electrical length of the original line segment, $Z_S$, $\theta_S$ are characteristic impedance and electrical length of the resulting line segment.

Susceptance $B_{o1}$ can be created either in the form discrete capacitor [8] or as capacitive stub [6], [7].

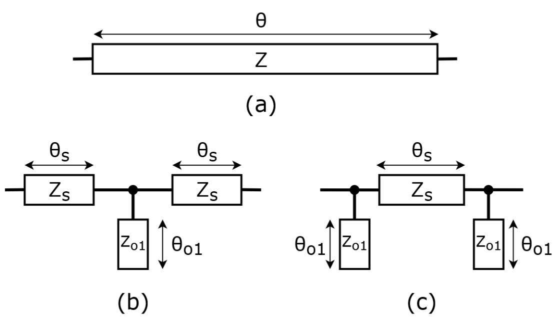



Fig. 2. Line segment and its shortened equivalents.

After simulating designs which used both Π and T configurations, capacitors and capacitive stubs, it became clear that while providing a substantial reduction in line length, it also brought several disadvantages.

Designs utilizing stubs increased circuit size in other directions due to the stub length. Designs with discrete capacitors would require obtaining components with the specific value of capacitance, which is difficult. Further, all designs increased the characteristic impedance of lines, thus narrowing the microstrip and increasing losses. Problems also arose when it came to connecting the stubs or capacitors to the microstrip, because such junctions always resulted in poor performance.

### B. Meandering of lines

Meandering of lines simply introduces bends into straight line segment, which allows better usage of surrounding surface while decreasing the overall length of the circuit [4]–[6].

For the reduction of the 35.36 Ω line it was chosen that its wide microstrip will be meandered using three semicircles, as shown in Fig. 3. Length of the line is defined by a curve running through the center of the microstrip.

The 50 Ω lines were meandered inward to reduce the unused area in the middle of the circuit. Fig. 3. shows how each line is made out of five straight segments connected by bends. It is important to note, that the portions of the line cannot be too close, otherwise parasitic coupling across the gap will occur.

Introducing so many curves into microstrips has the effect of decreasing their electrical length. After the first simulation of the design, the central frequency was shifted upwards. This can be easily corrected by first calculating length of microstrip at this new central frequency using design equations for microstrip. Lines in the model are then elongated by the difference between the original line length and this new one calculated from the first simulation result. Care must be taken to calculate for each line impedance separately because equations describing wavelength are influenced by this.

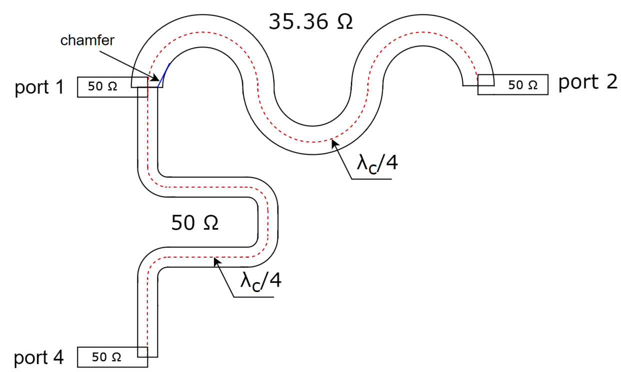



Fig. 3. Meandered branchline hybrid layout.

It became apparent that this meandered design is superior because it is very simple and does not introduce any drastic changes along the microstrip length, resulting in good performance.

## IV. Final assembly

A 3D model with meandered lines was modified by adding edge-mounted SMA connectors and solder mask. There was no noticeable negative influence on the simulated results.

The finished device would be placed inside a box from extruded aluminium acting as an EMI shielding. A rough calculation was performed showing that the rectangular waveguide cavity resonator [2] formed by the box has its lowest resonant frequency slightly above 2 GHz, which is well beyond the working frequency of the hybrid.

A 3D model of this box was created, as seen in Fig. 4. and it was included in the simulation without observing any significant degradation of performance.

Since the box is fully enclosed, there was a concern about overheating of the PCB. Thermal analysis was conducted in Ansys Icepak. The whole assembly was simulated at an ambient temperature of 20 °C while being placed on a flat surface. Surrounding air was allowed to naturally flow over the box due to convection.

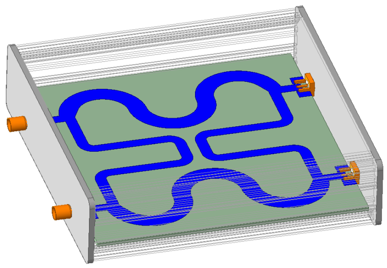


Fig. 4. Model of the completed circuit assembly.

Results in Fig. 5. show that the hottest parts of the whole setup are on the 35.36 Ω line section connecting the input port and its mirrored output port, the maximum temperature of the substrate and copper does not exceed 41 °C. The outer surface of the aluminium box rests at about 25 °C.

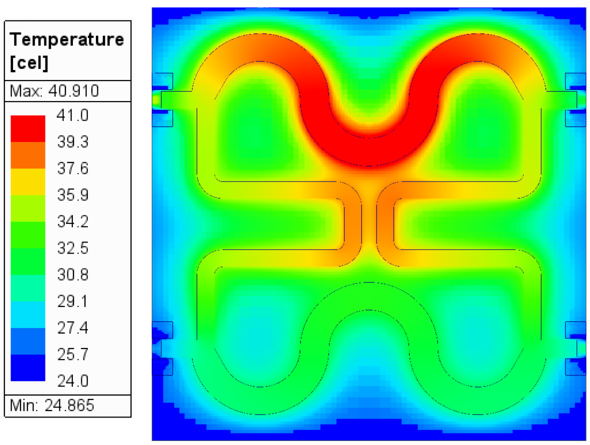



Fig. 5. Temperature distribution on top of the substrate and microstrip.

## V. Measurement

The manufactured PCB of the meandered branchline hybrid was fitted with SMA connectors, as shown in Fig. 6. This board was then measured using Rohde and Schwarz ZVL6 two-port network analyzer.

S-parameters were measured in 201 datapoints from 333 MHz to 533 MHz with respect to the input port, which stayed connected to the network analyzer's first channel, while the second channel was rotated between remaining ports on the board. Ports connected to the analyzer were matched with its internal 50 Ω impedance and the remaining two ports were matched by 50 Ω terminations for SMA connectors.

## VI. Comparison of Results

In Table I and II are displayed the compiled results for simulations of rat-race quadrature hybrid, simple branchline hybrid and its meandered modification. In addition, there are results for the meandered hybrid with connectors (both simulated and measured) and finally simulation of the complete assembly of meandered hybrid, connectors and aluminium box.

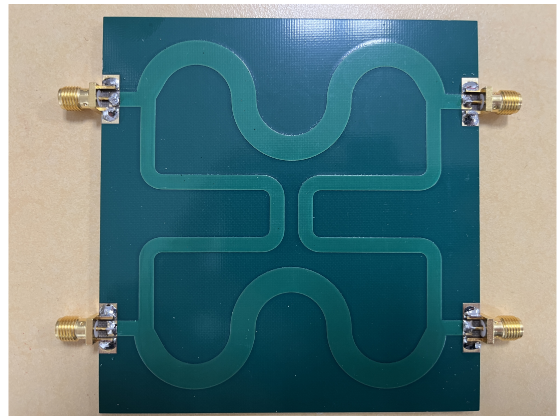


Fig. 6. Manufactured board of the meandered quadrature hybrid.

Table I displays the magnitudes of S-parameters and phase difference between output ports 2 and 3, values are recorded at central frequency $f_c$ (433 MHz) and at boundaries of the interval ± 10 MHz around $f_c$, therefore $f_l$ = 423 MHz and $f_u$ = 443 MHz, because only narrow bandwidth is required.

Table II shows simulated losses throughout the volume of all solids with input power being 100 W. Area covered by each design is also recorded.

It is clear that rat-race-based hybrid is too large and has an unacceptable drift of phase difference. Branchline structure solves these issues, it provides almost constant phase difference inside the considered frequency interval while being much smaller. The division of input power is almost in a 1:1 ratio with negligible imbalance.

The meandered modification enables further miniaturization of the circuit, it occupies less than half the area of the simple branchline hybrid. The S-parameters are mostly unaffected by this modification, only $S_{11}$ and $S_{41}$ specifically degrade only very slightly for this to be of any concern. The addition of connectors and of the box caused a slight increase in losses for the meandered hybrid but at the same time there were minor improvements of $S_{11}$ and $S_{41}$.

Measurement of the meandered board fitted with connectors very well matches the simulated (designed) central frequency, there was however improvement in $S_{11}$ magnitude by over 6 dB and of $S_{41}$ by about 2 dB at $f_c$. More importantly there was a slight increase in losses, most likely due to the connectors and their solder joints, nickel content in the gold plated soldering pads might play a role as well. Another reason could be surface roughness of the microstrip or losses from soldermask. All of these effects were not included in the simulations. Phase difference between outputs changed by about one degree for the worse at margins of the considered 20 MHz interval, the central frequency however maintained excellent phase difference.

## VII. Conclusion

With the help of numerical modeling, several designs of narrowband quadrature hybrid were proposed and tested.

TABLE I
SIMULATED S-PARAMETERS AND PHASE DIFFERENCE ($\phi(S_{21} - \phi(S_{31}))$ OF PROPOSED HYBRIDS

| | $\|S_{11}\|$ [dB] | | | $\|S_{21}\|$ [dB] | | | $\|S_{31}\|$ [dB] | | | $\|S_{41}\|$ [dB] | | | $\phi(S_{21} - \phi(S_{31})$ [°] | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Design** | $f_l$ | $f_c$ | $f_u$ | $f_l$ | $f_c$ | $f_u$ | $f_l$ | $f_c$ | $f_u$ | $f_l$ | $f_c$ | $f_u$ | $f_l$ | $f_c$ | $f_u$ |
| Rat-race with $\lambda_c/4$ line | -36.2 | -51.1 | -37.2 | -3.13 | -3.11 | -3.12 | -3.05 | -3.07 | -3.06 | -34.2 | -38.2 | -34.6 | -85.6 | -89.2 | -92.7 |
| Simple branchline | -26.6 | -45.3 | -28.8 | -3.19 | -3.14 | -3.14 | -3.07 | -3.08 | -3.10 | -25.1 | -29.8 | -26.1 | -270.4 | -270.2 | -270.1 |
| Meandered branchline | -26.4 | -31.5 | -25.0 | -3.10 | -3.06 | -3.09 | -3.09 | -3.10 | -3.11 | -26.6 | -34.3 | -25.9 | -269.7 | -269.9 | -270.1 |
| Meandered with connectors | -26.2 | -33.6 | -25.9 | -3.18 | -3.14 | -3.17 | -3.08 | -3.09 | -3.10 | -26.7 | -38.8 | -26.9 | -269.6 | -269.7 | -269.8 |
| Meandered with connectors measured | -26.9 | -40.2 | -26.6 | -3.23 | -3.16 | -3.22 | -3.15 | -3.13 | -3.17 | -26.6 | -40.7 | -27.4 | -268.7 | -269.8 | -271.1 |
| Meandered assembly | -25.6 | -33.3 | -26.7 | -3.11 | -3.07 | -3.10 | -3.09 | -3.09 | -3.10 | -26.0 | -37.2 | -27.5 | -269.9 | -270 | -270.1 |

TABLE II
SIMULATED LOSSES AT $f_c$ AND APPROXIMATE AREA TAKEN UP BY EACH DESIGN.

| **Design** | **Simulated losses** [W] | **Used area [mm$^2$]** |
|---|---|---|
| Rat-race with $\lambda_c/4$ line | 1.37 | 349 × 274 |
| Simple branchline | 1.35 | 150 × 133 |
| Meandered branchline | 1.48 | 89 × 98 |
| Meandered assembly | 1.59 | 115 × 105 |

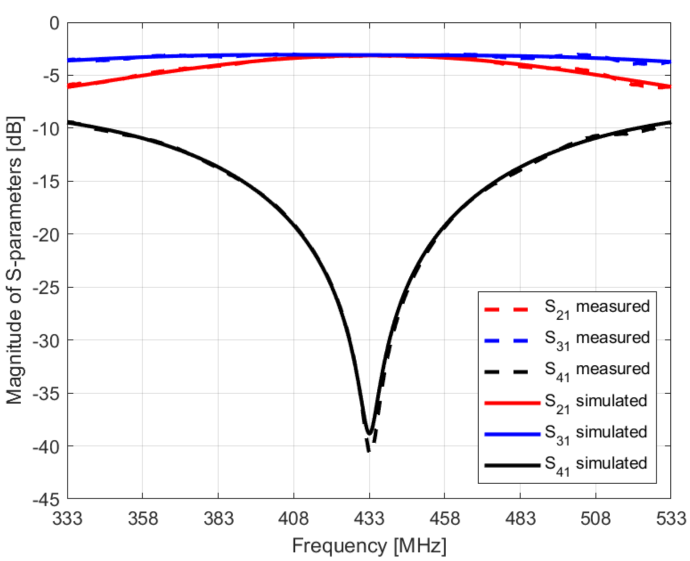



Fig. 7. Simulated and measured S-parameters of the meandered hybrid with connectors.

Branchline topology on microstrip was quickly selected for its simplicity and compactness. Substrate material with the lowest available loss tangent allowed very low wasted power in the circuit. Further attempts at size reduction have been made and by far the best proved to be the topology using meandered lines. The performance of the meandered circuit was also verified by measurement.

Thermal simulation showed that it is possible to run the circuit with the desired power input without risking overheating.

If it is determined in the future that different central frequency is needed, the whole design can be quickly changed to accommodate this need, because the design of meandered microstrip lines requires only simple calculations without the need for time-consuming optimization.

## ACKNOWLEDGMENT


Presented work was supported by the development project FEKT-S-23-8425.


## REFERENCES


[1] S. A. Maas, Practical microwave circuits. Norwood: Artech House, 2014.
[2] D. M. Pozar, Microwave engineering, 4th edition. John Wiley, 2012.
[3] J. F. White, High frequency techniques: an introduction to RF and microwave engineering. Hoboken: John Wiley, 2004.
[4] N. Orhanovic, R. Raghuram and N. Matsui, "Characterization of microstrip meanders in PCB interconnects," 2000 Proceedings. 50th Electronic Components and Technology Conference (Cat. No.00CH37070), Las Vegas, NV, USA, 2000, pp. 508-512, doi: 10.1109/ECTC.2000.853205.
[5] L. G. Maloratsky, "Using Modified Microstrip Lines to Improve Circuit Performance", High Frequency Electronics, no. 3, pp. 36-52, 2011.
[6] Young-Hoon Chun and Jia-Sheng Hong, "Compact wide-band branch-line hybrids," in IEEE Transactions on Microwave Theory and Techniques, vol. 54, no. 2, pp. 704-709, Feb. 2006, doi: 10.1109/TMTT.2005.862657.
[7] C. -W. Tang and M. -G. Chen, "Synthesizing Microstrip Branch-Line Couplers With Predetermined Compact Size and Bandwidth," in IEEE Transactions on Microwave Theory and Techniques, vol. 55, no. 9, pp. 1926-1934, Sept. 2007, doi: 10.1109/TMTT.2007.904331.
[8] T. Hirota, A. Minakawa and M. Muraguchi, "Reduced-size branch-line and rat-race hybrids for uniplanar MMIC's," in IEEE Transactions on Microwave Theory and Techniques, vol. 38, no. 3, pp. 270-275, March 1990, doi: 10.1109/22.45344.

# A Comparison of Popular Band-limited Signal Reconstruction Methods

1st Matyáš Telecký
*Department of Control and Instrumentation*
*Brno university of technology*
Brno, Czech Republic
Matyas.Telecky@vut.cz

2nd Ondrej Mihálik
*Department of Control and Instrumentation*
*Brno university of technology*
Brno, Czech Republic
Ondrej.Mihalik1@vut.cz

***Abstract*—The paper focuses on the reconstruction of continuous and discrete signals observed during a finite time interval. For the task of signal recovery, the Gerchberg-Papoulis (GP) method is compared with a proposal for an improved GP implementation and with Slepian reconstruction method. These three methods are implemented in the MATLAB and applied to selected band-limited signals. Their results are compared in terms of computational complexity and reconstruction accuracy. The improved GP method and Slepian method are recommendable, as they remove problems associated with spurious signal periodicity that is usually induced by the FFT in the original GP method.**

***Index Terms*—band-limited signal, Gerchberg-Papoulis Method, MATLAB, signal approximation, signal recovery**

## I. Introduction

Certain practical signal processing applications cope with the problems that a signal can be observed over a finite time interval. Furthermore, AD converters record only discrete signals. By discretising a finite signal segment, we lose some information and the original continuous waveform cannot be reconstructed using the well-know Nyquist-Shannon-Kotelnikov sampling theorem. Methods such as the Gerchberg-Papoulis (GP) algorithm [1] or Slepian extrapolation method [2] are commonly used to recover lost information in discrete signals [3]. Our aim is to propose a modification of the GP algorithm which addresses the issues induced by Fast Fourier Transform (FFT).

The paper is divided as follows. Section II described the GP algorithm in general terms. Its standard implementation based on FFT is discussed in Section III. This version introduces spurious periodicity in the extrapolated signal. Therefore, in Section IV, we propose the exact implementation of the original method without FFT. Strange as it may seem, the continuous GP algorithm is naturally reduced into a finite-dimensional problem that can be realised using a computer. In Section VI, both versions of the GP algorithm are compared to the Slepian method. Successful application of this method was recently reported in [2].

## II. Standard implementation of GP algorithm

In this section we will illustrate one iteration of the GP method described in [1]. To simplify the discussion, the method is typically presented as if we were working with continuous signals. Later, in Section IV, it will be applied to discrete signal samples.

Brno University of Technology

For better explanation, we may define a concrete signal $f(t) = \mathrm{sinc}(t)$. The extrapolation starts with a continuous-time signal; it is assumed that only the part $g_0(t)$ of the original signal, $f(t)$, is observed, see Fig. 1, a). The first step is calculation of the frequency spectrum $G_0(\omega)$ using FFT, Fig. 1 b). The second step is to limit the frequency spectrum. We obtain a band-limited signal $H_0(\omega)$ using

$$H_n(\omega) = \begin{cases} G_n(\omega) & \text{for} \quad |\omega| < \Omega, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

see Fig. 1, c). The next step is the application of the IFFT to obtain the first signal approximation $h_0(t)$, Fig. 1, d).

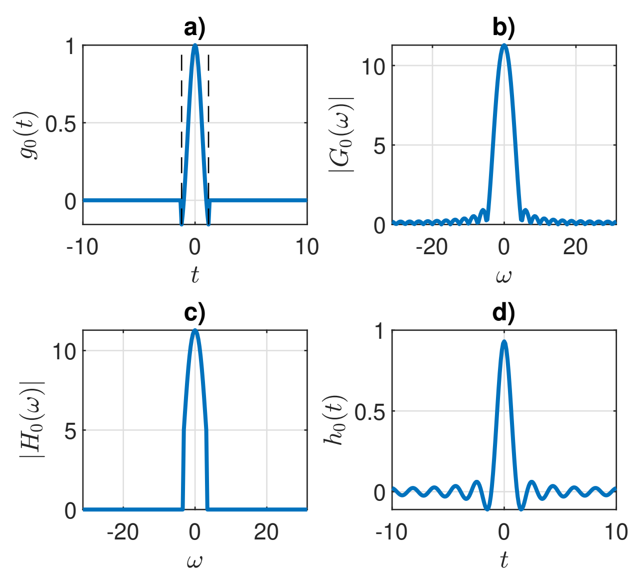



Fig. 1. a) a time limited signal $g_0(t)$ with black lines denoting the observation interval; b) its frequency spectrum $G_0(\omega)$; c) band-limited signal $H_0(\omega)$; d) band-limited signal in time domain $h_0(t)$.

In the last step of the first iteration, the known signal $g_0(t)$ is substituted within the known time interval $t \in \langle -1, 1 \rangle$, as denoted using dashed lines in Fig. 2. The general equation for the next signal iteration $g_n(t)$ can be written as

$$g_{n+1}(t) = [1 - w(t)]h_n(t) + g_0(t) \tag{2}$$

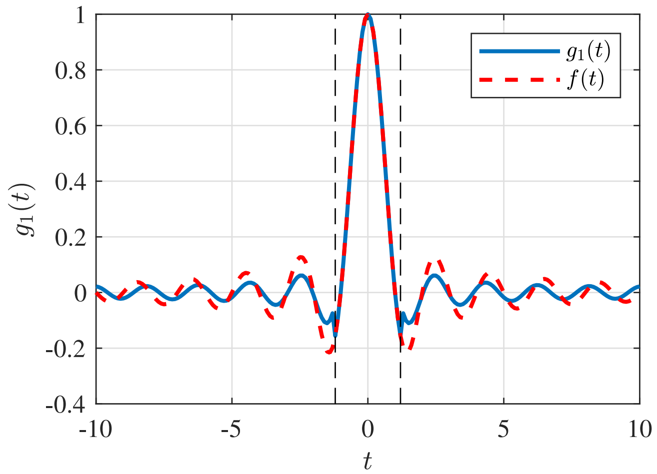



Fig. 2. The first iteration, $g_1(t)$, of the GP algorithm. The black lines denote the observation interval.

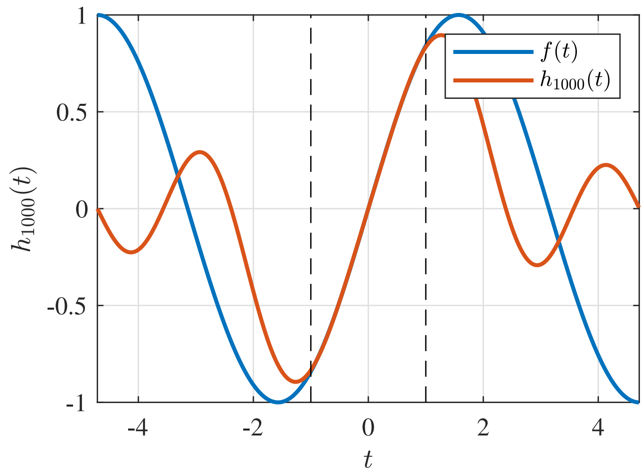



Fig. 4. Extrapolated smooth signal $h_{1000}(t)$ after $n = 1000$ iterations with black lines denoting the observation interval.

This method is iterative and can be used as many times as desirable by alternating between (1) and (2).

## III. Periodic Implementation

One option for implementation is to utilize MATLAB functions `fft` and `ifft` with operations from Section II resulting in extrapolation in Fig. 3. Signal $f(t) = \mathrm{sinc}(t-1)$ known in time interval from 1 to 3 seconds is shown. We can observe the extrapolation improving with successive iterations.

However, the `ifft` function considers only periodic signals (vectors). This affects the quality of approximation of the signal. It is shown in Fig. 4, where more extreme situation is used. The signal $\sin(t)$ approximated by vector in range from $\frac{-3\pi}{2}$ to $\frac{3\pi}{2}$ does not converge to zero at its edge values. This results in poor quality extrapolation. The size of $g_0(t)$ vector has no effect. Values from the interval $t \in \langle -1, 1\rangle$ are used.

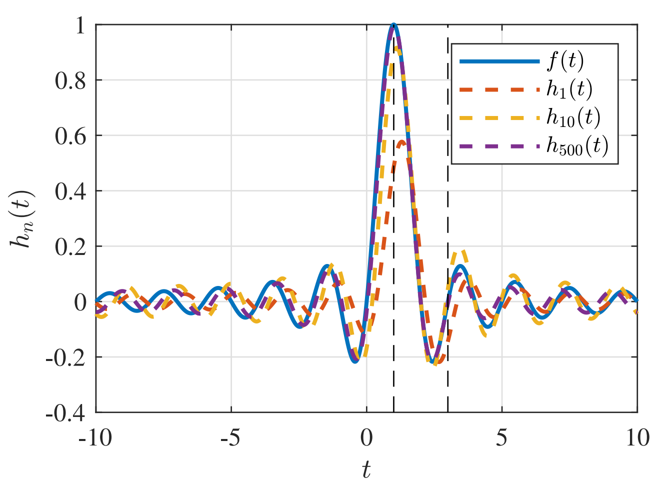



Fig. 3. Extrapolated smooth signal $h_n(t)$ after $n$ iterations with black lines denoting the observation interval.

## IV. Aperiodic Implementation

In this section we derive the formulae required for aperiodic implementation. Subsection IV-B modifies them so that they are more suitable for computer implementation.

### A. Direct aperiodic approach

Suppose we want to approximate the sampled signal $g_0(kT_\mathrm{s})$ from $\frac{-NkT_\mathrm{s}}{2}$ to $\frac{NkT_\mathrm{s}}{2}$, where $N$ is the number of samples of the signal. The frequency components are calculated by the Fourier transform.

$$G_n(\omega) = \mathcal{F}\{g_n(kT_\mathrm{s})\} = \sum_{k=-N/2}^{N/2} g_n(kT_\mathrm{s})\mathrm{e}^{-j\omega kT_\mathrm{s}} \tag{3}$$

We limit this signal by a window $W(\omega)$ in the frequency domain.

$$W(\omega) = \begin{cases} 1 & \text{for} \quad |\omega| < \Omega \\ 0 & \text{otherwise} \end{cases} \tag{4}$$

Then we convert it by inverse fourier transform into the time domain to obtain vector $h_n$ of the length $l$.

$$\begin{aligned} h_n(lT_\mathrm{s}) =& \mathcal{F}^{-1}\{G_n(\omega)W(\omega)\} \\ =& \frac{1}{2\pi}\int_{-\infty}^{\infty} G_n(\omega)W(\omega)\mathrm{e}^{j\omega lT_\mathrm{s}}d\omega \end{aligned}$$

Window $W(\omega)$ zeroes elements of $G(\omega)$ outside the interval $\langle -\Omega, \Omega\rangle$, so it can be expressed by the finite limits of the integral. Than we replace $G_n(\omega)$ with (3), reading

$$h_n(lT_\mathrm{s}) = \frac{1}{2\pi}\int_{-\Omega}^{\Omega} \sum_{k=-N/2}^{N/2} g_n(kT_\mathrm{s})\mathrm{e}^{-j\omega(k-l)T_\mathrm{s}}d\omega$$

Since the summation is not frequency dependent, it can be removed before the integral.

$$h_n(lT_\mathrm{s}) = \frac{1}{2\pi}\sum_{k=-N/2}^{N/2} g_n(kT_\mathrm{s})\int_{-\Omega}^{\Omega} \mathrm{e}^{-\mathrm{j}\omega(k-l)T_\mathrm{s}}d\omega \tag{5}$$

Then we calculate the integral using Euler's relation.

$$\int_{-\Omega}^{\Omega} \mathrm{e}^{-\mathrm{j}\omega(k-l)Ts} d\omega = \frac{1}{-j(k-l)T_s}(\mathrm{e}^{-\mathrm{j}\Omega(k-l)T_\mathrm{s}} - \mathrm{e}^{\mathrm{j}\Omega(k-l)T_s})$$
$$= \frac{1}{j(k-l)T_s}(\mathrm{e}^{\mathrm{j}\Omega(k-l)T_\mathrm{s}} - \mathrm{e}^{-\mathrm{j}\Omega(k-l)T_\mathrm{s}}) =$$
$$= \frac{2\sin\Omega(k-l)T_\mathrm{s}}{(k-l)T_\mathrm{s}}$$

The result is plugged into the equation (5):

$$h_n(lT_\mathrm{s}) = \sum_{k=-N/2}^{N/2} g_n(kT_\mathrm{s}) \frac{\sin\Omega(k-l)T_\mathrm{s}}{\pi(k-l)T_\mathrm{s}} \tag{6}$$

As mentioned before, the last step is to replace the known values with $g_0(kT_\mathrm{s})$ or $w(lT_\mathrm{s})f(lT_\mathrm{s})$ vector:

$$g_{n+1}(lT_\mathrm{s}) = w(lT_\mathrm{s})f(lT_\mathrm{s}) + [1-w(lT_\mathrm{s})]h_n(lT_\mathrm{s}) \tag{7}$$

### B. Modified aperiodic approach

For easy work in Matlab it is convenient to calculate everything in matrices. Thus, we introduce the operator B [3], [4], which represents the $\mathrm{sinc}(t)$ function from (6). Or rather, it represents the frequency constraint.

$$Bx(kT_\mathrm{s}) = \sum_{k=-\infty}^{\infty} x(kT_\mathrm{s}) \frac{\sin\Omega(t-kT_\mathrm{s})}{\pi(t-kT_\mathrm{s})} \tag{8}$$

The next introduced operator is D representing the time constraint. Using D we can select only $k$ elements from the vector $t$.

$$Dx(t) = \begin{cases} x(t) & \text{for} \quad t = kT_\mathrm{s} \\ 0 & \text{otherwise} \end{cases} \tag{9}$$

Using matrices $B$ and $D$ the resulting relation (7) can be written (using time vector $t$):

$$g_{n+1}(t) = Df(t) + h_n(t) - Dh_n(t) \tag{10}$$

Vector $h_n$ will be calculated as

$$h_{n+1}(t) = Bg_{n+1}(t) = Bh_n(t) - BD[h_n(t) - f(t)].$$

Since this vector is already frequency limited, we do not need to apply the B operator to it and the equation simplifies to:

$$h_{n+1}(t) = h_n(t) - BD(h_n(t) - f(t)) \tag{11}$$

In Fig. 5 we can observe the application of (11) to more complex signal $2\cos(t) + t$ with 33 known samples at time $kT_s$. Note that the resulting signal (vector) is not periodic and the approximation is accurate enough after 100 iterations. For this implementation, it is an interpolation rather than an extrapolation, since we know fewer values than in the previous implementation from Section III. You can see values of the operator B in (12) for this specific approximation.

$$B = \begin{bmatrix} \frac{\sin\Omega(-10+16T_\mathrm{s})}{\pi(-10+16T_\mathrm{s})} & \cdots & \frac{\sin\Omega(-10-16T_\mathrm{s})}{\pi(-10-16T_\mathrm{s})} \\ \vdots & \ddots & \vdots \\ \frac{\sin\Omega(10+16T_\mathrm{s})}{\pi(10+16T_\mathrm{s})} & \cdots & \frac{\sin\Omega(10-16T_\mathrm{s})}{\pi(10-16T_\mathrm{s})} \end{bmatrix} \tag{12}$$

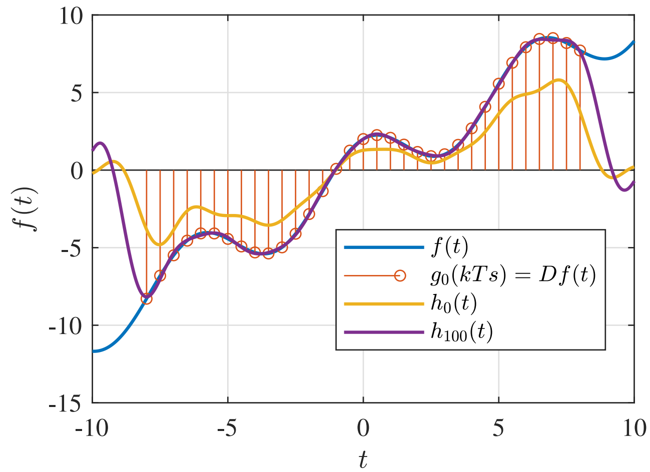



Fig. 5. Approximation of signal $2\cos(t) + t$ after 100 iterations.

## V. The effect of bandwidth

The choice of the maximum frequency $\Omega$ is an important factor. This value can affect the quality of the approximation. Let's define the signal

$$f(t) = e(t) + 0.1\sin(\frac{11\pi}{64}(4t-3))$$
$$\text{where } e(t) = \begin{cases} 1 \text{ for } t = 0 \\ \dfrac{\sin(\frac{11\pi}{128}4t)}{11\sin(\frac{\pi}{128}4t)} \text{ otherwise,} \end{cases} \tag{13}$$

and for one case, let's set $\Omega = 1$ and the other case $\Omega = \pi$. For this signal, as can be seen in Fig. 6, it is preferable to set $\Omega = 1$, since we observe better approximation here. In general, the more oscillatory the approximated signal is, the larger $\Omega$ should be.

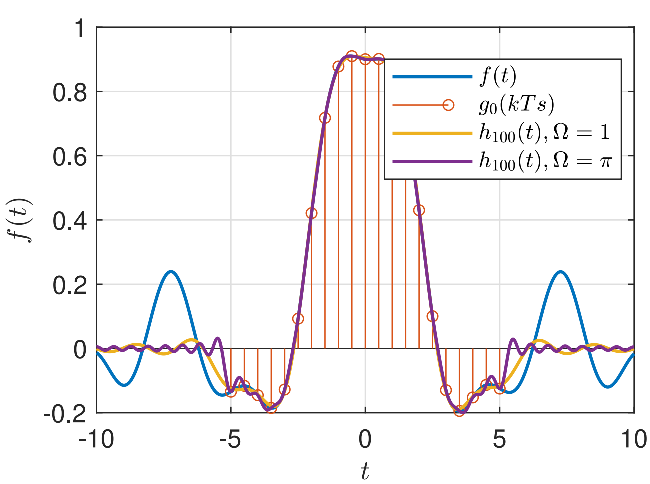



Fig. 6. The difference between setting $\Omega$ to 1 and $\pi$.

## VI. Slepian method comparison

In the last part, we compare the algorithm with reconstruction using Slepian basis described in more detail in papers [2], [5]. In simple terms, the Slepian method constructs a series consisting of Slepian sequences that represent the discrete signal; a concept similar to the use of sines and cosines in the Fourier series [6]. The inverse Slepian transform allows us to evaluate the signal approximation, the Slepian series, for any $t \in (-\infty, \infty)$. In Fig. 7 we can see the approximation of (13) by a periodic implementation $h_p(t)$, non-periodic implementation $h_n(t)$ and the approximation by Slepian sequences $s(t)$ from 21 known samples $g(kT_\mathrm{s})$.

One factor evaluating the quality of approximation is the implementation time complexity described in Table I. Next, it is the global quadratic error [7], shown in Table II, which is calculated as

$$E = \int_{-\tau}^{\tau} |f(t) - \hat{f}(t)|^2 dt \tag{14}$$

where $f(t)$ is the true signal and $\hat{f}(t)$ is the approximation constructed either by the Slepian or GP method. Table II compares the errors, $E$, for different methods and for both interpolation, $\tau = 5$ and extrapolation in a finite interval for $\tau = 10$.

The periodic implementation has been modified for better comparison, so that iterpolation can be seen. Only 41 samples were worked with and after the GP algorithm was finished, the signal samples were densified by using `fft` and zero padding. Therefore, there is a better time consumption but a worse accuracy of approximation and interpolation than the algorithm using operators. Note that the next element after the last one of the signal $h_p(t)$ would correspond to the first one and we observe a periodic signal.

Comparing the non-periodic implementation with reconstruction with Slepian sequences, notice that Slepian method is better in both aspects, even in pure interpolation.

TABLE I
Time comparison

| | |
|---|---|
| time-consuming for the periodic implementation.* | $4.2 \pm 0.6$ ms |
| time-consuming for the non-periodic implementation.* | $139 \pm 4$ ms |
| time-consuming for approximation with Slepian sequences.* | $1.7 \pm 0.2$ ms |

*The measurements were taken 10 times.

TABLE II
Global error $E$ computed in $t \in \langle -10, 10 \rangle$.

| | |
|---|---|
| Global error of the periodic implementation | 0.185 |
| Global error of the non-periodic implementation | 0.131 |
| Global error of the approximation with Slepian sequences | $4.4 \cdot 10^{-3}$ |
| *Global error of the periodic implementation | $3.9 \cdot 10^{-3}$ |
| *Global error of the non-periodic implementation | $5.7 \cdot 10^{-6}$ |
| *Global error of the approximation with Slepian sequences | $10^{-15}$ |

*Error computed in $t \in \langle -5, 5 \rangle$.

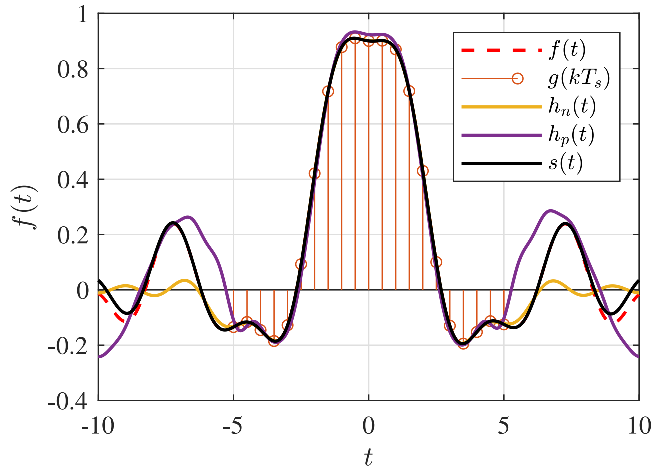



Fig. 7. Approximation of the signal $f(t)$ using GP implementation, $h_p(t)$, $h_n(t)$, and approximation using Slepian sequences $s(t)$.

## VII. Conclusion

In this paper we revisited the original GP algorithm and proposed a modification which avoids the spurious periodic extrapolation of the signal caused by FFT. Three distinct methods of reconstruction were implemented in Matlab: FFT-based GP algorithm, proposed modification of the GP algorithm and the Slepian method. Our comparison of their complexity and accuracy shows that the Slepian method is in all aspects preferable to the tested variants of the GP method. Slepian method surpassed GP algorithm both in accuracy and computational speed.


## Acknowledgment

The completion of this paper was made possible by the grant No. FEKT-S-23-8451 – “Research on advanced methods and technologies in cybernetics, robotics, artificial intelligence, automation and measurement” financially supported by the Internal science fund of Brno University of Technology.


## References


[1] A. Papoulis, “A new algorithm in spectral analysis and band-limited extrapolation,” *IEEE trans. circuits and syst.*, vol. 22, no. 9, 1975. DOI:10.1109/TCS.1975.1084118

[2] D. Slepian, “Prolate spheroidal wave functions, fourier analysis, and uncertainty - V: the discrete case,” *Bell Syst. Tech. J.*, vol. 57, no. 5, pp. 1371–1430, 1978.

[3] O. Mihálik, “Demystifying Band-Limited Extrapolation,” in *Proc. 27th Conf. STUDENT EEICT 2021.* Brno: Vysoké učení technické v Brně, Fakulta elektrotechniky a komunikačních technologií. 2021, pp. 576–580.

[4] O. Mihálik, “Reconstruction of non-uniformly sampled signals using gerchberg-papoulis method.” in *Proc. II 26th Conf. STUDENT EEICT 2020.* Brno: Vysoké učení technické v Brně, Fakulta elektrotechniky a komunikačních technologií. 2020, pp. 136–140.

[5] V. Šimeček, “Diskrétní Slepianova ortogonální báze a její využití pro získání spektra signálů.” B.S. thesis, Dept. Control and Instrum., Brno Univ. Tech., Brno, 2023.

[6] A. Kufner and J. Kadlec, *Fourierovy řady,* 1st ed. Praha: Academia, 1969.

[7] A. Papoulis, *Signal analysis.* New York: McGraw-Hill Book Company, 1977.

# Experimental workplace for analysis of batteries

Karel Kvasnička
*Department of Electrical Power Engineering*
*Brno University of Technology*
Brno, Czech Republic
238474@vutbr.cz

***Abstract*—To find out the condition and the capacity of a battery they are discharged and charged, while the voltage and the current of the battery are measured. Parameters of batteries are changing with aging and usage. There are many types of batteries that are divided into two main categories, primary batteries and secondary batteries. Their properties depend on used materials and production technique. An experimental workplace was built for battery testing equipped with a regulated power supply, an electronic load, a data acquisition unit and safety components. The workplace is able to test various types of batteries. Measurements are automated with computer program.**

***Keywords*— Batteries, analysis of batteries, testing of batteries, discharging and charging of batteries**

***Introduction***

Batteries are becoming more important as they are increasingly used in vehicles, handheld devices and other electrical devices. Batteries use chemical reactions to produce electrical energy. Parameters such as energy capacity, voltage and current capabilities depend on materials and production technologies. Each type of a battery has different properties.

During discharge the voltage of the battery is changing. With usage and aging of batteries their energy capacity is decreasing. To measure energy capacity of a battery we have to discharge the battery, while measuring current and voltage. Simple measuremts of a battery discharge shows only total energy capacity. With continuous measuments and data recording it is posible to represent voltage and current changes during discharge.

## I. Batteries and their sorting

### A. Basic construction of a battery

Each battery consist of one or more cells. These cells are connected in series or parallel to get higher voltage or current. Each cell consists of a positive and a negative electrode. Between electrodes is a electrolyte and a separator, which create electrical isolation but conduct ions. Electrical energy is obtained by chemical reactions between electrodes. [1]

### B. Basic parameters of batteries

Used materials create main diferences in batteries. Materieals influence energy capacity, weigth, voltage, current capabilities, recharging, operating temperaturs and self discharge. Main two groops are primary and secondary batteries. Primary batteries are nonrechargable. When their energy is depleted, they can´t be charged to replenish the chemical energy. On the other hand secondary batteries can be charged to restore their chemical energy.[2]

Important parameters of a battery are capacity and nominal voltage. Nominal voltage is mean voltage during discharge of the battery. Capacity, used mainly with secondary batteries, is determined by discharging a battery according to norms.[1][2]

### C. Primary zinc batteries

Primary zinc batteries use zinc metal in the negativ electrode. Manganese oxide, oxygen or silver oxide are used as the positive electrode. [2]

Widely used zinc-carbone and akaline batteries have electrodes made of zinc and manganese oxide, the difference between them is in the construction and electrolytes. Nominal voltage is 1.5V. [2][3]

Energy density of an alkaline battery is around 200Wh/kg.[4]

Energy density of a zinc-carbon battery is around 40Wh/kg.[5]

### D. Primary lithium batteries

Primary lithium batteries use metal lithium as their negative electrode. Materials used as the positive electrode are for example pyrite, manganese oxide, iodine.[3][4]

Commonly used material is manganese oxide [4]. Lithium salts are used as an electrolyte [3]. Nominal voltage per cell is from 1.5V to 4V. Energy density of primary lithium batteries is usually about 300Wh/kg. Primary lithium batteries using Thionyl chloride have energy density up to 500Wh/kg.[4]

### E. Secondary lithium batteries

Secondary lithium batteries have lithium in their negative electrode. Lithium can be used in two forms as a metal or as an ion in a carbon grid. Metal lithium has risks with high reactivity and wrong metal deposition while charging. For higher safety lihium as an ion in a carbon grid, known as lithium-ion, is used, although with lower energy density as a trade-off.[3]

For positive electrode are used oxide salts with lithium. As an electrolyte are used lithium salts dissolved in a dissolvent. [6][7] Nominal voltage per cell is 1.5 to 3.75V [2][7]. Ordinary lithium-ion batteries have energy density 50 to 260 Wh/kg [6]. Experimental batterie claims to have more than 700Wh/kg.[8]

Combination of constant curent and constant voltage sources are used to charge litium-ion batteries. Battery is firstly charge with constant current to set voltage, than it´s charge constantly with set voltage. Charging ends after set amount of time or current falling to set minimum.[2]

### F. *NiCd and NiMH secondary batteries*

NiCd and NiMH batteries have the positive electrode made of nickel oxides and metal. NiCd uses metal cadminum as the negative electrode. NiMH uses metal alloys which store hydrogen as the negative electrode. Both NiMH and NiCd use alkaline electrolyte. [3]

Nominal voltage per cell is 1.2V. NiMH is succesor to NiCd, because NiMH lacks poisonous cadmium [3]. Energy density of NiCd batteries is betweem 45 to 80Wh/kg, NiMH batteries have 60 to 120Wh/kg [6].

For charging of NiCd and NiMH batteries is used normal and conservative current. Normal charging end after certain time, voltage drop or temperatur rise. Charging with conservative current uses smaller current than normal current charging, but the charging is continuous.[1][2]

### G. *Lead acid secondary batteries*

Lead acid secondary batteries have the positive electrode made of lead oxide and the negative electrode made of metal lead. As the electrolyte sulfuric acid is used. [3]

Differences in lead acid batteries are mostly in construction of the batteries. Lead acid batteries have two main categories, open lead acid batteries and valve regulated lead acid batteries (VRLA). Open lead acid batteries need occasional water top-up to counter water evaporation. Valve regulated lead acid batteries use gelled or glass fiber thicken electrolyte. Thanks to this water top-up are not needed and electrolyte can not spill. A valve is used to protect the battery from overpressure mainly during charging.[2][3]

Nominal voltage per cell is 2V [1]. Energy density is 30 to 50Wh/kg [6].

To charge the lead acid batteries multiple types of powersupply can be used. Most common are constant voltage chargers.[1][9]

## II. EXPERIMENTAL WORK PLACE

### A. *Measuring instruments and safety relay*

Agilent N3301A with the modul N3302A is used as a electronic load for discharging batteries. The N3302A module is able to load a battery up to 150W. Maximum voltage is 60V and maximum current is 30A. The electronic load uses constant current or constant voltage modes.

Agilent E3631A is used as a power supply for charging the batteries. Its voltage can be set between 0-25V and current between 0-1A. It automatically switches between constant voltage and constant current modes.

Agilent 34970A data aquisition unit with two modules 34901A is used to measure multicell batteries, temperature and exact total battery voltage. First module is used solely to measure voltage of individual cells of a battery. Up to 20 cells can be measured. Second module is used for temperature and total battery voltage measurements.

Safety relay is used to disconnect a battery from measuring instruments to ensure safety when unexpected behaviour occures. The relay is controled by Arduino Uno board, which checks temperature of battery and functioning of control program. If temperature rises above 42 °C or the control program does not communicate set amount of time, Arduino disconnects the battery automatically.

### B. *Instruments connection of the experimental workplace*

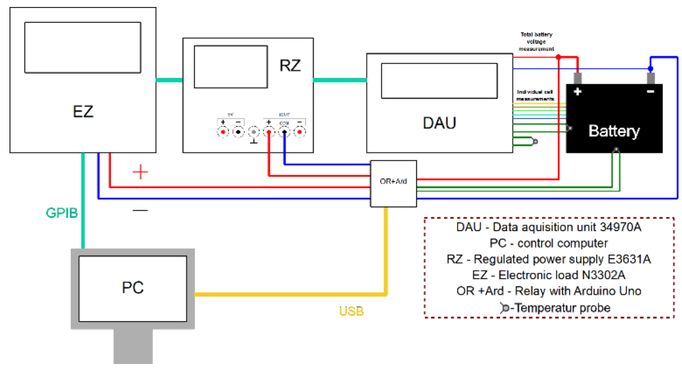



*Fig. 1: Connection scheme*

Connection of instruments to the battery is shown in Fig. 1. The power supply and the electronic load are connected into the safety relay. The safety relay connects to battery terminals. Agilent 34901A is connected to battery directly.

Measuring instruments comunicate with computer through GPIB to USB interface. GPIB is connected to electronic load, power supply and data aquisiton unit. The relay has its separate 12V powersupply. Arduino Uno is connected through USB to the computer.

### C. *Control program*

The control program is written in Matlab. It has UI interface to control charging and discharging of batteries. User chooses current limits, minimal and maximal voltages, number of cells and switch-off conditions.

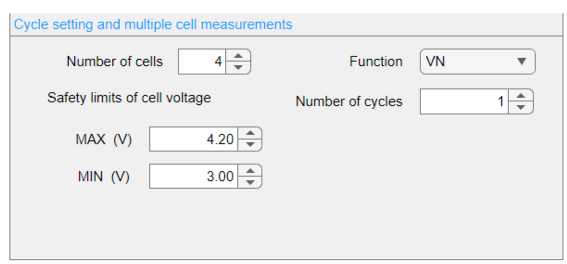



*Fig. 2: Cycle setting*

In cycle setting (Fig. 2) firstly type of the cycle is chosen under function. The first type is discharging then charging. The second type is charging then discharging. The third type is only charging. The fourth type is only discharging. The first and second type can run multiple time with setting cycle counts. For voltage measurements of individual cells of multiple cell batteries, number of cells must be set. Also minimum and maximum voltage of cell is set to protect cells againts overcharging or overdischarging.

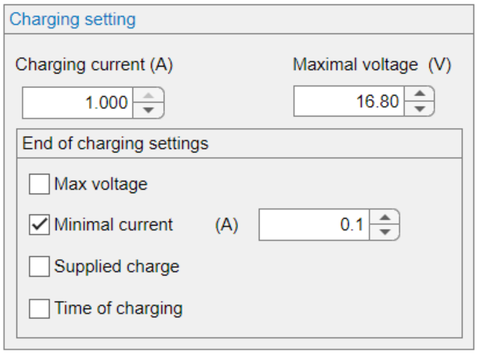



*Fig. 3: Charging setting*

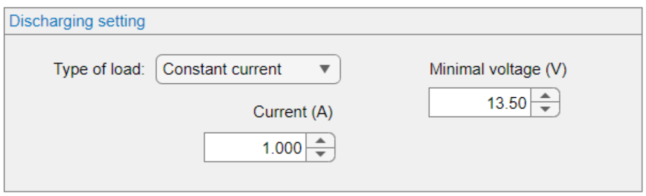



*Fig. 4: Discharging setting*

In charging setting (Fig. 3) and in discharging setting (Fig. 4) progress of charging/discharging is set. For discharging user sets minimum voltage, which is used to end discharging of a battery, and a discharging methods, which are constant current or constant resistance. For charging user sets maximum voltage, maximum current and end of charge conditions. User can use multiple conditions like minimum current, supplied charge, achieving maximum voltage or set time of charging.

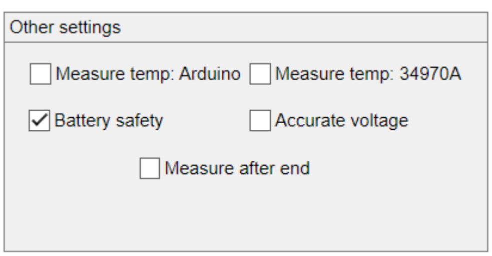



*Fig. 5: Other settings*

In other settings (Fig. 5) additional measurements can be switched on like temperature and accurate total battery voltage measurements.

After setting all values, control program asks user for a name of the measurement. Then the measurement can be started.

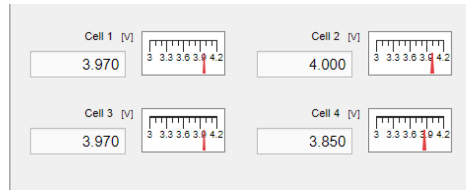



*Fig. 6: Cell measurements values*

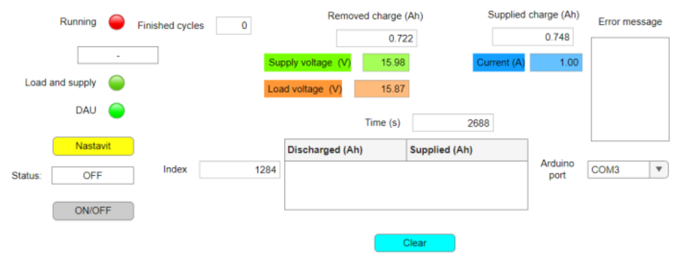



*Fig. 7: Other measurements values and control buttons*

Currently measured values are shown in control program as seen in Fig. 6 and Fig. 7. Each value has its corresponding description and unit of measurement. Each measurement takes 1 to 2 seconds to complete. The values are continuously stored in .txt file under the chosen name of the measurement. If an error occurs, corresponding error message is shown. If the whole control program crashes, measured values are not lost.

### D. *Measurements inaccuracy*

Measurements of the battery are burdened with inaccuracies of the measuring instruments, shown in TABLE I. TABLE II. and TABLE III.

TABLE I. ELECTRONIC LOAD ACCURACY [10]

| | | | |
|---|---|---|---|
| **Constant current setting** | ***range*** | 3 A | 30 A |
| | ***accuracy*** | ± 0,1 % +5 mA | ± 0,1 % +10 mA |
| **Current measurement** | ***range*** | 3 A | 30 A |
| | ***accuracy*** | ± 0,05 % +3 mA | ± 0,05 % +6 mA |
| **Voltage measurement** | ***range*** | 6 V | 60 V |
| | ***accuracy*** | ± 0,05 % +3 mV | ± 0,05 % +8 mV |

TABLE II. POWER SUPPLY ACCURACY [11]

| **output** | 1 | 2 | 3 |
|---|---|---|---|
| **Voltage setting** | ± 0,1 % +5 mV | ± 0,05 % + 20 mV | |
| **Current setting** | *± 0,2 % +10 mA* | ± 0,15 % + 4 mA | |
| **Voltage measurement** | *± 0,1 % +5 mV* | ± 0,05 % + 10 mV | |
| **Current measurement** | ± 0,2 % +10 mA | ± 0,15 % + 4 mA | ± 0,2 % + 10 mA |

TABLE III. DAC UNIT VOLTAGE MEASUREMENTS ACCURACY [12]

| **range** | Accuracy after 1 year at 23±5 °C |
|---|---|
| 100 mV | ± 0,0050 %+0.0040 % |
| 1 V | ± 0,0040 %+0.0007 % |
| 10 V | ± 0,0035 %+0.0005 % |
| 100 V | ± 0,0045 %+0.0006 % |
| 300 V | ± 0,0045 %+0.0030 % |

Sampling of the data by control program, which takes 1 to 2 seconds, also cause a little inaccuracy by itself. Charge computation takes time difference between current and last measurement to determine the time which is used to multiply currently measured current to calculate the charge which is added to the total charge.

The resistance of the connections between the measuring instruments and battery causes voltage drop which affects voltage measurements by the electronic load and the power supply. This voltage drop rises with higher currents. The unwanted resistence is caused by cables, relay contacts and connectors. Effect of this can be mitigated with use of DAU to measure exact voltage near battery. When using constant resistence mode for discharging, then this unwanted resistence must be accounted for. Constant current discharging in combination with DAU accurate voltage measurements is uneffected by this. Charging curves can be slightly distorted. This resistance can differ with use of adapters, which are used to connect to the battery.

For accurate measurent the battery must be connected with its appropriate connectors or suitable alternative. Resistance between the battery and its connections can be considered as a representation as in a real device, thereby not taken as an inaccuracy.

## ACKNOWLEDGEMENT

I would like to thank Mr. Ing. Jiří Majzner Ph.D. for consultations and help with solving problems.

## ABBREVIATIONS

GPIB-General purpose interface bus

## REFERENCES

[1] ARENDÁŠ, Miroslav a RUČKA, Milan. *Nabíječky a nabíjení*. 2. aktualiz. vyd. Praha: BEN - technická literatura, 1997. ISBN 80-860-5629-5.

[2] BATTEX, SPOL. S R.O. *Abeceda baterií a akumulátorů*. Online. 2009, 2023. From: www.battex.info. [cit. 2024-03-01].

[3] HUGGINS, Robert A. *Energy storage: fundamentals, materials and applications*. Second edition. Cham: Springer, 2016. ISBN 978-3-319-21238-8.

[4] MICROPOWER BATTERY COMPANY. *PRIMARY BATTERY OPTIONS: A LOOK AT LITHIUM BATTERIES*. Online. MICROPOWER BATTERY COMPANY. MicroBattery.com. 2021, 2023. From: https://www.microbattery.com/primary-lithium-batteries. [cit. 2024-03-02].

[5] LI, Lucky. *Principle and performance analysis of zinc carbon battery*. Online. In: TYCORUN ENERGY. Takomabattery. 2022, 2023. From: https://www.takomabattery.com/zinc-carbon-battery/. [cit. 2024-03-02].

[6] CLOUD, Maria. *What is the Energy Density of a Lithium-Ion Battery?* Online. In: FLUX POWER. Flux Power. California: Flux Power, 2024. From: https://www.fluxpower.com/blog/what-is-the-energy-density-of-a-lithium-ion-battery. [cit. 2024-03-04].

[7] PISTOIA, Gianfranco. *Lithium-Ion Batteries - Advances and Applications*. Online. Oxford: Elsevier, 2014. ISBN 978-0-444-59513-3. From: https://app.knovel.com/hotlink/pdf/id:kt00U81K71/lithium-ion-batteries/front-matter. [cit. 2024-03-04].

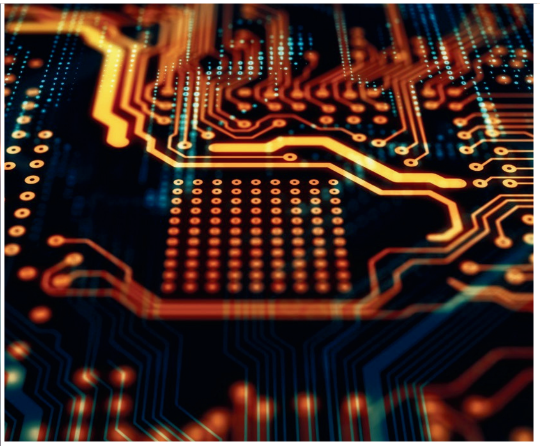
[8] HANLEY, Steve. *Chinese Researchers Announce 711 Wh/kg Lithium Battery*. Online. In: CLEANTECHNICA. CleanTechnica. 2024. From: https://cleantechnica.com/2023/04/25/chinese-researchers-announce-711-kwh-kg-lithium-battery/. [cit. 2024-03-04].

[9] *Technical-Manual.pdf: SEALED LEAD-ACID BATTERIES*. Online. In: POWER-SONIC CORPORATION. Power Sonic. 2024. From: https://www.power-sonic.com/wp-content/uploads/2018/12/Technical-Manual.pdf. [cit. 2024-03-04].

[10] *N3300A Series DC Electronic Loads.pdf: N3300 Series DC Electronic Loads*. Online. In: KEYSIGHT TECHNOLOGIES. Keysight. 2000, 2024. From: https://www.keysight.com/us/en/assets/7018-06712/data-sheets/5980-0232.pdf. [cit. 2024-03-27].

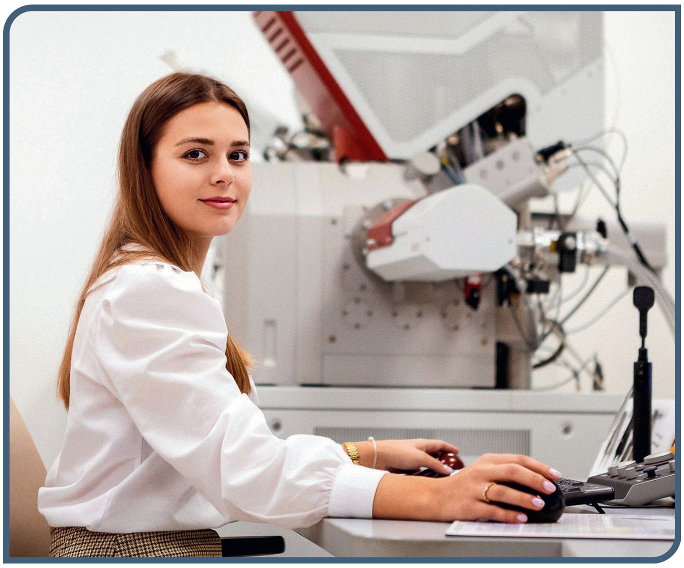
[11] *5968-9726.pdf: Keysight E363xA Series Programmable DC Power Supplies*. Online. In: KEYSIGHT TECHNOLOGIES. Keysight. 2000, 2024. From: https://www.keysight.com/us/en/product/E3631A/80w-triple-output-power-supply-6v-5a--25v-1a.html. [cit. 2024-03-27].

[12] *34970A Data Acquisition_Switch Unit Family.pdf: 34970A Data Acquisition/ Switch Unit Family*. Online. In: KEYSIGHT TECHNOLOGIES. Keysight. 2000, 2024. From: https://www.keysight.com/us/en/product/34970A/34970a-data-acquisition-control-mainframe-modules.html. [cit. 2024-03-27].



onsemi™

# VÁHÁŠ, JAKÝM SMĚREM SE VYDAT?



## VRHNI SE DO SVĚTA POLOVODIČŮ A UTVÁŘEJ NOVÉ TECHNOLOGICKÉ TRENDY!

**onsemi** je světový **technologický lídr** v oblasti **polovodičů**, a to od jejich **vývoje až po finální výrobu součástek**. Jsme mezinárodní společnost s bohatou historií v České republice, kde působíme v **Rožnově pod Radhoštěm** a **Brně**. Zabýváme se **návrhem** integrovaných obvodů, **výrobou** desek z křemíku i karbidu křemíku a také výrobou polovodičových čipů, ale také samotným **vývojem technologií** výroby polovodičových materiálů i polovodičových součástek.

**~60** stážistů za rok

**7 mld Kč** investic do výroby v Rožnově (2021-2023)

zaměstnanců v Rožnově a Brně **2200**

patentů v ČR **>300**

# STUDUJEŠ OBOR V OBLASTI?

materiály · fyzika · elektro · chemie · IT · technologie

## PŘIJĎ K NÁM NA STÁŽ

- Můžeš k nám na letní a celoroční **stáž** Jako stážistu tě ohodnotíme **200–280 Kč/hod.** (podle ročníku a zkušeností)
- Můžeš získat stipendium až **5 000 Kč/měs.**
- Napiš si u nás svou bakalářskou nebo diplomovou práci
- U dlouhodobých stáží si můžeš vyzkoušet práci v různých týmech
- Zajistíme ti **bezplatné ubytování** (Rožnov p. R.) nebo **příspěvek na bydlení** (Brno)

## PŘIJĎ K NÁM PRACOVAT

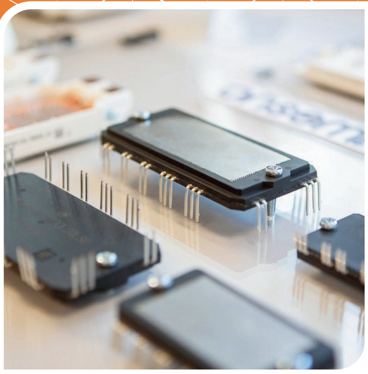
- Nabízíme ti velmi atraktivní startovní plat a **zrychlený platový růst** (navýšení platu až o **20 %** po 18 měsících)
- 25 dní dovolené + 3 dny sick days + bridge days
- **Přispějeme** ti na dopravu, bydlení, stravování, penzijko do výše 6 % z tvého platu a dostaneš od nás výhodný mobilní tarif
- Podpoříme tvé **další vzdělávání**, jazykové dovednosti, možnost získat zkušenosti v zahraničí a mnoho dalších výhod

## Zažij, jaké je to být součástí moderní firmy.

Výroba polovodičových materiálů (SiC a Si)

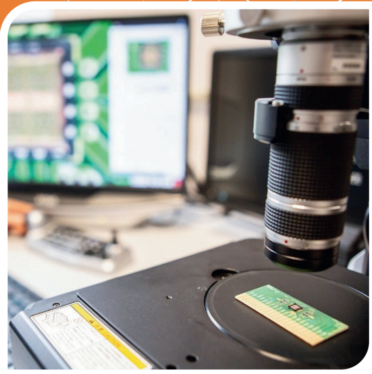
Špičkově vybavené laboratoře

Výroba čipů



Středisko návrhu součástek

IT oddělení



Vlastní výzkum a vývoj

onsemi™

Pro více informací si načti:

# Bacterial Identifier: Accelerating Bacterial Genome Detection

1st Julie Nejezchlebová
*Department of Biomedical Engineering*
*Faculty of Electrical Engineering and Communication,*
*Brno University of Technology*
Brno, Czech Republic
xnejez10vutbr.cz

2nd Ivan Rychlík
*Department of Microbiology and Antimicrobial Resistance, Veterinary Research Institute*,
Brno, Czech Republic
ivan.rychlik@vri.cz

3rd Jana Schwarzerová
*Department of Biomedical Engineering*
*Faculty of Electrical Engineering and Communication,*
*Brno University of Technology*
Brno, Czech Republic
*Department of Molecular and Clinical Pathology and Medical Genetics, University Hospital Ostrava, Ostrava, Czech Republic*
*Molecular System Biology (MOSYS), University of Vienna,*
Vienna, Austria
Jana.Schwarzerova@vut.cz

***Abstract* — Bacterial identification is crucial for effectively monitoring and controlling the spread of infectious diseases. In addressing this critical need, our study introduces an engineered application designed to expedite the analysis of bacterial sequencing data, thus providing a streamlined method for species identification. The Bacterial Identifier underwent thorough testing using a significant dataset obtained from the Veterinary Research Institute. Central to the app's functionality is the integration of three essential tools. Implemented through a cohesive bash script, these tools are seamlessly combined to ensure optimal performance and accuracy in bacterial identification. Furthermore, to enhance user experience, a user-friendly interface was developed using Python 3, facilitating intuitive navigation and efficient utilization of the application's capabilities.**

***Keywords — Microbiota, Genetic-level identification, Bioinformatic analysis, Bacterial identification app***

## I. Introduction

Identifying bacterial strains remains an ongoing challenge for human health, especially those carrying antibiotic resistance genes or displaying heightened virulence [1]. Consequently, there is a growing need for genetic-level identification of bacterial strains [2][3].

Besides traditional laboratory techniques used for bacterial identification, bacterial identification can be performed using tools that compare bacterial genomes directly with a reference database. Among these tools, BLAST (Basic Local Alignment Search Tool) is widely used in bioinformatics, especially for smaller databases. However, for larger databases, BLAST [4] may not be suitable due to its long running time. For example, in the study by Kent [5], it is mentioned that to analyze genomes efficiently, rapid mRNA/DNA and cross-species protein alignments are essential. Consequently, the study by Kent [5] introduces a new tool, BLAT, which surpasses popular existing tools such as BLAST in both accuracy and speed, being 500 times faster [5].

This study focuses on the direct development of a software application featuring a user-friendly interface integrated with the bacterial database of the Veterinary Research Institute. The application utilizes BLAST, a traditional and widely used tool in the field, in conjunction with faster tools offering greater heuristic freedom.

Despite its limitations, BLAST remains important in veterinary medicine and future drug research, which is why it is an integral part of our designed application. Additionally, our software app is tailored for rapid bacterial identification, particularly for large databases, and it utilizes the Cd-hit tool [6], which has demonstrated and discussed suitability in our previous research [7][8].

## II. Datasets

The robustness of the bacterial database, for which the software application is designed, is broad, as confirmed by a brief summary analysis of the tested dataset. This bacterial database was initially published in the study by Medvedcky et al [9] and expanded with additional pigs samples in the study by Schwarzerova et al [3]. The summary analysis of dataset was constructed using the Up-to-date Bacterial Core Genome (UBCG) [9] by phylogenetic methods and design by IToL v6 [11], see Fig. 1. It represents 452 genome sequences of 8 bacterial phyla – *Firmicutes, Bacteroides, Verrumicrobiota, Actinobacteria, Elusimicrobiota, Proteobacteria, Synergistes* and *Fusobacteria*, providing a comprehensive perspective on the microbiome.

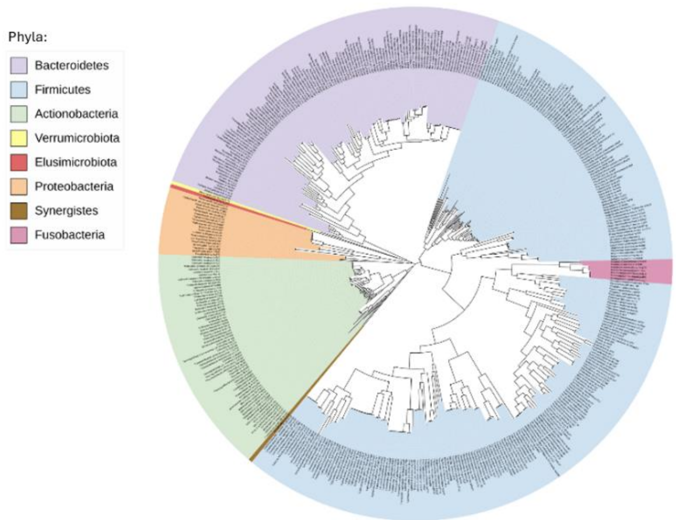



Fig. 1. Phylogenetic tree of 452 sequenced genomes obtained from chicken caecum and pig feces. Bacteria from the phylum Firmicutes (254 genomes) are shown in blue, bacteria from the phylum Bacteroidetes (113 genomes) are shown in purple, the green color represents bacteria from the Actinobacteria phylum (65 genomes), the yellow color represents the Verrumicrobiota (1 genome), the red color represents the Elusimicrobiota (1 genome), the orange color represents bacteria from the Proteobacteria strain (19 genomes), the brown color represents the Synergistes (1 genome) and the pink color represents the Fusobacteria (7 genomes).

## III. Methods

The Bacterial Identifier, a core stone in this study, has been implemented in the bash scripting language. This tool utilizes a combination of specialized software tools, specifically Barrnap [12], Cd-hit [6], and BLAST [4][13]. It was implemented through a cohesive bash script.

The backend of the Bacterial Identifier is structured into multiple phases. In the initial phase, bacterial identification is conducted using specialized tools. Subsequently, in the second phase, a detailed analysis of the output files from Barrnap, Cd-hit, and BLAST tools takes place, enabling a comprehensive understanding of bacterial genomes. In the final phase, a comparative analysis is executed, providing users with definitive identification results.

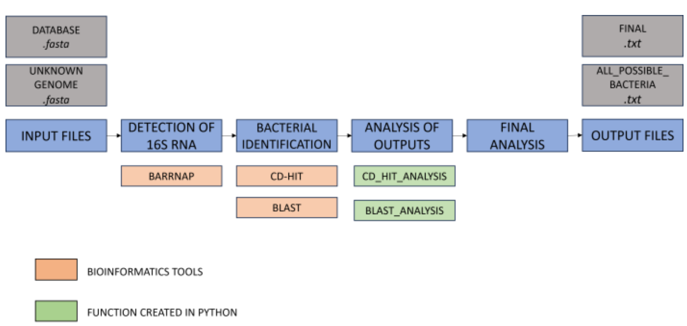



Fig. 2. Block diagram of the Bacterial Identifier. Blue color represents the process of the algorithm, the gray color represents the input and output files, the green color represents functions created in Python 3, and the orange color represents the previous implemented bioinformatics tool.

### 1) *Default Settings of the Bacterial Identifier*

To ensure optimal performance and quality results from the Bacterial Identifier, configuring parameters for individual tools such as Barrnap, Cd-hit, and BLAST is crucial. Tables I, II, and III provide a detailed overview of all configured parameters, their functions, and default settings within the Bacterial Identifier.

TABLE I: Barrnap Parameter Settings

| Parameter name | Description | Default Setting |
|---|---|---|
| **kingdom** | Kingdom of organisms to be analyzed | bac (bacteria) |
| **evalue** | E-value threshold | 1e-6 |
| **lencutoff** | Ratio of minimum predicted rRNA length to standard rRNA length | 0.8 (80%) |
| **reject** | Threshold for rejecting rRNA (matches with rRNA below this value will be excluded) | 0.25 (25%) |
| **outseq** | Name of the output file where the predicted sequences will be saved | rRNA.fasta |
| **threads** | Number of CPU threads used | 6 |

TABLE II: Cd-hit Parameter Settings

| Parameter name | Description | Default Setting |
|---|---|---|
| **I** | Input file (unknown genome) | rRNA.fasta |
| **i2** | Input multifasta file with reference genomes | multifasta.fasta |
| **O** | Output file | Cd-hit_output |
| **T** | Number of CPU threads used will be excluded) | 6 |
| **M** | Maximum memory used by Cd-hit (MB) | 0 (unlimited) |
| **C** | Sequence similarity threshold | 0.95 (95%) |

TABLE III: BLAST Parameter Settings

| Parameter name | Description | Default Setting |
|---|---|---|
| **evalue** | Minimum match for similarity | 1e-5 |
| **query** | File with unknown bacterium sequences | „$input_file“ |
| **db** | File with reference bacterium sequences | blast_database |
| **out** | Output file | blast.txt |
| **outfm** | Output file format | 6 qseqid sseqid pident length evalue bitscore |
| **num_threads** | Number of CPU threads used will be excluded) | 6 |

### 2) *Output File Analysis*

After executing Barrnap, Cd-hit, and BLAST, an analysis of the output files is performed. Due to the different formats of the output files from each tool, two separate Python 3 functions are developed for analyzing Cd-hit and BLAST outputs. No specific function is created for Barrnap; instead, it is checked whether the output file is empty, ensuring the presence of known rRNA genes in the unknown genome.

Despite the creation of two separate functions – *cd_hit_analysis* and *blast_analysis* – the underlying principle of both functions remains consistent. These functions parse the respective output files, saving only the names of bacteria exhibiting similarity with the unknown bacterium above a predefined threshold. These newly generated files then proceed to the final analysis.

### 3) *Final analysis*

The final steps of the process are conducted using the core backend of the application in a bash script and vary based on whether the user opts to identify the unknown bacterium using one or more tools after initiating the Bacterial Identifier. If only one tool is selected, the output files of the Bacterial Identifier consist of the results from the functions *cd_hit_analysis* and *blast_analysis*. In the case of multiple tools being chosen, the Bacterial Identifier generates two new output files.

The first file - final.txt, stores bacteria predicted by all selected tools, while the second file - all_possible_bacteria.txt, contains all bacteria predicted by at least one of the chosen tools.

### 4) *Frontend App Implementation*

To make the Bacterial Identifier accessible to a wider scope of users, a simple and intuitive user interface has been created using the Tkinter library [14] in Python 3. Upon launching the application, the main window appears, serving as a platform to configure parameters influencing the analysis, see Fig. 3.

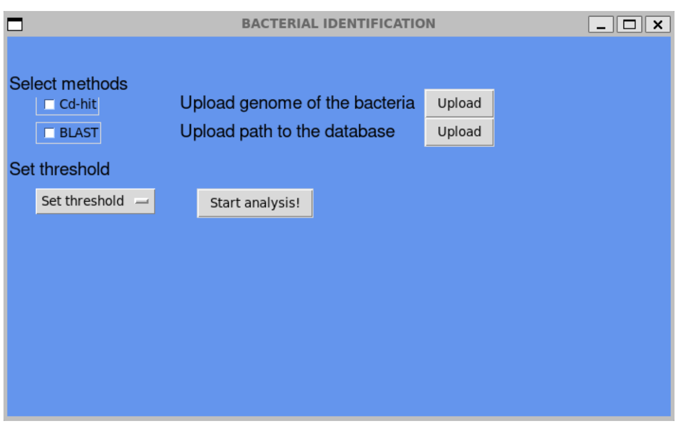



Fig. 3. The window that appears to the user when the application is launched.

In the Selection methods section, users have to the option to choose a method form bacterial identification, either the Cd-hit tool, the BLAST tool, or both simultaneously. If both tools are selected, both are automatically initiated for bacterial identification. Simultaneously, the Barrnap tool is automatically launched to detect 16S rRNA.

Another adjustable parameter is the similarity threshold (*Set threshold*), which determines how similar a sequence must be to be considered the same species. This parameter is optional; if the user does not enter a value, the application automatically sets the similarity threshold to 95%.

On the right side of the app, there are two buttons allowing users to set path to analyzed files. The first button facilitates uploading files with unknown bacteria in *.fasta* format, while the second button enables users to specify the path to reference genomes.

If all parameters are properly configured, users can press the *Start* analysis button, initiating the analysis itself. If the user fails to select a bacterial identification method or set file paths, an error message is displayed, alerting the user to missing configuration, Fig. 4.

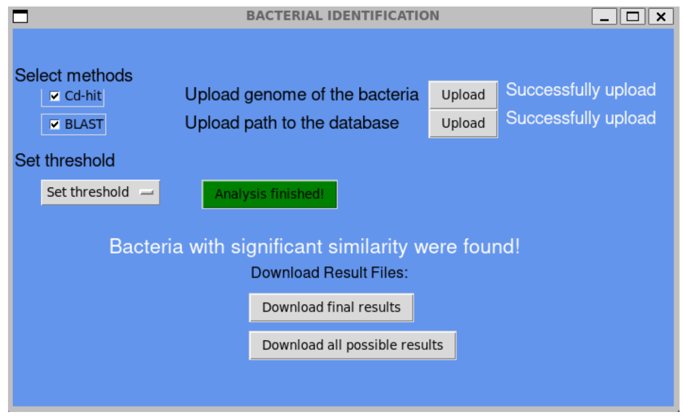



Fig. 4. Example of the message that is displayed if the user does not specify the path to the file with the unknown bacteria and the reference files.

When the user initiates the analysis and it completes successfully, an information message is displayed on the screen, providing the user with a preliminary result of the analysis. This message informs the user whether a match with any of the reference bacteria was found or if the unknown bacteria does not match any of them. Additionally, buttons are displayed that allow the user to download the resulting analysis files, which are crucial for further processing and interpretation of the results, see Fig. 4.

## IV. RESULTS AND DISCUSSION

### 1) *Experiment 1*

To test our app, we randomly selected a bacterium from our analyzed dataset, *Absiella dilichum* ET39, for identification purposes. After running the app, *Absiella dilichum* ET39 was identified as a bacterium from the species *Firmicutes*. Therefore, both Cd-hit and BLAST correctly identified this bacterium.

### 2) *Experiment 2*

In the second experiment, we uploaded *Bacteroides gallinaceum_25* as an unknown bacterium that was not in the analyzed dataset. *Bacteroides gallinaceum_121* was identified as the most similar bacterium by the Bacterial Identifier because significant matches with this bacterium were found by both BLAST and Cd-hit.

Cd-hit identified only this bacterium as similar, whereas BLAST identified another bacterium as similar. Specifically, BLAST identified *Bacteroides gallinaceum_121*, *Bacteroides gallinaceum_84*, *Bacteroides gallinaceum_An558*, *Bacteroides gallinaceum_An825*, and *Bacteroides gallinaceum_ET474* as similar bacteria to *Bacteroides gallinaceum_25*.

### 3) *Comparative Analysis of Computational Efficiency: Cd-hit vs. BLAST in Bacterial Identification*

During the development of the app, a temporal analysis was conducted. According to available sources [15][16], the asymptotic complexity of BLAST is $O(n^2)$, where $n$ represents the number of sequences. As the number of sequences increases, there is a substantial increase in the running time of BLAST. Conversely, the Cd-hit tool is expected to deliver faster results when working with larger databases compared to BLAST.

The testing of the Bacterial Identifier was performed on the Veterinary Research Institute database, comprising 452 bacterial genomes. For the linear model, the R2 value was 0.99626 for Cd-hit and 0.94138 for BLAST. However, when we calculated the R2 values for the exponential model, the R2 value for Cd-hit was only 0.76779, whereas for BLAST it was 0.94406. From these calculated R2 values, it is evident that the runtime of Cd-hit increases linearly with the database size, whereas for BLAST, it increases exponentially. This insight suggests that when working with a more extensive test dataset, Cd-hit would represent a faster alternative to BLAST.

## V. CONCLUSION

In conclusion, this study has presented the development and evaluation of the Bacterial Identifier application, designed to streamline bacterial identification processes. By integrating tools such as BLAST and Cd-hit and incorporating a user-friendly interface, the Bacterial Identifier offers an efficient solution for researchers and practitioners in the field of microbiology. Through testing on the Veterinary Research Institute database, consisting of 452 bacterial genomes, we have demonstrated the effectiveness of the Bacterial Identifier in accurately identifying bacterial strains. Our analysis revealed insights into the computational performance of BLAST and Cd-hit, with Cd-hit exhibiting linear scalability and potentially faster runtime when handling larger datasets compared to the exponential increase observed with BLAST.

The results underscore the importance of considering computational efficiency when selecting tools for bacterial identification, particularly in scenarios involving extensive datasets. The Bacterial Identifier, with its emphasis on speed, accuracy and user-friendly enviroment, presents a valuable resource for researchers and practitioners seeking to expedite the identification process while maintaining high-quality results.

## ACKNOWLEDGMENT

Computational resources were provided by the e-INFRA CZ project (ID:90140), supported by the Ministry of Education, Youth and Sports of the Czech Republic

This work has been supported by the National Centre for Biotechnology in Veterinary Medicine – NaCeBiVet.

## REFERENCES

[1] TERRENI, Marco; TACCANI, Marina; PREGNOLATO, Massimo. New antibiotics for multidrug-resistant bacterial strains: latest research developments and future perspectives. *Molecules*, 2021, 26.9: 2671.

[2] LI, Wenjun; RAOULT, Didier; FOURNIER, Pierre-Edouard. Bacterial strain typing in the genomic era. *FEMS microbiology reviews*, 2009, 33.5: 892-916.

[3] SCHWARZEROVA, Jana, et al. Detecting horizontal gene transfer among microbiota: an innovative pipeline for identifying co-shared genes within the mobilome through advanced comparative analysis. *Microbiology Spectrum*, 2024, 12.1: e01964-23.

[4] ALTSCHUL, Stephen F., et al. Basic local alignment search tool. *Journal of molecular biology*, 1990, 215.3: 403-410.

[5] Kent, W.J., 2002. BLAT—the BLAST-like alignment tool. *Genome research*, *12*(4), pp.656-664.

[6] Li, W. and Godzik, A., 2006. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics*, *22*(13), pp.1658-1659.

[7] SCHWARZEROVÁ, Jana, et al. Systems biology approach for analysis of mobile genetic elements in chicken gut microbiome. In: *2022 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. IEEE, 2022. p. 2865-2870.

[8] SCHWARZEROVÁ, Jana, et al. A minireview on the bioinformatics analysis of mobile gene elements in microbiome research. *Frontiers in Bacteriology*, 2023, 2: 1275910.

[9] MEDVECKY, Matej, et al. Whole genome sequencing and function prediction of 133 gut anaerobes isolated from chicken caecum in pure cultures. *BMC genomics*, 2018, 19.1: 1-15.

[10] Kim, J., Na, S.I., Kim, D. and Chun, J., 2021. UBCG2: Up-to-date bacterial core genes and pipeline for phylogenomic analysis. *Journal of Microbiology*, *59*(6), pp.609-615.

[11] LETUNIC, Ivica; BORK, Peer. Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation. *Bioinformatics*, 2007, 23.1: 127-128.

[12] SEEMANN, T.; BOOTH, T. Barrnap: basic rapid ribosomal RNA predictor. *GitHub repository*, 2018.

[13] Madden, T., 2013. The BLAST sequence analysis tool. *The NCBI handbook*, *2*(5), pp.425-436.

[14] LUNDH, Fredrik. An introduction to tkinter. *URL: www. pythonware. com/library/tkinter/introduction/index. htm*, 1999.

[15] CHAN, Alexander. An Analysis of Pairwise Sequence Alignment Algorithm Complexities: Needleman-Wunsch, Smith-Waterman, FASTA, BLAST and Gapped BLAST. URL: https://biochem218.stanford.edu/Projects%202004/Chan.pdf.

[16] *What is the time complexity of the given BLAST?* Online. In: Stack overflow. URL: https://stackoverflow.com/questions/59454358/what-is-the-time-complexity-of-the-given-blast.

# The Overlooked Fact: the Critical Role of Bioinformatics Pipeline in Microbiome Analysis

Petra Polakovičová
*Department of Biomedical Engineering*
*FEEC, Brno University of Technology*
Brno, Czech Republic
*Department of Experimental Medicine*
*Institute for Clinical and Experimental Medicine*
Prague, Czech Republic
xpolak37@vutbr.cz

***Abstract*—The research on the human microbiome has gained significant popularity over the past couple of years across both scientific and public communities. This area has revealed considerable knowledge, mainly thanks to next-generation sequencing. Even though technology possibilities evolve faster than ever, and there is an enormous amount of publicly available data, there are not yet precise procedures, methods, and bioinformatics tools to work accurately enough with microbiome data, and there are no standardized recommendations for bioinformatics pipelines to assess the most accurate results.**

**Here, we demonstrate the potential inaccuracies that can arise from various bioinformatics processing of amplicon 16S RNA sequencing data by performing a comparative analysis of diverse pipeline configurations using mock community samples and biological material originating from biopsy samples obtained during regular colonoscopy from patients with primary sclerosing cholangitis.**

***Index Terms*—microbiome, bioinformatics, comparative analysis, 16S rRNA**

## I. Introduction

Whether of the human body or other habitats, the microbiome study relies primarily on next-generation sequencing. The most precise method of exploring the microbiome is utilizing metagenomic shotgun sequencing. However, it is also an expensive and computationally demanding method. These reasons are still strong factors nowadays; therefore, the utilization of 16S rRNA gene sequencing is the most common at the moment. This gene is relatively short, covering approximately 1500 bp. It includes nine hypervariable regions (V1-V9) that exhibit significant sequence diversity across various species. These regions are flanked by conserved regions, which play a crucial role in facilitating PCR amplification with universal primers, the main step of the library preparation process.

However, many challenges arise in 16S rRNA amplicon sequencing which could be divided into two main categories. ”Wet lab” issues concern sample handling, DNA isolation methods, library preparation, and sequencing. The researchers are very much aware that virtually any step in this pipeline significantly influences the result, and several policies are put into practice to standardize this process.

In the bioinformatics community, different approaches are used to process FASTQ files and researchers often turn to freely available pipelines, typically relying on recommendations from relevant scientific papers or replicating processes from other studies. Today, such pipelines are mainly DADA2 [1], VSEARCH-UNOISE3 [2], and Deblur [3]. These tools extract true amplicon sequences (ASVs) from error-laden reads. Many studies on the microbiome often mention the tool they used without explaining its parameters and how it might bias the data. This approach has potential consequences, as various tools with different parameters can lead to different outcomes, and therefore, different conclusions can be deduced.

This study points out the possible biases caused by the bioinformatics processing itself and emphasizes the necessity of cautious presentation of results from similar studies without considering possible variations caused by data processing. A comparative analysis of the custom dataset of 16S rRNA amplicon sequencing reads is presented, which is processed with various settings across the three commonly used pipelines.

## II. Materials and Methods

### A. Dataset

This study is based on data from the research on the pathology of primary sclerosing cholangitis (PSC). The biological material was obtained during routine protocolary colonoscopies from patients who had undergone liver transplantation (LTX) for PSC (PSC group) or alcohol cirrhosis (control group) and regularly attended follow-up visits at Institute for Clinical and Experimental medicine (IKEM), Prague, CR. The study was performed according to the Declaration of Helsinki, including the changes accepted during the 59th WMA General Assembly, and approved by the Joint Ethics Committee of IKEM and Thomayer Hospital. All patients provided dedicated informed consent. In addition, mock community samples (ZymoBIOMICS Microbial Community DNA Standard[1]) were used as internal standards (Mock group).

[1] https://zymoresearch.eu/collections/zymobiomics-microbial-community-standards/products/zymobiomics-microbial-community-dna-standard

A dataset comprising 62 samples was created, including 22 internal standards, 20 cecum biopsy samples from the PSC group, and 20 cecum biopsy samples of controls. DNA was isolated using a Qiagen Power Fecal DNA kit, and the V3-V4 region of the 16S rRNA gene (primers: 341F, 806R) was sequenced on an Illumina Miseq with 2x250 bp read length. Table I provides basic characteristics of the dataset, information about the analysed groups, and sequencing read details. Sequencing data are available upon justified request.

TABLE I
BASIC CHARACTERISTICS OF THE ANALYZED DATASET, SUBDIVIDED ACCORDING TO THE SAMPLE GROUPS

| Characteristic | PSC<br>N = 20 | Control<br>N = 20 | Mock<br>N = 22 |
|---|---|---|---|
| Female sex [-] | 10 (50%) | 16 (80%) | - |
| Age<br>(min, med, max) [years] | 34<br>52.5<br>81 | 49<br>59<br>67 | - |
| Time from LTX<br>(min, med, max) [months] | 0<br>107.5<br>324 | 18<br>72<br>150 | - |
| Number of raw reads<br>(min, med, max) [-] | 34472<br>50466<br>145781 | 32127<br>55782<br>66939 | 30234<br>52670<br>88954 |
| Forward reads Q-score >= Q30<br>(min, med, max) [%] | 91.86<br>94.71<br>99.27 | 93.12<br>95.69<br>98.71 | 91.19<br>97.27<br>99.11 |
| Reverse reads Q-score >= Q30<br>(min, med, max) [%] | 79.02<br>87.29<br>91.35 | 40.81<br>90.68<br>94.47 | 63.72<br>81.68<br>93.70 |

## B. Data Processing

All samples underwent initial quality checks using FastQC v0.11.9 in combination with MultiQC v1.12 [4]. Next, primers were trimmed using CutAdapt v3.5 [5], with any untrimmed reads discarded.

Regarding bioinformatics processing, 28 different parameter settings across three pipelines for reads denoising, namely DADA2, VSEARCH-UNOISE3, and Deblur, were employed to process the data. The specific parameters are listed in Table II and Table III.

Utilizing DADA2, the paired reads were processed separately. After filtering and trimming with default parameters except for expected errors, the error model was trained for error rate estimation, and ASVs were inferred. The final step involved merging the paired reads, with default settings except for the maximum allowed mismatches in the overlap region. Chimeras were removed using the 'removeBimeraDenovo' function and taxonomic composition was obtained using IDTAXA [6] from the DECIPHER R package v2.30.0.

VSEARCH-UNOISE3 and Deblur operate with merged reads, unlike DADA2. Therefore, the initial common step for these two pipelines was merging the paired reads. This was conducted using the 'vsearch –fastq_mergepairs' command, with different values of minimum overlap region length and maximum allowed mismatches.

The subsequent steps differed for each pipeline. In VSEARCH, the reads underwent reorientation and filtering, with various settings for expected error and truncation length. These filtered reads were then denoised using UNOISE3, which is implemented in VSEARCH. Chimeras were discarded using 'uchime3_denovo'. Finally, taxonomic assignment was performed using SINTAX.

Following the merging of reads, as explained above, the QIIME2-Deblur plugin [6] was utilized, with the different sequence trim lengths, consistent with those used in VSEARCH. The taxonomic composition was obtained using IDTAXA, similar to the approach employed with DADA2. Each classifier used the SILVA v138.1 database [7] for taxonomic assignment.

TABLE II
DADA2'S PARAMETERS CONFIGURATIONS

| Setting | Expected Errors | Maximum Mismatches | Setting | Expected Errors | Maximum Mismatches |
|---|---|---|---|---|---|
| S1 | (2,2) | 0 | S8 | (2,5) | 1 |
| S2 | (2,2) | 1 | S9 | (1,3) | 0 |
| S3 | (2,3) | 0 | S10 | (1,3) | 1 |
| S4 | (2,3) | 1 | S11 | (1,4) | 0 |
| S5 | (2,4) | 0 | S12 | (1,4) | 1 |
| S6 | (2,4) | 1 | S13 | (1,5) | 0 |
| S7 | (2,5) | 0 | S14 | (1,5) | 1 |

## C. Comparative Analysis

The whole analysis was performed on the genus taxonomic level. All mock community samples processed with different pipelines were compared to the reference relative abundance of ZymoBIOMICS Microbial Community DNA Standard, which is illustrated in Fig. 1. Since the output of the SINTAX taxonomic classifier is also a probability matrix of classification, those with a probability of less than 0.7 were reassigned as 'unclassified', while others remained unchanged.

The downstream analysis utilized R v4.2.2. Precision and recall were computed for each pipeline setting to assess taxon detection accuracy. Bray-Curtis distance was calculated with vegan v2.6.4 [8] and Phyloseq v1.46.0 [9] for relative abundance comparison. Results were visualized using ggplot2 v3.4.4 [10]. Statistical differences in Bray-Curtis distances were evaluated with Kruskal-Wallis and post-hoc Dunn's test.

Furthermore, the biopsy samples were investigated focusing only on the most relevant pipeline settings. The top two performing settings for each pipeline were selected based on precision (>0.98), recall (>0.75), and median Bray-Curtis

TABLE III
VSEARCH'S AND DEBLUR'S PARAMETERS CONFIGURATIONS

| | VSEARCH-UNOISE3 + Deblur | | | VSEARCH -UNOISE3 | Deblur |
|---|---|---|---|---|---|
| Setting | Maximum Mismatches | Minimum Overlap | Truncation Length | Expected Errors | Mean Error Per Base |
| S1 | 2 | 10 | 250 | 1 | 0.005 |
| S2 | 2 | 12 | 350 | 1 | 0.005 |
| S3 | 3 | 12 | 350 | 1 | - |
| S4 | 3 | 12 | 350 | 2 | 0.005 |
| S5 | 2 | 12 | 350 | 2 | - |
| S6 | 2 | 12 | 400 | 2 | 0.005 |
| S7 | 2 | 12 | 400 | 1 | - |
| S8 | 2 | 10 | 400 | 1 | 0.005 |
| S9 | 2 | 10 | 400 | 2 | - |

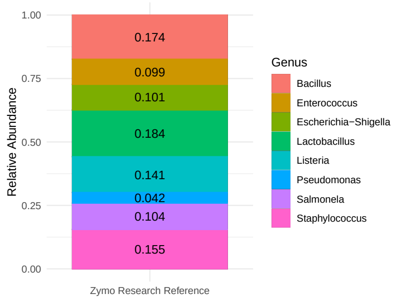



Fig. 1. The theoretical composition of ZymoBIOMICS Microbial Community Standard in terms of 16S rRNA gene abundance.

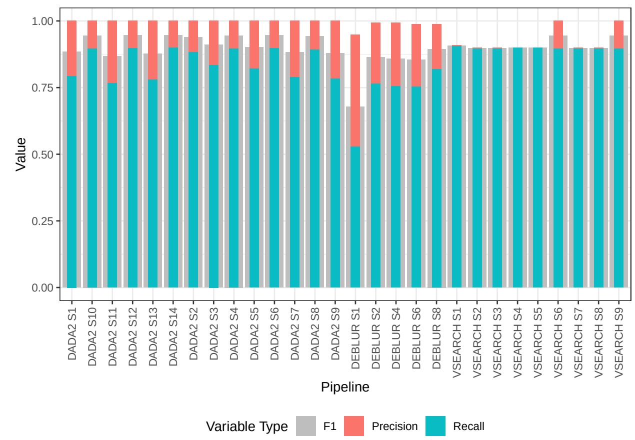



Fig. 2. Visualization of mean precision, recall, and F1 score of all of each pipeline setting, the average is calculated across all mock community samples.

distance (<0.4) criteria. Additionally, default settings for each pipeline were included in the comparison.

Statistical analysis began with filtration, removing the very low abundant (abundance <0.01% within a sample) and very low prevalent (prevalence <10% of all samples) genera. Next, to compare these two groups, the Shannon index was calculated and tested using the Mann-Whitney U test. Subsequently, the multivariate analysis was performed, testing differences in beta diversity (Aitchison distance) via PERMANOVA. Finally, univariate analysis was done utilizing Aldex2 v1.34 [11] using 128 Monte-Carlo samples.

To answer the question to what degree pipelines differ from each other, a simple comparison was made (i) by plotting PCA on clr-transformed data and (ii) by comparing outputs of taxonomic assignments based on different pipelines. To highlight the different performances of the taxonomic classifiers, the VSEARCH results were re-classified in the same way as the other two pipelines using IDTAXA.

The proposed analysis and a more comprehensive report are available at https://github.com/xpolak37/CA-Microbiome.

## III. Results and Discussion

Regarding the presence/absence of target genera in mock communities, the DADA2 pipeline has shown the best results, with five settings achieving F1 scores higher than 0.945 (S14, S12, S6, S4, S10). All settings of DADA2 demonstrated a precision of 1. The three settings of VSEARCH exhibited recall exceeding 0.90 (S1, S4, S5). However, only two of its settings (S6, S9) achieved a precision of 1, resulting in comparatively lower F1 scores compared to DADA2. Deblur showed the lowest performance, with most settings ranking last in recall and F1 score, although Deblur's precisions outperformed most VSEARCH settings. All metrics are visualized in Fig. 2.

Investigating the relative taxonomic composition also mirrors the results described above. However, as depicted in Fig. 3, DADA2 displays instability with different parameters, primarily with 'Maximum Mismatches' in the merging step. This observation is significant as the DADA2's default setting does not allow any mismatches; however, it does not perform well on the presented data. In contrast, VSEARCH demonstrates robustness in results with different parameters, and although it did not outperform the best result of DADA2, it showed similar behavior with different settings.

In Deblur, the lowest performance was achieved by setting S1, which can be attributed to over-trimming to length 250, thereby reducing the resolution of the denoising algorithm and the taxonomic classifier. Interestingly, over-trimming did not cause problems in VSEARCH S1.

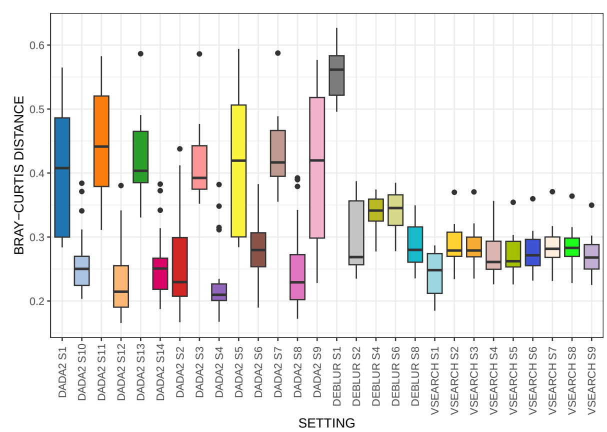



Fig. 3. Distribution of Bray-Curtis distances between individual mock community samples and reference abundances across different pipelines and parameters.

Overall, a significant difference between the Bray-Curtis distances of tested pipelines was observed ($P$<0.001), with 226 of 406 pairs of pipelines being significant in post-hoc testing at the significance level of 0.05. Out of 28, 13 pipelines met the performance criteria for being used for processing the biopsy samples, namely DADA2 S2, S3, S4, S6, S8, S10, S12, S14; VSEARCH S6, S9; and Deblur S2, S4, S8. From each pipeline, two configurations were picked according to their F1 score (DADA2 S12, S14; VSEARCH S6, S9; Deblur S2 S8) together with their default settings (noted as DADA2 S1, VSEARCH DEFAULT, and Deblur DEFAULT).

The comparative analysis of biopsy samples revealed that the biggest difference among the compared pipelines was introduced by the taxonomic classifier. As observed in the PCA analysis in Fig. 4, an apparent cluster of samples was present for the pipeline that utilized the SINTAX taxonomic classifier. Only 47.94% of the revealed genera were common across all 12 tested pipelines, and 56.70% were common among pipelines that employed the same classifier. The Venn diagram (Fig. 5) depicts how dissimilar the default settings of individual pipelines were, with 98 of 187 genera common to all four pipelines.

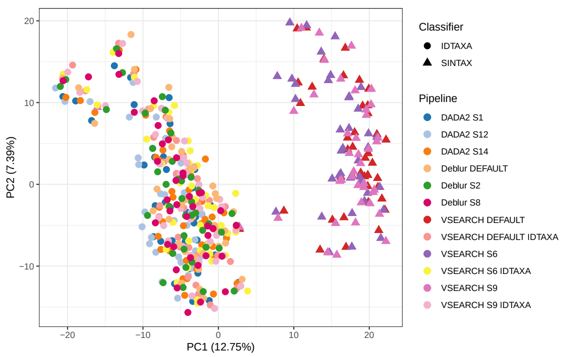



Fig. 4. Results of PCA analysis of biopsy samples processed by different pipelines.

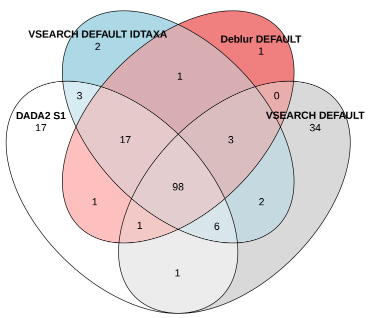



Fig. 5. A Venn diagram showing the shared and unique genera revealed by four pipelines with default settings.

The alpha and beta diversity measures in PSC and control samples assessed according to data obtained from different pipelines were similar. The Shannon index showed no significant difference between the PSC and control samples. However, PERMANOVA revealed significant differences between these two groups, except with the Deblur pipeline set to default settings, where the PERMANOVA was not significant ($P$=0.052).

However, this consistency does not apply to univariate analysis. After FDR correction, no genus appeared to be differentially abundant. This is a common phenomenon in microbiome analysis, where taxa distinguishing groups may not be identified despite the significance of beta diversity. Before FDR correction, nine genera were differentially abundant in the two groups, but none matched all the pipelines. The *Anaerostipes* genus was selected by all DADA2's settings and Deblur DEFAULT. The *Cutibacterium* was detected in all DADA2's settings and VSEARCH settings (both in combination with SINTAX and IDTAXA). However, others, like *Dialister* or *Blautia* were detected by some of the VSEARCH settings only.

## IV. Conclusion

This paper brings a new perspective on microbiome data processing and points out that more emphasis should be put on the proper selection and setup of the bioinformatics pipeline, as considerably different results were revealed on mock community samples. While the statistical significance of alpha and beta diversity on the genus level seemed relatively unaffected by pipeline selection for real microbiome data, other types of statistical tests, like univariate or functional analysis, may be notably influenced due to slight variations in taxonomic compositions between pipelines.

A possible solution would be to use multiple settings when processing the samples, compare the results with internal standards and describe the processing step in detail in the study. It is crucial that significant differences between groups remain consistent across various parameters to enhance the robustness of the reported findings.

## Acknowledgment


The data were acquired as a part of the project supported by the Ministry of Health of the Czech Republic grant nr. NU21J-06-00027.

Computational resources were provided by the e-INFRA CZ project (ID:90140), supported by the Ministry of Education, Youth and Sports of the Czech Republic.

Many thanks to doc. RNDr. Monika Cahová, Ph.D. for her valuable feedback and insights that led to the reported outcomes.


## References


[1] B. J. Callahan et al., "DADA2: High-resolution sample inference from Illumina amplicon data", Nature Methods, vol. 13, no. 7, pp. 581-583, 2016.

[2] T. Rognes et al., "VSEARCH: a versatile open source tool for metagenomics", PeerJ, vol. 4, 2016.

[3] A. Amir et al., "Deblur Rapidly Resolves Single-Nucleotide Community Sequence Patterns", MSystems, vol. 2, no. 2, pp. e00191-16, Apr. 2017.

[4] P. Ewels, M. Magnusson, S. Lundin, and M. Käller, "MultiQC: summarize analysis results for multiple tools and samples in a single report", Bioinformatics, vol. 32, no. 19, pp. 3047-3048, Oct. 2016.

[5] M. Martin, "Cutadapt removes adapter sequences from high-throughput sequencing reads", EMBnet.journal, vol. 17, no. 1, pp. 10-12, Aug. 2011.

[6] A. Murali et al., "Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2", Nature Biotechnology, vol. 37, no. 8, pp. 852-857, 2019.

[7] E. Pruesse et al., "SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB", Nucleic Acids Research, vol. 35, no. 21, pp. 7188-7196, Nov. 2007.

[8] J. Oksanen et al., "vegan: Community Ecology Package," R package version 2.6-4, 2022.

[9] P. J. McMurdie, S. Holmes, and M. Watson, "Phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data", PLoS ONE, vol. 8, no. 4, Apr. 2013.

[10] H. Wickham, Ggplot2: elegant graphics for data analysis, 2nd edition. Cham: Springer - Verlag, 2016.

[11] A. D. Fernandes et al., "ANOVA-Like Differential Expression (ALDEx) Analysis for Mixed Population RNA-Seq", PLoS ONE, vol. 8, no. 7, Jul. 2013.

# Implementation of a deep learning model for segmentation of multiple myeloma in CT data


Pavel Gálík
*Department of Biomedical Engineering*
*FEEC, Brno University of Technology*
Brno, Czech Republic
xgalik06@vutbr.cz

Michal Nohel
*Department of Biomedical Engineering*
*FEEC, Brno University of Technology*
Brno, Czech Republic
xnohel04@vutbr.cz



***Abstract*—This paper deals with the implementation of a deep learning model for spinal tumor segmentation of multiple myeloma patients in CT data. Deep learning is becoming an important part of developing computer-aided detection and diagnosis systems. In this study, a database of 25 patients who were imaged on spectral CT and for whom different parametric images (conventional CT, virtual monoenergetic images, calcium suppression images) were reconstructed, was used. Three convolutional neural network models based on the nnU-Net framework for lytic lesion segmentation were trained on the selected data. The results were evaluated on a test database and the trained models were compared.**



***Index Terms*—multiple myeloma, computed tomography, deep learning, nnU-Net, segmentation, monoenergetic image, calcium suppress image**


## I. Introduction

The human spine forms the axis of the upright body and consists of vertebrae that are connected by ligaments, joints, and intervertebral discs. The spine consists of 33 – 34 vertebrae, which divide into 7 *vertebrae cervicales* (C1 – C7), 12 *vertebrae thoracicae* (Th1 – Th12), 5 *vertebrae lumbales* (L1 – L5), 5 vertebrae that create *os sacrum* (S1 – S5), 4 – 5 *os coccygis*. Vertebrae are made of three parts, *corpus vertebrae* (the body), *arcus vertebrae* (vertebral arch), which protect the spinal cord, and is attached to the vertebral body from behind, and *processus vertebrae* (vertebrae process) are attached to the vertebral arch and serve for vertebral mobility. The structure of the vertebrae varies according to their position in the spine, it differs in the size of the vertebral corpus, size, and count of processes [1], [2].

Multiple myeloma (MM) stands as an incurable and biologically diverse disease rooted in the plasma cells. This disease is characterized by the unregulated proliferation of monoclonal plasma cells within the bone marrow, resulting in the excessive production of nonfunctional intact immunoglobulins or their constituent chains. The microenvironment, that supports the growth of myeloma cells and their survival consists of cellular and non-cellular environment compartments. The cellular compartment consists of hematopoietic and non-hematopoietic cells. The first includes myeloid cells, T and B lymphocytes, natural killer cells, and osteoclasts. The second one includes bone marrow stromal cells, bone marrow-derived mesenchymal stromal cells, fibroblasts, osteoblasts, adipocytes, endothelial cells, and blood vessels. Multiple myeloma ranks as the second most prevalent hematologic malignancy in the medical field. It is essentially a hematological malignancy characterized by clonal expansion of plasma cells [3], [4].

One of the typical features of multiple myeloma is its extensive damage to the skeletal system, which affects 80 – 90 % of patients during the course of the disease [5]. These myeloma-induced bone lesions are exclusively osteolytic and contribute to severe and incapacitating bone pain, pathologic fractures, hypercalcemia, spinal cord compression, and heightened mortality rates. The most prevalent symptom observed in individuals with multiple myeloma is bone pain, which is experienced by more than two-thirds of patients. A significant proportion of myeloma patients fall into the elderly demographic population, with a median age at diagnosis of 69 years and a median age at the time of death of 74 years [5]. Over the past three decades, there has been notable progress in treatment, resulting in an increased 5-year survival rate. Nevertheless, it's important to note that multiple myeloma remains an incurable disease [5], [6].

According to the European Myeloma Network and the European Society for Medical Oncology, the whole body low dose CT is recommended as the initial reference standard procedure for the diagnosis of lytic bone disease in patients with MM, as it is able to detect lesions with less than 5 % trabecular bone destruction [7].

Several authors are dealing with the segmentation of tumors: Yao et al. aimed to develop computer-aided detection and diagnosis (CAD) systems, in order to help detect any suspicious abnormalities in a large amount of data. Authors introduced a system for bone metastasis detection [8]. Based on the author's observation and knowledge about bone metastasis, they devised a set of 26 quantitative features in three categories: location, shape, and density. The authors developed a CAD system to detect bone lytic metastases in the thoracolumbar spine using routine CT images [8].

Cheng et al. developed a system for the identification of the metastasis of prostate cancer in whole-body scan images by using a deep convolutional neural network [9]. The developed system exhibited satisfactory performance for a small dataset of 205 cases, 100 of which were of bone metastasis. The sensitivity and precision for bone metastasis in the chest were

$0.82 \pm 0.08$ and $0.70 \pm 0.11$. The developed system has the potential to provide a prediagnostic report for physicians' final decisions [9].

J. Chmelik et al. dealt with the segmentation and classification of lytic and sclerotic metastatic lesions that are difficult to define by using spinal 3D CT images obtained from highly pathologically affected cases. The authors considered many approaches and uses of different CNN architecture, however, they would always have to overcome some unnecessary difficulties. The authors described a method for voxel-wise segmentation and classification of lytic and sclerotic bone tissues in whole-spine CT scans, based on the CNN architecture. This method can be applied to variously affected patient data acquired 29 with different CT acquisition parameters [10].

## II. Dataset

For this paper, data from 15 multiple myeloma patients and 10 patients without spinal pathology were used. The data was acquired with the help of the University Hospital Brno, Department of Radiology and Nuclear Medicine, using Philips Healthcare IQon spectral CT. The acquisition parameters included a peak tube voltage of 100 kV, tube current of 10 mA, a matrix size of $512 \times 512$, slice thickness of 0.9 mm, sharp reconstruction kernel, and hybrid iterative reconstruction technique (iDose4, set to level 4). The data was obtained with the Ethics Committee´s approval under application registration number NU23J-08-00027, and all patients gave informed consent. Two separate readers, one of whom was board certified, reviewed the data utilizing a specialized workstation (Intellispace Portal version 12.1; Philips Healthcare). The diagnosis of Multiple Myeloma (MM) was confirmed through elevated levels of monoclonal immunoglobulin in the blood and an augmented plasma cell count in the bone marrow. Initial disease staging was conducted using spectral CT with a low-dose protocol, adhering to guidelines set forth by the International Myeloma Working Group (IMWG).

The final database consists of 25 patients: 10 without pathologies in the spine and 15 diagnosed with MM and lytic lesions in the spine. Since spectral CT was used, conventional, virtual monoenergetic images at 40 keV, and calcium-suppressed CT images with index 25 were acquired as shown in Fig. 2. For every CT scan, a mask of a segmented spine was available, which was created using the nnU-Net machine learning model presented in [11]. An example of whole-body CT data is shown in Fig. 1. Also, a manually created mask for lytic lesions was available, the mask was labeled on VMI 40 keV images using MITK software [12], as the contrast is best on these images. Images with lesion mask are shown in Fig. 2(b).

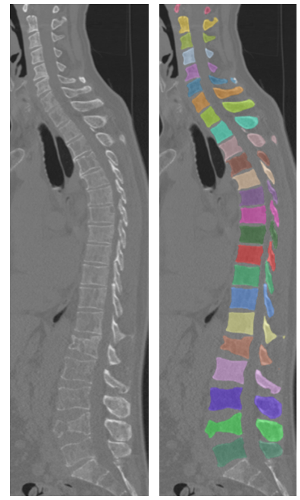
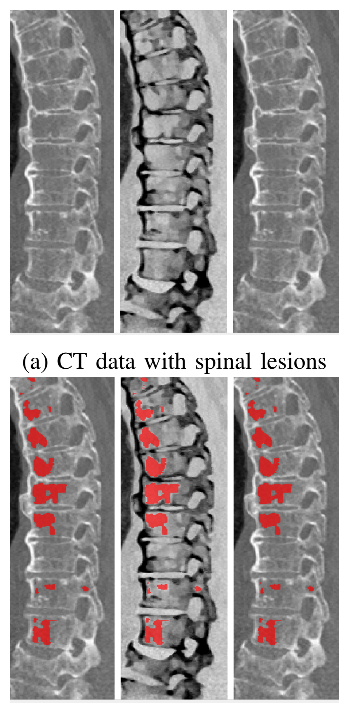


Fig. 1: On the left a whole body conventional CT scan of a patient with multiple myeloma, on the right an example of the available spinal segmentation mask.

As the CT scans are whole-body, the images are very large, thus preprocessing is needed. According to the available spinal segmentation mask, shown in Fig. 1, the images were cropped, and the size significantly decreased, resolving in much faster model training. The model was trained on the described dataset with professionally labeled MM lesions, the example of available training data is shown in Fig 2.



(a) CT data with spinal lesions

(b) CT data with lesions annotations

Fig. 2: Example of multiple myeloma CT data, from left conventional, calcium-suppressed with index 25 and monoenergetic 40 keV CT scans

## III. Model description

Machine learning and its quick development have the potential to automate the work of radiologists by automatically detecting and segmenting bone lesions from obtained images. The former segmentation methods often used image segmentation techniques, such as thresholding, region-growing, edge-based segmentation, active contour models, snakes, etc [13].

The nnU-Net [14] is a robust and self-adapting framework based on 2D and 3D vanilla U-Nets. This framework is a fast and effective deep-learning method for biomedical image segmentation. The nnU-Net automatically adapts its architectures to the given image geometry. For each task, the nnU-Net automatically runs a five-fold cross-validation for three different automatically configured U-Net models and the model with the highest mean foreground dice score is chosen for final submission. This paper uses the approach of this model structure, with the code available at https://github.com/MIC-DKFZ/nnUNet [14].

One of the main advantages of the nnU-Net is the ability to handle input images of different sizes or optimize the preprocessing step. As was mentioned earlier, the dataset used for this paper consists of 25 patients, however, ground truth labels were available only for 20 of them. On those, the model was trained.

If the structure of the data for nnU-Net is met, it automatically sets the optimal architecture and learning parameters, with a learning rate set at 0.01 and gradually reduced during training. The batch size was set at 2, patch size of $112 \times 192 \times 112$, a kernel size of $3 \times 3 \times 3$, and a total of 1000 learning epochs with the ReLu activation function were performed. For model validation, 5-fold cross-validation was utilized. For the model training, the Metacenter was used, which had 12 cores, 32GB of RAM, and a dedicated graphics card.

## IV. Metrics

For the evaluation of multiple myeloma segmentation, the Dice criterium was used, according to the Equation:

$$Dice(P,T) = 2\frac{|P \bigcap T|}{|P| + |T|} \tag{1}$$

Where $P$ represents the predicted pixel value, and $T$ represents the correct values (ground truth). The value of the Dice score is a dimensionless quantity [–] in the range of 0 and 1 (1 being the absolute match and 0 being no match).

For further evaluation, the false negative rate was used according to the Equation:

$$Sensitivity = \frac{TP}{TP + FN} \tag{2}$$

$$FNR = 1 - sensitivity \tag{3}$$

Where TP is true positive count, FN is false negative count.

## V. Results

Since true labels are available only for the 10 MM and 10 healthy patients, for the 5 remaining MM patients, the evaluation was done only by visual checkup. The result of a chosen segmentation is shown in Fig 3.

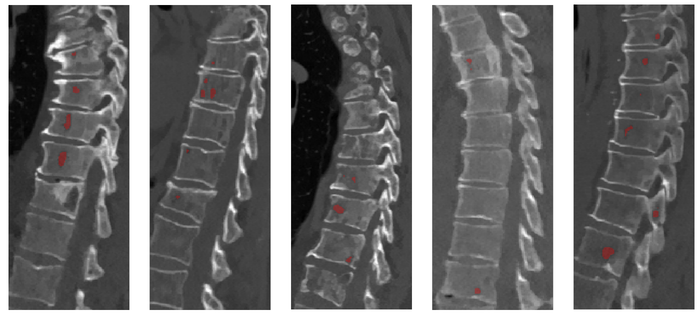


Fig. 3: Results of lesions segmentation on VMI dataset

As can be seen, detected lesions are located correctly, however, not many of the MM lesions were detected by the trained model.

The dice was computed for a validation set, and the results are shown in the following TABLE I. Model 1 represents a model trained on conventional CT data, model 2 represents a model trained on calcium-suppressed data, and model 3 represents a model trained on virtual monoenergetic images.

TABLE I: Dice coefficients for different models

| Patient | Model 1 | Model 2 | Model 3 |
|---|---|---|---|
| Myeloma_001 | – | 0.272 | – |
| Myeloma_004 | 0.272 | – | 0.268 |
| Myeloma_006 | – | 0.333 | – |
| Myeloma_007 | 0.301 | – | 0.215 |

Also, false negative rates were computed for the validation set, and the results are shown in TABLEII. Models represent the same datasets as in the TABLE I.

TABLE II: False negative rate for different models

| Patient | Model 1 | Model 2 | Model 3 |
|---|---|---|---|
| Not_Myeloma_001 | – | 0 | – |
| Not_Myeloma_004 | 0 | – | 0 |
| Not_Myeloma_009 | 0 | – | 0 |
| Myeloma_001 | – | 0.49 | – |
| Myeloma_002 | – | 0.84 | – |
| Myeloma_004 | 0.83 | – | 0.84 |
| Myeloma_006 | – | 0.59 | – |
| Myeloma_007 | 0.81 | – | 0.88 |

According to the results shown in the tables, the models mostly under-segmented and did not detect most of the lesions. This could be given to the small training dataset, which contains mainly patients with low MM occurrence. An example of segmented lesions of different models is shown in Fig 4.

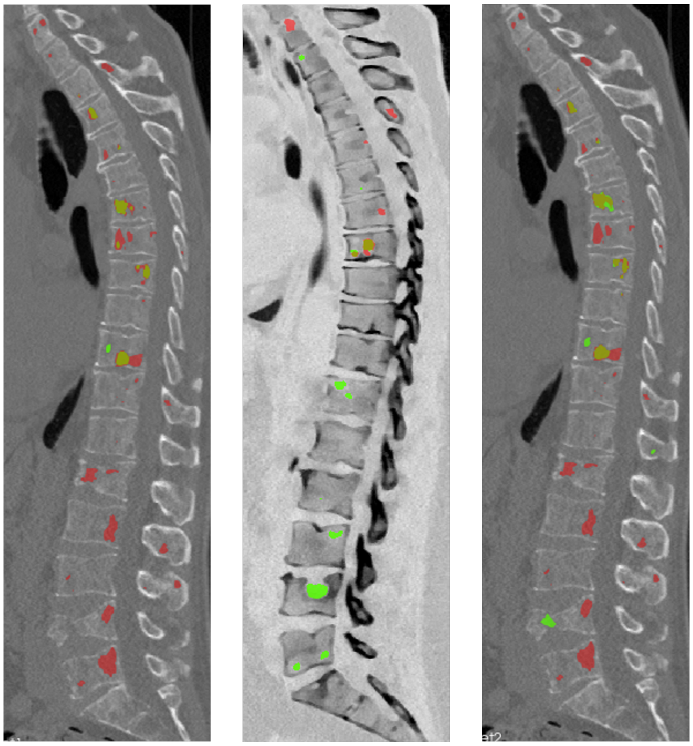


Fig. 4: CT scans with labeled lesions (green color) and with predicted lesion (mix of red and green color). From left conventional, calcium-suppressed with index 25, monoenergetic 40 keV CT scans

Fig 4 shows, that the models mostly under-segment or wrongly detect the MM lesions. A bigger training dataset would ensure better prediction results, as well as training a model on different training datasets (e.g. combined conventional with calcium-suppressed images) could improve the model performance. Also, training only on patients with MM could improve the results. Despite these shortcomings, improved models could very well serve as a checkup tool for skilled professionals.

## VI. Conclusion

This paper aims to implement and evaluate the deep learning segmentation model which would be a good help, with diagnosing multiple myeloma, and simplify radiologists' and doctors' work.

Three different nnU-Net models were trained on datasets consisting of 20 patients and their conventional, calcium-suppressed CT images with index 25, and virtual monoenergetic images at 40 keV CT scans. For evaluation, the dice criterium was used with the values above 0.21. Although a relatively small training data set was available, the results shown are quite promising and provide a good basis for further research and development.

## Acknowledgment

The paper and the research were supported by Philips Healthcare company and the University Hospital Brno, Department of Radiology and Nuclear Medicine. Computational resources were provided by the Ministry of Education, Youth, and Sports of the Czech Republic under the Projects CESNET (Project No. LM2015042) and CERIT-Scientific Cloud (Project No. LM2015085) provided within the program Projects of Large Research, Development, and Innovations Infrastructures.

## References

[1] FIALA, Pavel; VALENTA, Jiří and EBERLOVÁ, Lada, 2015. Stručná anatomie člověka. Online. Praha: Univerzita Karlova v Praze, nakladatelství Karolinum.

[2] ČIHÁK, Radomír, 2016. Anatomie. Online. Třetí, upravené a doplněné vydání. Praha: Grada.

[3] BRIGLE, Kevin and ROGERS, Barbara, 2017. Pathobiology and Diagnosis of Multiple Myeloma. Seminars in Oncology Nursing. vol. 33, no. 3, pp. 225-236. Available at: https://doi.org/10.1016/j.soncn.2017.05.012.

[4] SIMEONE, F. Joseph; HARVEY, Joel P.; YEE, Andrew J.; O'DONNELL, Elizabeth K.; RAJE, Noopur S. et al., 2019. Value of low-dose whole-body CT in the management of patients with multiple myeloma and precursor states. Online. Skeletal Radiology. vol. 48, no. 5, pp. 773-779. Available at: https://doi.org/10.1007/s00256-018-3066-6.

[5] SILBERMANN, Rebecca and ROODMAN, G. David, 2013. Myeloma bone disease: Pathophysiology and management. Online. Journal of Bone Oncology. vol. 2, no. 2, pp. 59-69. Available at: https://doi.org/10.1016/j.jbo.2013.04.001.

[6] HILLENGASS, Jens; USMANI, Saad; RAJKUMAR, S Vincent; DURIE, Brian G M; MATEOS, María-Victoria et al., 2019. International myeloma working group consensus recommendations on imaging in monoclonal plasma cell disorders. Online. The Lancet Oncology. vol. 20, issue 6, pp. e302-e312. Available at: https://doi.org/10.1016/S1470-2045(19)30309-2.

[7] ORMOND FILHO, Alípio G.; CARNEIRO, Bruno C.; PASTORE, Daniel; SILVA, Igor P.; YAMASHITA, Sâmia R. et al., 2019. Whole-Body Imaging of Multiple Myeloma: Diagnostic Criteria. Online. RadioGraphics. vol. 39, no. 4, pp. 1077-1097. Available at: https://doi.org/10.1148/rg.2019180096.

[8] YAO, Jianhua; O'CONNOR, Stacy and SUMMERS, Ronald, 2007. Computer Aided Detection of Lytic Bone Metastases in the Spine using Routine CT Images. Online. 2007 4th IEEE International Symposium on Biomedical Imaging: From Nano to Macro. pp. 512-515. Available at: https://doi.org/10.1109/ISBI.2007.356901.

[9] CHENG, Da-Chuan; LIU, Chia-Chuan; HSIEH, Te-Chun; YEN, Kuo-Yang and KAO, Chia-Hung, 2021. Bone Metastasis Detection in the Chest and Pelvis from a Whole-Body Bone Scan Using Deep Learning and a Small Dataset. Online. Electronics. vol. 10, no. 10. Available at: https://doi.org/10.3390/electronics10101201.

[10] CHMELIK, Jiri; JAKUBICEK, Roman; WALEK, Petr; JAN, Jiri; OUREDNICEK, Petr et al., 2018. Deep convolutional neural network-based segmentation and classification of difficult to define metastatic spinal lesions in 3D CT data. Online. Medical Image Analysis. vol. 49, pp. 76-88. Available at: https://doi.org/10.1016/j.media.2018.07.008.

[11] M. Nohel, R. Jakubicek, L. Blazkova, V. Valek, M. Dostal, P. Ourednicek, and J. Chmelik, "Comparison of Spine Segmentation Algorithms on Clinical Data from Spectral CT of Patients with Multiple Myeloma", in *MEDICON-23 and CMBEBIH-23*, 2024, pp. 309-317.

[12] I. Wolf, R. L. Galloway, Jr., M. Vetter, I. Wegner, M. Nolden, T. Bottger, M. Hastenteufel, M. Schobinger, T. Kunert, and H. -P. Meinzer, "The medical imaging interaction toolkit (MITK): a toolkit facilitating the creation of interactive software by extending VTK and ITK", p. 16-.

[13] RICH, Joseph M.; BHARDWAJ, Lokesh N.; SHAH, Aman; GANGAL, Krish; RAPAKA, Mohitha S. et al., 2023. Deep learning image segmentation approaches for malignant bone lesions: a systematic review and meta-analysis. Online. Frontiers in Radiology. 2023-8-8, vol. 3. Available at: https://doi.org/10.3389/fradi.2023.1241651.

[14] ISENSEE, Fabian, Paul F. JAEGER, Simon A. A. KOHL, Jens PETERSEN a Klaus H. MAIER-HEIN. NnUNet: a self-configuring method for deep learning-based biomedical image segmentation. Nature Methods. 2021, 18(2), 203-211. ISSN 1548-7091.

# Training zones estimation

1st Tobiáš Goldschmidt
*Department of Biomedical Engineering*
*FEEC, Brno University of Technology*
Brno Czech Republic
220487@vutbr.cz

2nd Oto Janoušek
*Department of Biomedical Engineering)*
*FEEC, Brno University of Technology*
Brno Czech Republic
janouseko@vut.cz

***Abstract*—Training zones, or exercise intensity zones, are a vital part of modern sport. Assessing these zones is not only costly but also needs to be done quite frequently. This paper proposes a new way of assessing training zones to overcome these challenges - based on DFA alpha1. It shows that laboratory assessment of training zones by DFA alpha1 is quite accurate. However, its usage to monitor exercise intensity in daily use is not as accurate, especially during low and medium intensity exercises.**

***Index Terms*—detrended fluctation analysis, heart rate variability, training zones**

## I. Introduction

In modern-day sport, every athlete wants to train as efficiently as possible. To achieve the best efficiency, training zones are used. A training zone tells the athlete how intense their workout is based on some easily measurable unit. For example, a distance runner who wants to maximize the effect of endurance training knows that training should be held in Zone 2, which, for them, means a heart rate range of 132-147 bpm (beats per minute). Usually, these zones are determined from a graded exercise test in a laboratory, during which exercise intensity is controlled, and metrics like lactate levels in blood or respiratory gas exchange parameters are measured. Based on physiological thresholds in lactate levels in blood or gas exchange parameters, zones are then calculated. However, equipment for this type of testing is quite expensive (respiratory gas exchange analyzers can cost thousands of euros) or invasive (for blood lactate level analysis, a blood sample needs to be taken), therefore not very attractive for the common user. Also, the results of these tests are translated into parameters that can be easily measured outside the laboratory, like heart rate or workload. However, these parameters have no physiological thresholds, so if any physiological change occurs in an athlete's body (e.g., the right training leads to a lower heart rate during the zone which maximizes endurance training), they have no way of adjusting training without another laboratory test. This paper proposes a way to overcome these two obstacles in training zone estimation by creating algorithm to show exercise intensity zone in real time based on DFA (detrended fluctuation analysis) alpha1 coefficient of HRV (heart rate variability). Based on papers [1], [2], physiological thresholds that occur in lactate levels in blood or gas exchange parameters also occur in DFA alpha1. For measuring DFA alpha1, only a heart rate belt is needed (costs around 75 euros). DFA alpha1 can also be calculated in pseudo-real-time (with a delay of approximately 25-45 seconds), so based on it, athletes can adjust the intensity of training during training and do not need to use heart rate or workload.

## II. Trainig zones estimation

Training zones are usually estimated from a graded exercise test in the laboratory. During this test, exercise intensity is controlled (usually as speed and slope of the treadmill or load on the ergometer) and increases until the tested athlete can no longer withstand it. This test covers the whole range of possible exercise intensities that the human body can experience. By monitoring lactate levels in blood or respiratory gas exchange during the test, energy metabolism is observed. To break down nutrients into energy, the human body needs oxygen and produces carbon dioxide. When there is not enough oxygen, the human body still generates energy from nutrients but less efficiently and creates lactate as a byproduct of this process. In human energy metabolism, there are two thresholds - AT (aerobic threshold) and AnT (anaerobic threshold) [3], [4]. The aerobic threshold represents the state at which metabolism becomes mixed, meaning that the body does not get enough oxygen and lactate starts to appear in the blood [3], [4]. The anaerobic threshold represents the state at which lactate in the blood cannot be utilized, and therefore the concentration of lactate in the blood starts to quickly rise [3], [4]. From these two thresholds, exercise intensity zones can be derived. The most intuitive approach is to split exercise intensity into three zones (because of two thresholds), but this division is rarely used. The most commonly used division is into 5 zones - compensation, long-time endurance, short-time endurance, development, and maximal intensity. Each specialist derives zones from AT and AnT in a slightly different way, but the most commonly used way is to have AT as the border between short and long-term endurance zones and AnT as the border between development and maximal intensity zones.

## III. Detrended fluctation analysis

DFA (Detrended Fluctuation Analysis) is a nonlinear method of HRV (heart rate variability) analysis. It was first used by [5] in 1995 and has since become a popular method with many applications. It is based on the fact that fluctuations in HRV are caused by the inner workings of the heart as well as by external uncorrelated stimuli. These two types of fluctuations should have different types of correlation properties. Fluctuations caused by external stimuli can be seen as trends in the data that show over different time scales. In the end, DFA quantifies the presence or absence of these fractal correlations. From DFA, two parameters can be obtained - alpha1 and alpha2. Alpha1 denotes short-term fluctuations, and alpha2 denotes long-term fluctuations.

### A. *Computation of detrended fluctation analysis*

The original computation of DFA by [5] can be split into 5 steps, as seen below. This approach was also used in the computation of DFA in the proposed algorithm.

1) Firstly, RR intervals (time between two R waves in ECG) sereis of length N is integrated (see equation below where $RR(i)$ is the current RR interval, and $RR_{avg}$ is the average RR interval) .

$$y(k) = \sum_{i=1}^{k}(RR(i) - RR_{avg}) \tag{1}$$

2) Next step is to divide signal into non-overlaping boxes of equal length $n$ (see Figure 1).
3) By fitting least squares line is fitted to each box from previous step and therefore, local tredn in each box is calculated (see in Figure 1). The local trend is denoted as $y_n(k)$.

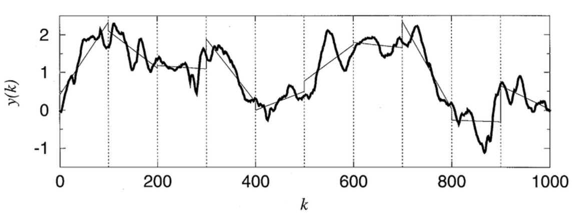



Fig. 1. Example of division of RR intervals into boxes of equal length and local trends in these boxes.

4) Local trend is then substracted and root-mean-square is calculated through all samples. The result is denoted as $F(n)$. The calculation can be seen in Equation 2.

$$F(n) = \sqrt{\frac{1}{N}\sum_{k=1}^{N}[y(k) - y_n(k)]^2} \tag{2}$$

5) Last step is to repeat steps 2-4 over different windoe sizes. From this $F(n)$ is obtained sa function of of teh size of the window $n$. $F(n)$ and $n$ are then plotted in log-log graph and linear relation of this graph is calculated. Slope is then denoted as alpha coefficient. For window sizes 4-16 is this coeffcinet called alpha1 and for window sizes 16-64 is called alpha2.

### B. *Detrended flucatation analysis and exercise intensity*

In last years there has been incresing number of papers published on topic DFA alpha1 an exercise intensity, especially its relation to AT and AnT. This activity culminated in two papers [1], [2] claiming that there are thresholds in DFA alpha1 highly correlating with AT and AnT. These thresholds are DF alpha1 = 0.75 for AT, and DFA alpha1 = 0.5 for AnT.

## IV. Proposed Algorithm

The algorithm for live display of training zones based on DFA alpha1 consists of 4 parts - connecting to the heart rate belt (namely Polar H10 by Polar Electro Oy, Kempele, Finland), collecting data from the heart rate belt (see Process A in Figure 2), calculating exercise intensity (see Process B in Figure 2), and user interface (see Process C in Figure 2). More detailed descriptions of each part (except for the user interface - as this paper is based on the first author's diploma thesis, the user interface will be implemented in the final stages of diploma thesis creation) can be found in further chapters. As shown in Figure 2, three of these parts are done as asynchronous processes. This is because the timely computation of DFA alpha1 may take longer than the time needed for new sample acquisition. The proposed algorithm is written in the Python language.

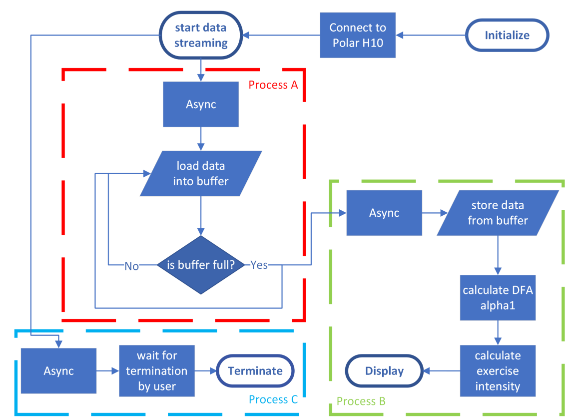



Fig. 2. Flowchart of implemented algorithm with marked three different asynchronous processes

### A. *Connecting to heart rate belt*

From variety of available hear rate belts Polar H10 by Polar Electro Oy, Kempele, Finland has been selected. It also shows excelent agreement with ECG based measuring of RR intervals [6]. It communicates with other devices via BLE (Bluetooth Low Energy). To connecting to it, algorithm based on [7] and [8] has been used. It is mainly based on Bleak library for Python.

### B. *Collecting data from heart rate belt*

After the connection with Polar H10 is established and the data stream started, incoming data need to be collected. In this part, two buffers are created. The first buffer controls the rate of computing exercise intensity, and the second one controls the sample size over which DFA alpha1 and therefore exercise intensity will be computed.

The first buffer is established to control the computational load on the device. As mentioned before, DFA alpha1 calculation is a computationally difficult process and may take some time, therefore computing it with each incoming sample may be overkill for the device. The size of this buffer is set to 10 samples, meaning that DFA alpha1 will be computed with every 10th incoming sample of data. Based on this, DFA alpha1 and exercise intensity will be calculated every 3-5 seconds, depending on exercise intensity.

The second buffer is established because [5] states that the minimal number of samples over which DFA can be computed is 5 times bigger than the maximal window used in computation. Because in this case DFA alpha1 is calculated, the maximal window size is 16, and therefore the minimal number of samples is 80. However, the need to accumulate 80 samples before DFA alpha1 calculation causes a delay of 25-40 seconds, depending on exercise intensity. This delay would be even bigger if the number of samples was larger. During initial testing, a bigger sample size did not show significantly better results, so with the aim to use this algorithm in real-time applications, the size of this buffer was set to 80 samples.

### C. Exercise intensity computation

To compute exercise intensity, firstly DFA alpha1 needs to be calculated. To compute DFA alpha1, RR intervals are needed. RR intervals suffer from distortion, and these distortions can cause significant disturbances in DFA alpha1 computation [9]. To filter this distortion, a simple filter was implemented - affected RR intervals are replaced by the mean of their neighboring unaffected RR intervals. After filtering, DFA alpha1 is computed. This computation is based on the method of DFA computation described by [5]. Further information about DFA computation can be found in Chapter III-A. From the DFA alpha1 value, the exercise intensity zone is found by thresholding. As mentioned in Chapter III-B, there are certain thresholds in DFA alpha1 corresponding to AT and AnT, while AT and AnT correspond to borders between intensity zones (see Chapter II). That means that proposed thresholds for the 5-zone model are DFA alpha1=1.0 (between compensation and long-term endurance), DFA alpha1=0.75 (between long and short-term endurance, also AT), DFA alpha1=0.6 (between short-term endurance and development), and DFA alpha1=0.5 (between development and maximal intensity, also AnT).

## V. Results

For all plots below, a color coding of exercise zones is used as follows - blue represents compensation, green represents long-term endurance, yellow represents short-term endurance, red represents development, and magenta represents maximal intensity. The plots also utilize thresholds T1, T2, T3, and T4. T1 denotes the threshold between compensation and long-term endurance, T2 between long and short-term endurance, T3 between short-term endurance and development, and T4 between development and maximal intensity.

Testing of the proposed algorithm has been conducted in two different runs by the same person (23 years old, male, active sportsman). Exercise intensity zones are translated to heart rate and were measured by a laboratory test in a specialized laboratory. Testing of proposed algorithm for laboratory usage was done by same person. As reference values were used values of thresholds found from gas exchange parameters.

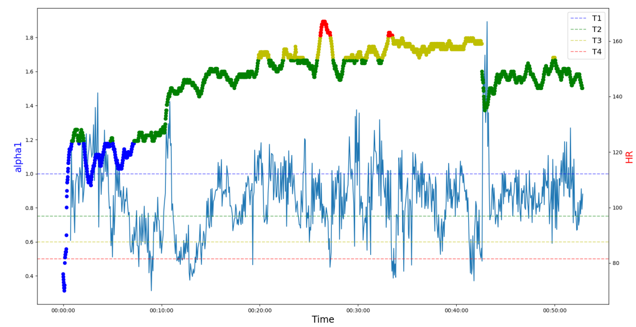



Fig. 3. Graph of HR and DFA alpha1 over time during tempo run. Begining and end of run are warm-up and cool down phases, tempo phase (30 minutes of neraly same intensity) is between 10th and 40th minute.

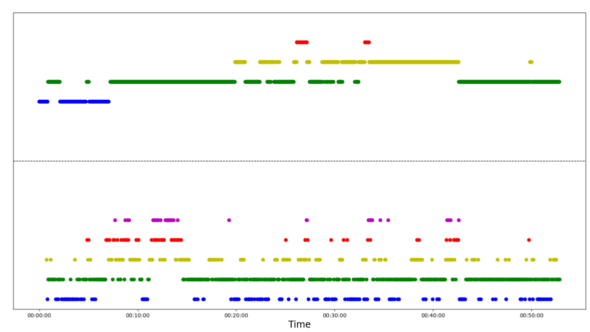



Fig. 4. Graph of exercise intensity zones absed on DFA alpha1 and HR over time for tempo run. Upper par of plot are zones based on HR, bottom on HR. Tempo part of run is between 10th and 40th minute.

As observed in Fig. 3, during the 30-minute tempo run (when runner tries to keep same pace and percieved efford for certain period of time), DFA alpha1 does not follow the same trend as heart rate (HR). While HR slightly elevates during the run, DFA alpha1 remains nearly constant, oscillating around approximately 0.8. Additionally, as depicted in Fig. 4, zones based on HR transition from long-term intensity to short-term intensity with occasional dips into the development zone, whereas zones based on DFA alpha1 are scattered across the scale from compensation to maximal intensity. However, it's notable that the most common zone based on both metrics is the same - long-term endurance.

In the intervals session (a run where the runner increases intensity for periods of time multiple times per run from a fixed distance or tempo with pauses in between), in the first 3 intervals, the DFA alpha1 value oscillates around 1.2

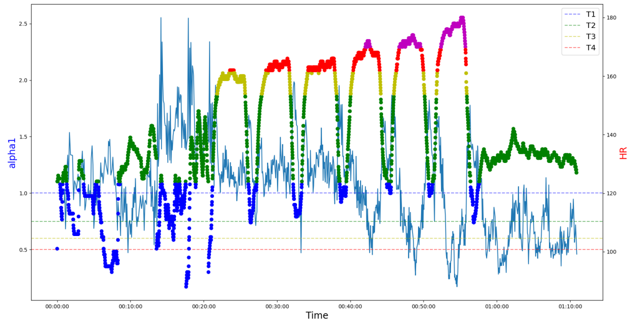



Fig. 5. Graph of HR and DFA alpha1 over time during intervals run. Begining and end of run are warm-up and cool down phases, intervals phase (6 times cca 4 minutes of increased intensity with pauses) is between 22th and 57th minute.

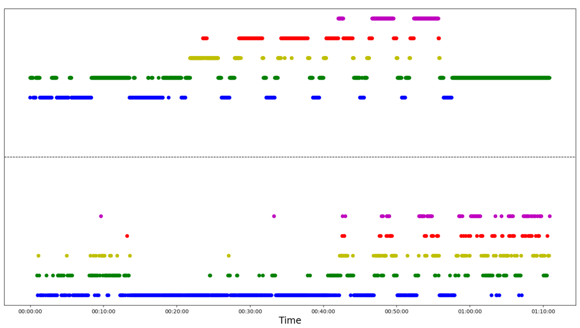



Fig. 6. Graph of exercise intensity zones absed on DFA alpha1 and HR over time for intervals. Upper par of plot are zones based on HR, bottom on HR. Intervals part of run is between 22th and 57th minute.

(in the compensation zone), while the heart rate (HR) value remains between the short-range endurance and development zones (see Fig. 5). Both metrics maintain a similar trend, with minimal changes observed. However, for the last three intervals, both metrics exhibit opposite trends - HR increasing and DFA alpha1 dropping. As shown in Fig. 6, for the first three intervals, the metrics do not match in their zone assumptions. However, in the last three intervals, when the exercise intensity reached maximal intensity, both metrics assume similar exercise intensity zones, although the intensity zones based on DFA alpha1 are much more fluctuating.

DFA alpha1 can also be used in laboratory assessment of exercise intensity zones. To find thresholds for each zone, the mean of the last 10 samples under that threshold was calculated. From there, the time of thresholds has been determined, and from the time of thresholds, the HR value at these thresholds has been found (see Fig. 7). The mean difference over all thresholds in HR was 1.5 bpm (beats per minute), and in time of test, it was 30 seconds. DFA alpha1 followed an opposite trend to the HR trend during exercise. T2 and T4 were more similar between metrics than other thresholds.

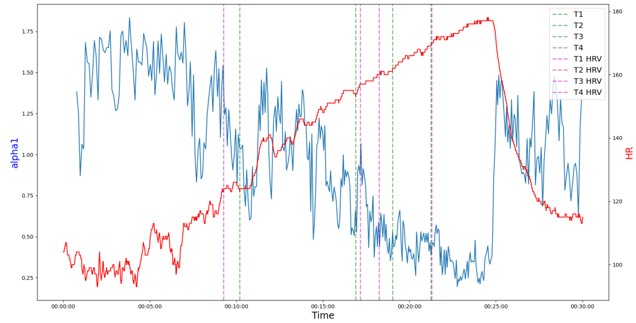



Fig. 7. Comparison of DFA alpha1 and HR in time during laboratory test

## VI. Conclusion

The proposed algorithm demonstrates good accuracy in laboratory usage, although it was not the primary aim of its creation. In outdoor training, it shows the best accuracy in workouts with increasing intensity and in workouts with higher intensity, preferably near maximal. Further testing on more subjects is needed for fine-tuning parameters as well as for validation. In conclusion, the proposed algorithm shows great potential in planned use but requires more work on app creation and fine-tuning. Even with these improvements, it may not be able to distinguish between low and medium exercise intensity.

## References


[1] B. Rogers, D. Giles, N. Draper, O. Hoos, and T. Gronwald, "A New Detection Method Defining the Aerobic Threshold for Endurance Exercise and Training Prescription Based on Fractal Correlation Properties of Heart Rate Variability", FRONTIERS IN PHYSIOLOGY, vol. 11, 2021.

[2] B. Rogers, D. Giles, N. Draper, L. Mourot, and T. Gronwald, "Detection of the Anaerobic Threshold in Endurance Sports: Validation of a New Method Using Correlation Properties of Heart Rate Variability: Validation of a New Method Using Correlation Properties of Heart Rate Variability", Journal of Functional Morphology and Kinesiology, vol. 6, no. 2, 2021.

[3] W. Kindermann, G. Simon, and J. Keul, "The significance of the aerobic-anaerobic transition for the determination of work load intensities during endurance training", European journal of applied physiology and occupational physiology, vol. 42, pp. 25-34, 1979.

[4] D. Keir, F. Fontana, T. Robertson, J. Murias, D. Paterson, J. Kowalchuk, and S. Pogliaghi, "Exercise Intensity Thresholds: Identifying the Boundaries of Sustainable Performance: Identifying the Boundaries of Sustainable Performance", Medicine & Science in Sports & Exercise, vol. 47, no. 9, 2015.

[5] C. -K. Peng, S. Havlin, H. E. Stanley, and A. L. Goldberger, "Quantification of scaling exponents and crossover phenomena in nonstationary heartbeat time series", Chaos: an interdisciplinary journal of nonlinear science, vol. 5, no. 1, pp. 82-87, 1995.

[6] M. Schaffarczyk, B. Rogers, R. Reer, and T. Gronwald, "Validity of the Polar H10 Sensor for Heart Rate Variability Analysis during Resting State and Incremental Exercise in Recreational Men and Women", SENSORS, vol. 22, no. 17, 2022.

[7] M. M. Span, "PolarBand2lsl", GitHub, 02023.

[8] W. van Ham, "Polar", Wilbert's website at SocSci, 02012.

[9] B. Rogers, D. Giles, N. Draper, L. Mourot, and T. Gronwald, "Influence of Artefact Correction and Recording Device Type on the Practical Application of a Non-Linear Heart Rate Variability Biomarker for Aerobic Threshold Determination", SENSORS, vol. 21, no. 3, 2021.

# Co-expression analysis of small RNAs and untranslated regions in Rhodospirillum rubrum

Kristyna Hermankova
*Department of Biomedical Engineering*
*FEEC, Brno University of Technology*
Brno, Czech Republic
ORCID: 0009-0002-6720-0499

Karel Sedlar
*Department of Biomedical Engineering*
*FEEC, Brno University of Technology*
Brno, Czech Republic
ORCID: 0000-0002-8269-402

***Abstract* — Bacterial small RNAs (sRNAs) are regulatory elements often mentioned in the context of stress response. So are polyhydroxyalkanoates (PHAs), a possible replacement for conventional plastics, are frequently produced by bacteria facing stress conditions. PHAs then serve as the source of the carbon as well as the defense mechanism against further stress. Here, we present sRNA analysis of PHA-producing bacterium *Rhodospirillum rubrum*, especially the co-expression analysis of inferred sRNA candidates from RNA-Seq data. Weighted correlation network analysis (WGCNA) was used to compute clusters of genes with similar expression profiles, resulting in 28 modules. Further, the functional annotation of genes in each module was applied and the overall functional profiles were mentioned using a 'guilt by association' approach. Additionally, functional enrichment of modules was obtained through over-representation analysis (ORA). These results provide an overview of regulatory elements and their predicted functions in *R. rubrum*, which further unveil its dynamic properties and can help with the assessment of its industrial application.**

***Keywords — RNA-Seq, PHA, sRNA, functional annotation, co-expression analysis***

## I. Introduction

New approaches are still in the process of substituting conventional plastics with those that can be produced from renewable sources. One such option is the replacement with polyhydroxyalkanoates (PHAs) originating from living organisms such as bacteria [1]. Nutrient-limiting stress conditions are often applied to induce the production of PHAs [2] in the form of small granules which then serve as a source of carbon and contribute to the organism's overall resistance against other stress factors [3].

The Next generation industrial biotechnology [4] approach aims for sustainable bacterial production of PHA bioplastics, so they are competitive with the cheap production of fossil fuel-based plastics. One way to reduce production costs is to find promising bacterial candidates that can efficiently produce PHA without the need for expensive conditions for batch production [4].

There are plenty of bacteria whose genome has been already sequenced and thus information on whether bacterium contain genes of PHA synthesis machinery is available. However, the genetic information is only static and isn't enough for the assessment of bacterial capabilities that are of interest for industrial applications. Thus, further approaches revealing more dynamic properties of bacteria are required.

Small RNAs (sRNAs) are mainly non-coding elements involved in regulatory processes as they are binding to their mRNA target and the fate of the mRNA molecule then depends on whether the bonded sRNA discards or enables the translational process. Their transcription is induced mostly under stress conditions as they provide fast response thanks to their co-transcriptional as well as post-transcriptional regulatory capability [5]. sRNAs are transcribed from various genomic regions including intragenic regions (IGR), but also 5' and 3' untranslated regions (UTRs) of genes that might regulate the gene of interest [6]. Also, when speaking of 3' UTR regions, the promotor sequence responsible for its transcription doesn't have to be present [7]. This makes the search for sRNAs even trickier. Predictions are then primarily based on sequencing technologies other than genome sequencing. RNA-Seq [8] has become the state-of-the-art technique for producing high-throughput expression data. Even though the price of high-throughput sequencing technologies has become more available, there are still attempts to get as much information as possible from a single experiment. Thus, the idea of utilization of RNA-Seq data to predict sRNAs became obvious as there is a chance to capture actively transcribed sRNA molecules.

The accumulation of PHA in bacterial cells is frequently associated with the response of bacteria to the stress conditions. This fact concludes that sRNAs may be involved in the PHA synthesis pathway as molecules regulating the expression profile of genes associated. As predicting sRNA *in silico* is a non-trivial task, their annotations are often absent in automated pipelines. This results in a gap in our understanding of bacterial regulatory processes and thus further analyses need to be performed to get insight into the underlying mechanisms.

Here, we present results from co-expression analysis of *Rhodospirillum rubrum* DSM 467$^T$, bacterium producing PHAs. This provides insight into non-coding regulatory processes of PHA-producing bacterium as well as functional annotations of candidate sRNAs with the 'guilt by association' approach. Results offer overview of the functional properties of bacterial non-coding elements. This can help in gaining further knowledge about the processes emerging under different conditions and improve the assessment of the bacterium's utility in biotechnological applications.

## II. Materials and Methods

### A. RNA-Seq data

Data were obtained under project accession number PRJNA742260 from the NCBI Sequence Read Archive (SRA). An overview of RNA-Seq samples used is summarized in Table I. Data contain gene expression of two *R. rubrum* strains – wild type (WT) and knock-out (KO), cultivated on two substrates – acetate and fructose, each in 3 time points – T1, T2, T3 (except KO on acetate containing only T1 and T2) and in 3 biological replicates (range in column **Sample**).

TABLE I. RNA-Seq samples

| Sample | Time | Substrate | Genotype |
|---|---|---|---|
| 1-3 | T1 | Acetate | WT |
| 4-6 | T2 | Acetate | WT |
| 7-9 | T3 | Acetate | WT |
| 10-12 | T1 | Acetate | KO |
| 13-15 | T2 | Acetate | KO |
| 16-18 | T1 | Fructose | KO |
| 19-21 | T2 | Fructose | KO |
| 22-24 | T3 | Fructose | KO |
| 25-27 | T1 | Fructose | WT |
| 28-30 | T2 | Fructose | WT |
| 31-33 | T3 | Fructose | WT |

### B. Prediction of sRNAs and UTRs

Bam files of mapped RNA-Seq reads to the reference genome of wild-type strain *R. rubrum* DSM 467$^T$ were used as input to the baerhunter tool [9] for the sRNA and UTR prediction. Normalized sequencing depths of genes using the *sizeFactors* function from the DESeq2 package [10] were used for the estimation of the *high_cut_off* threshold for the baerhunter prediction. The rest of the thresholds remained set to default. Subsequently, for the computation of the genomic features' count table, baerhunter along with *featureCounts*, R Subreads [11] package was used.

To obtain a count table appropriately transformed for the co-expression analysis, the *varianceStabilizingTransformation* (*vst*) function in the DESeq2 package was used.

### C. Co-expression analysis

Counts of genomic features (further referred to genomic elements already annotated along with predicted sRNAs and UTRs) in the count table transformed with the *vst* were further processed with the WGCNA (weighted gene co-expression network analysis) package [12]. For the *unsigned* network, the *soft thresholding power* of 12 was selected as it is the first value that satisfied the scale-free topology criterion set on the 0.8 value. The *power* value of 12 was further used when computing modules representing clusters of genomic features with similar expression profiles. The *minModuleSize* was set to a minimum of 30 genomic features per module and *mergeCutHeight* was set to 0.2 to cut the dendrogram for merging modules.

### D. Functional annotation and Over-representation analysis

Consequently, Clusters of Orthologous Groups (COGs) were assigned to coding sequences (CDSs) using eggNOG-mapper [13]. To decide whether are modules enriched with specific COG, Over-representation analysis (ORA) was performed. Using the clusterProfiler package [14] and function *enricher*, a statistically significant overrepresentation of a functional group was examined. A *pvalueCutoff* of 0.05 was used for significant enrichment, and the Benjamini-Hochberg procedure was used to correct for multiple testing.

Plots were generated using the ggplot2 package [15] and colour palettes are from the ggsci package [16].

## III. Results and Discussion

Based on pairwise correlations of genomic features' expression profiles, 28 modules were obtained. These modules then represent 28 groups of clustered genomic features with correlated expression profiles. The *networkType* of the generated co-expression network was set to an *unsigned* network. Expression profiles in modules can thus highly correlate even in a negative manner, as sRNAs can negatively regulate their mRNA targets.

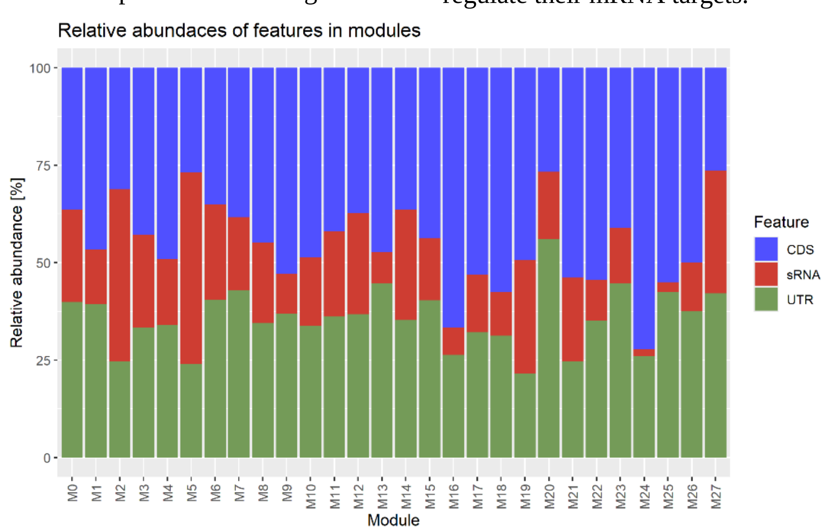



Fig. 1. The representation of genomic features in modules.

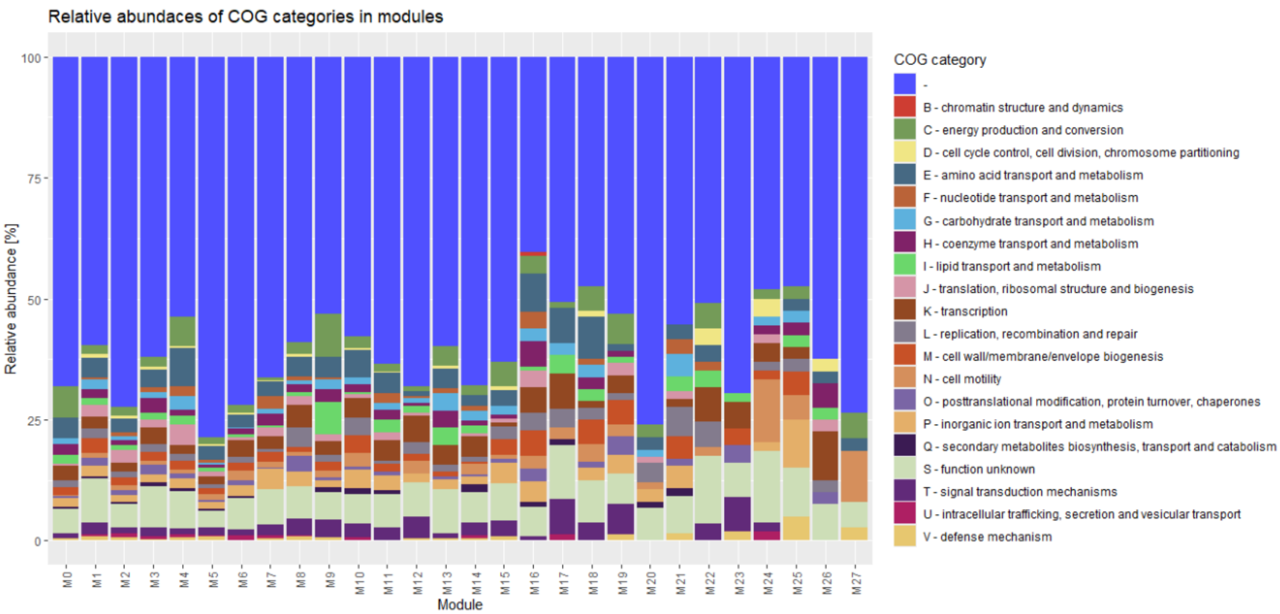



Fig. 2. The abundance of COG categories in each module.

Therefore, we assumed that there could be a negative correlation between the sRNA and its target, even though they have a similar biological function. The distribution of genomic features in each module is shown in Figure 1.

Predominant features in modules are mainly CDSs and their UTRs that are transcriptionally active enough to be detected by the baerhunter. Fluctuations of sRNA abundances throughout modules could suggest that modules with a higher percentage of sRNAs refer to functional pathways that are active under applied cultivation conditions. Higher percentages of sRNAs are in modules M2, M5, and M27 modules, suggesting higher regulatory activity. On the contrary, M16, M24 and M25 modules contain low numbers of sRNAs and thus could represent slightly active pathways.

The 'guilt by association' functional annotation approach resulted in functional annotation of modules based on the abundance of COGs assigned to CDSs in each module. The functional assignment to each module is then inferred from its COG distribution, see Figure 2.

The most abundant COG in each module is the 'dash' group which contains predicted sRNAs and UTRs, as the annotation of these features isn't possible with eggNOG-mapper, and further CDS that weren't assigned to any COG at all. The other abundant group is 'S' which refers to a cluster of similar proteins, but their function is unknown.

When comparing Figures 1. and 2., it is evident that modules with a high percentage of sRNAs have a small proportion of functionally annotated features. This is due to the lower percentage of CDSs that can be annotated. This principle also applies in reverse. Further, module M27 has only four COGs, except the 'S – unknown' and 'dash' groups, 'C - energy production and conversion', 'E - amino acid transport and metabolism', 'N - cell motility' and 'V - defense mechanisms' group. Such a combination could then suggest that the pathway encoded in this module is responsible for the motility as a response to induced stress.

More accurate suggestions can be concluded from ORA results. Enrichment with COGs resulted only in few modules with a statistically significant abundance of CDSs with specific functions. The M4 module contains the most COGs with significant frequency, see Figure 3. More precisely, the groups are - 'E – amino acid transport and metabolism', 'C – energy production and conversion', 'J – translation, ribosomal structure and biogenesis', 'G – carbohydrate transport and metabolism' and 'F – nucleotide transport and metabolism'. Even though these terms are very general, the main pathway ensuring the basic operation of the bacterium could be suggested.

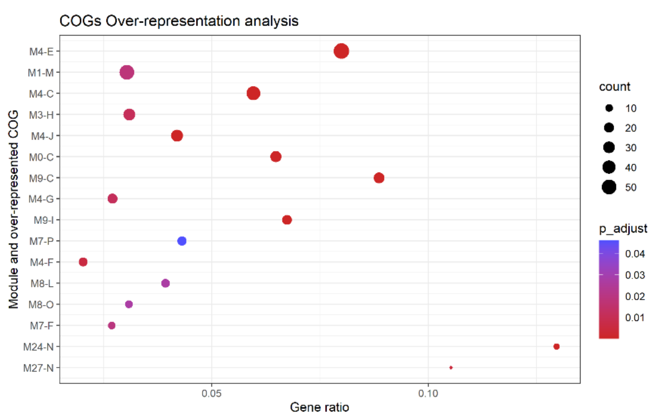



Fig. 3. Over-represented COGs in modules.

As we are talking about PHA-producing bacterium, we would expect genes and their regulators involved in the PHA synthesis pathway to cluster together. When considering all Figures, the M9 module seems as a suitable candidate for the PHA synthesis pathway as the ORA shows a significant overrepresentation of 'C' and 'I' COG, both playing crucial roles in PHA biosynthesis. However, the proportion of sRNAs isn't as high as we would expect, together with UTRs, also representing sites for the transcription of regulatory elements, constitute almost half of clustered genomic regions in specific module.

## IV. Conclusions

In this paper, we performed co-expression analysis of CDSs together with predicted non-coding elements of R. rubrum and provided results from following functional annotation with a ‘guilt by association’ approach.

After the pre-processing of transcriptomic data involving inference of non-coding elements, 28 modules referring to clusters of genomic features with similar expression profiles were found. Further, with the functional annotation of CDS, we were able to assign multiple categories to each module and thus provide initial insight into the functional annotation of whole modules that could stand for functional pathways. Subsequent ORA provided a more precise overview of whether modules are enriched with specific functional terms and preliminary pathway suggestions were made.

This work offers further perspectives, such as adding Gene Ontology or KEGG pathways enrichment analysis, which can lead to more precise functional annotations with more specific terms included. Additionally, this work can improve the functional annotations of non-coding elements or at least make a preliminary proposal of their targets. Overall, co-expression analysis of non-coding regulatory elements, combined with downstream analyses, can contribute to the assessment of bacterial utilization in industrial applications by providing advanced knowledge of the dynamic properties of the bacteria studied.

## Acknowledgement

This work was supported by a grant project from the Czech Science Foundation [GA21-15958L].

## References

[1] Steinbüchel, A., & Hein, S. (2001). Biochemical and molecular basis of microbial synthesis of polyhydroxyalkanoates in microorganisms. *Advances in Biochemical Engineering/Biotechnology*, *71*, 81–123. https://doi.org/10.1007/3-540-40021-4_3/COVER

[2] Kadouri, D., Jurkevitch, E., Okon, Y., & Castro-Sowinski, S. (2005). Ecological and Agricultural Significance of Bacterial Polyhydroxyalkanoates. *Critical Reviews in Microbiology*, *31*(2), 55–67. https://doi.org/10.1080/10408410590899228

[3] Ruiz, J. A., López, N. I., Fernández, R. O., & Méndez, B. S. (2001). Polyhydroxyalkanoate degradation is associated with nucleotide accumulation and enhances stress resistance and survival of Pseudomonas oleovorans in natural water microcosms. *Applied and Environmental Microbiology*, *67*(1), 225–230. https://doi.org/10.1128/AEM.67.1.225-230.2001

[4] Chen, G. Q., & Jiang, X. R. (2018). Next generation industrial biotechnology based on extremophilic bacteria. *Current Opinion in Biotechnology*, *50*, 94–100. https://doi.org/10.1016/J.COPBIO.2017.11.016

[5] Reyer, M. A., Chennakesavalu, S., Heideman, E. M., Ma, X., Bujnowska, M., Hong, L., Dinner, A. R., Vanderpool, C. K., & Fei, J. (2021). Kinetic modeling reveals additional regulation at co-transcriptional level by post-transcriptional sRNA regulators. *Cell Reports*, *36*(13), 109764. https://doi.org/10.1016/j.celrep.2021.109764

[6] Chao, Y., Papenfort, K., Reinhardt, R., Sharma, C. M., & Vogel, J. (2012). An atlas of Hfq-bound transcripts reveals 3’ UTRs as a genomic reservoir of regulatory small RNAs. *The EMBO Journal*, *31*(20), 4005–4019. https://doi.org/10.1038/EMBOJ.2012.229

[7] Chao, Y., & Vogel, J. (2016). A 3’ UTR-Derived Small RNA Provides the Regulatory Noncoding Arm of the Inner Membrane Stress Response. *Molecular Cell*, *61*(3), 352–363. https://doi.org/10.1016/J.MOLCEL.2015.12.023

[8] Wang, Z., Gerstein, M., & Snyder, M. (2009). RNA-Seq: a revolutionary tool for transcriptomics. *Nature Reviews Genetics 2008 10:1*, *10*(1), 57–63. https://doi.org/10.1038/nrg2484

[9] Ozuna, A., Liberto, D., Joyce, R. M., Arnvig, K. B., & Nobeli, I. (2020). baerhunter: an R package for the discovery and analysis of expressed non-coding regions in bacterial RNA-seq data. *Bioinformatics*, *36*(3), 966–969. https://doi.org/10.1093/BIOINFORMATICS/BTZ643

[10] Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, *15*(12), 1–21. https://doi.org/10.1186/S13059-014-0550-8/FIGURES/9

[11] Liao, Y., Smyth, G. K., & Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*, *30*(7), 923–930. https://doi.org/10.1093/BIOINFORMATICS/BTT656

[12] Langfelder, P., & Horvath, S. (2008). WGCNA: An R package for weighted correlation network analysis. *BMC Bioinformatics*, *9*(1), 1–13. https://doi.org/10.1186/1471-2105-9-559/FIGURES/4

[13] Cantalapiedra, C. P., Herņandez-Plaza, A., Letunic, I., Bork, P., & Huerta-Cepas, J. (2021). eggNOG-mapper v2: Functional Annotation, Orthology Assignments, and Domain Prediction at the Metagenomic Scale. *Molecular Biology and Evolution*, *38*(12), 5825–5829. https://doi.org/10.1093/MOLBEV/MSAB293

[14] Wu, T., Hu, E., Xu, S., Chen, M., Guo, P., Dai, Z., Feng, T., Zhou, L., Tang, W., Zhan, L., Fu, X., Liu, S., Bo, X., & Yu, G. (2021). clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *The Innovation*, 2(3), 100141. https://doi.org/10.1016/J.XINN.2021.100141

[15] Wickham H (2016). ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York. ISBN 978-3-319-24277-4, https://ggplot2.tidyverse.org

[16] *CRAN - Package ggsci*. (n.d.). Retrieved March 10, 2024, from https://cloud.r-project.org/web/packages/ggsci/index.html

# Integration of Simulator into AAS for Heat Transfer Station

Petr Cernocky
*Brno University of Technology*
*Department of Control and instrumentation*
Brno, Czech republic
xcerno29@vut.cz

Ing. Jakub Arm, Ph.D.
*Brno University of Technology*
*Departmen of Control and instrumentation*
Brno, Czech republic
jakub.arm@vut.cz

***Abstract*— This work examines the Asset Administration Shell (AAS) in the role of modern management of production units, which is one of the key enablers of the Industry 4.0 phenomenon. The main idea is to connect the simulator with AAS and create a functional architecture that will correspond with the real system. The next step would then be to connect this system with the real one and manage it using this architecture. The open-source program OpenModelica was chosen for the simulation because of its versatile command line use. As an example, a heat transfer system is modeled and simulated. As the results indicate, the OpenModelica simulator takes time to run a simulation that depends on hyperparameters such as samples per interval. Also, submodel for simulation adapter could be templated.**

***Keywords—AAS, Modelica, Industry 4.0, OpenModelica***

## I. Introduction

AAS is used for the encapsulation of a production unit (asset). Usually, it holds its data in a structured way using submodels [5]. It acts as a server waiting for user requests which are then served. Recently, AAS is used in the active form to provide advanced features, such as optimization and pattern detection.

This paper presents an architecture of AAS, which incorporates a simulator providing simulations of the system model. The OpenModelica tool is described the open-source tool for subsequent implementation. A heat transfer unit (Figure 1) is used as the system which can be difficultly modelled by the traditional analytic methods. The results also exhibit the simulator results and properties such as time of execution.

## II. AAS – ASSET ADMINISTRATION SHELL

Asset Administration Shell is a conceptual model and architectural framework created as part of the Industry 4.0 initiative, which deals with the digitization and automation of industrial processes. AAS serves to represent, manage, and share information about industrial assets and devices in digital form.

AAS can also be considered a specific version of a digital twin. AAS can also be considered a key unit in the field of the idea of Industry 4.0 and its efforts for decentralized production management - service oriented architecture.

### *A. Structure of AAS*

The Administration Shell consists of two parts. Header, which contains the basic unique identifiers of both the Asset and the digital envelope. And the body (Body), it contains additional information - submodels, e.g. for communication or about the current state of the product. Submodels can be nested and hold data elements which could provide a property, method, or event. Every data element could be extended using modifiers, such as HasSemantic. [1]

Fig. 1. The physical heat transfer system.

### *B. Types of AAS*

The form of AAS is determined by its features in the information model.

- **Passive AAS** - This type of AAS focuses on passive monitoring and surveillance of industrial facilities. Passive AAS only collects data, it does not otherwise interfere with the operation of the device. This type is used in situations where there is no need to control the equipment, but only to monitor and obtain information for the purposes of diagnostics, maintenance and planning of the production process. [2]

- **Reactive AAS** The second type of AAS uses a communication interface to provide access to the passive part. The communication interface is standardized and uses one of the TCP/IP based technology. [2]
- **Proactive AAS** The third type of AAS uses active management and response to changes or events concerning industrial assets and equipment. Proactive AAS is designed to interact with the environment and perform actions based on various stimuli and scenarios. This increases AAS's ability to effectively manage industrial processes and optimize equipment performance. [2]

## III. MODELING TOOLS

Manufacturing process simulators are tools or software that allow you to model, simulate, and analyze various aspects of a manufacturing process. These simulations are designed to help companies, developers and engineers better understand and optimize manufacturing processes. There is an abundance of such tools on the market, whether they are paid (Siemens) or freely downloadable (OpenModelica).

The other approach is to use a mathematical model to simulate each device of a system, such as [3] for the heat exchangers. Also, other components such as pipes, valves, etc need to be modelled from the perspective of the control loop.

### A. *Modelica*

Modelica is a cyber-physical system modeling language that supports the acausal coupling of components governed by mathematical equations to facilitate modeling from first principles. It provides object-oriented constructs that facilitate model reuse and can be conveniently used to model complex systems containing, for example, mechanical, electrical, electronic, magnetic, hydraulic, thermal, control, electrical, or process-oriented subcomponents. [3]

### B. *OpenModelica*

OPENMODELICA is an open-source modeling and simulation environment based on the Modelica language intended for industrial and academic use. Its long-term development is supported by a non-profit organization - the Open Source Modelica Consortium (OSMC). The goal of the project is to create a complete Modelica modeling, compilation and simulation environment based on free software distributed in source code and executable form intended for research, teaching and industrial use. [4]
OpenModelica includes a user-friendly graphical environment for model implementation. However, this model can also be created using a text editor using the Modelica language.

## IV. USAGE PROCEDURE

In the first step, the passive AAS part of the given system is created using the AAS Package Explorer. This is then stored with the help of Basyx in the CouchDB database. In the next step, the initial data is read from the database and used to create a file that OpenModelica continues to work with. The simulation results are further stored in a specific submodel as explained in the next section. The architecture (Figure 2) therefore implements an adapter for the simulator application.

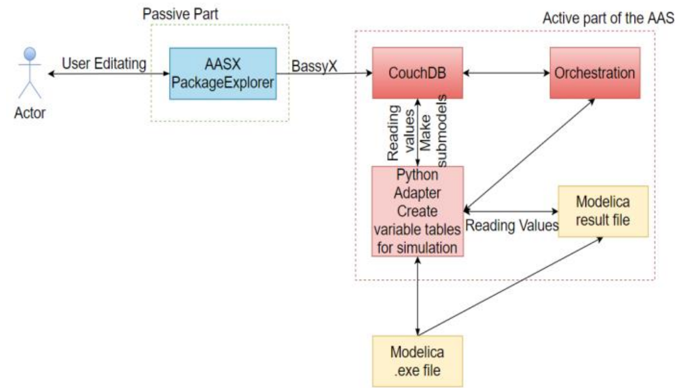



Fig. 2. Architecture of simulator incorporated into AAS

## V. HEAT TRANSFER STATION

As a demonstration system for control with the help of AAS, the demonstration system of the exchanger station, which is in the UAMT FEKT laboratory, was chosen (Figure 1). This station is currently controlled by Siemens PLC. Exchange stations are used in industry to transfer thermal energy between different media. Heat exchangers are used for this purpose. The central component of this station is the Alfa Laval CB14 heat exchanger. Furthermore, this system contains two Grundfos Alpha+ 25-40 circulation pumps, thermal and pressure sensors. The three-way valves RV111 with servo are the most important for control options. They are used to regulate the water temperature in the primary and secondary circuits.

## VI. OPENMODELICA MODEL

The model of the exchanger system was created in the OpenModelica program, where libraries with ready-made models of individual devices are freely available. These models can be easily made using the drag and drop function. In these ready-made models, all initialization values need to be set correctly. To make the simulation as fast as possible.

The parameters of the model, for example efficiency of the heat exchanger, were obtained from the datasheet of the individual components, which is included in the heat exchanger station.
The simulation is used to verify various control scenarious, and the results of this simulation are further saved in a csv file, which is further worked with in AAS. A graphical representation of the user interface and the exchange station model is shown in Figure 3.

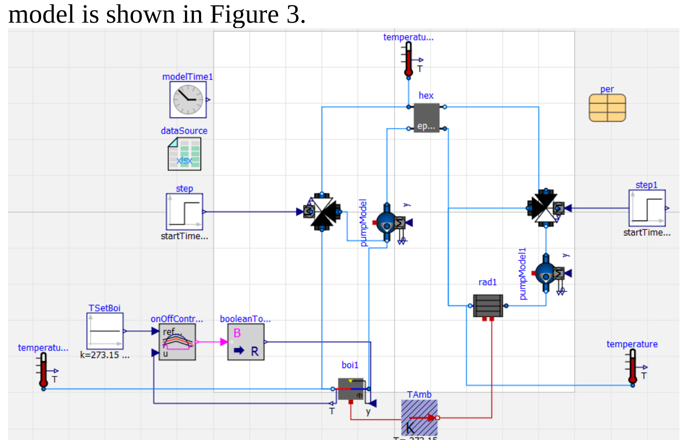


Fig. 3. Heat Trasfer Station OpenModelica Model

### A. *OpenModelica communication with AAS*

The created .xlsx file is used for communication with AAS. Which contains the basic data that we read from the real system (PLC) with the help of AAS.

Modelica offers possible libraries for further communication, which include both MQTT (Message Queuing Telemetry Transport) and, for example, TCP/IP (Transmission Control Protocol Internet Protocol) communication. However, for MQTT there was a problem with compatibility with internal libraries in Windows.

### B. *Result of the simulation*

The model was run for different time intervals and with different sampling rates. Simulations were performed with times from 1s up to 100000s. Sampling was chosen from 100 to 10000 samples. The resulting simulation times are shown in the graph in Figure 4.

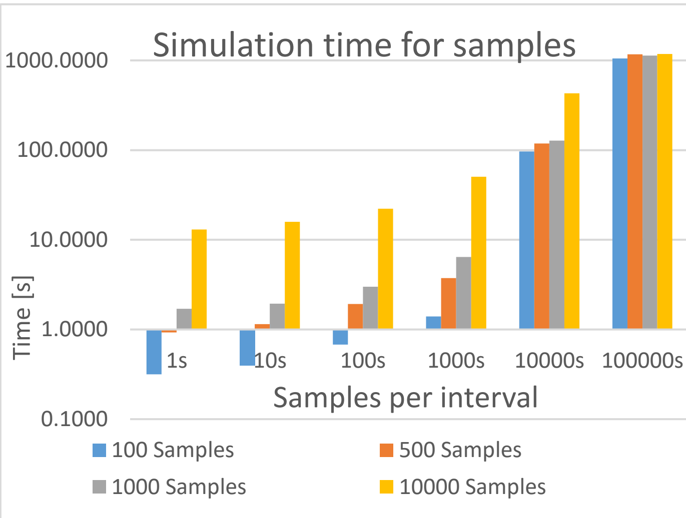



Fig. 4. Simulation times

Figure 5 shows the time course for heating the liquid using the boiler. The second curve shows the gradual heating of the radiator - cooler. The initial drop is caused by the init values - the selected ambient temperature. If other parameters are selected, this curve straightens.

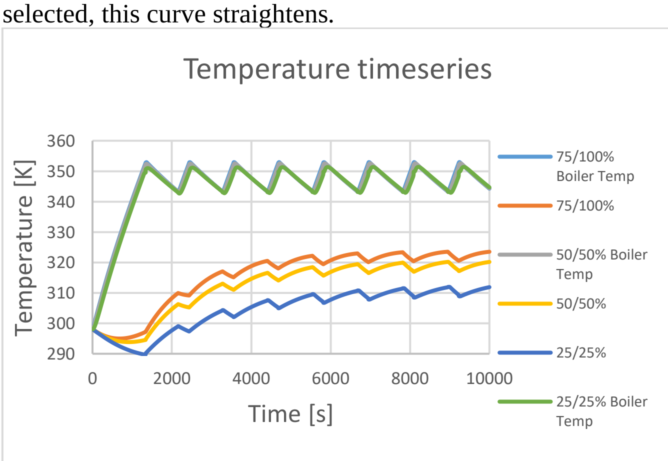



Fig. 5. Graphic representation of the liquid heating process for different scenarious

## VII. Simulation Scenarious

As part of the simulation of the entire system, three-way valves are selected as regulating elements. The simulation is selected as a combination for valve opening percentages. That's 25, 50, 75 and 100% that means a total of 16 simulations. The speeds of these simulations are different due to different nonlinearities of the system. Furthermore, the results of these simulations will be stored in AAS and the most suitable setting will then be evaluated and then applied to the real system.

## VIII. AAS Simulation Submodel

The simulation submodel is filled from a csv file after the simulation is finished using the Basyx tool and libraries. The names of the variables are determined in OpenModelica, but they must be modified due to the fact that, according to the IDTA (Industrial Digital Twin Association) group standard, the ShortId must not contain any special characters or spaces.

After the submodel is inserted into CouchDB, there is no need to create additional submodels. Then we just update the values stored in individual variables and possibly create an array of values.

## Conclusion

This paper presents the architecture of a simulator incorporated into AAS. As the demonstration, the model of the heat exchanger station was created and the simulation speeds for different time intervals were measured and evaluated. Moreover, the communication between the simulator and AAS is presented here, whereas the possibility of loading information from the created document is also discussed. Another direction in which the work will go is the creation of the AAS user interface and its deployment in the possibilities of simulating and controlling real equipment.

## Acknowledgment


The completion of this paper was made possible by the grant No. FEKT-S-23-8451 - “Research on advanced methods and technologies in cybernetics, robotics, artificial intelligence, automation and measurement” financially supported by the Internal science fund of Brno University of Technology.


## References


[1] J. Arm *et al.*, “Automated Design and Integration of Asset Administration Shells in Components of Industry 4.0”, *Sensors*, vol. 21, no. 6, 2021, doi: 10.3390/s21062004.

[2] C. Diedrich, “Asset Administration Shell”, *I40 - Industrie 4.0*, 2020. https://www.i40.ovgu.de/i40/en/Asset+Administration+Shell.html (accessed Nov. 25, 2023).

[3] Modelica® – A Unified Object-Oriented Language for Systems Modeling, Version 3.7. Modelica Association, 2023.

[4] P. Fritzson, "The Openmodelica Environment for Building Digital Twins of Sustainable Cyber-Physical Systems," *2021 Winter Simulation Conference (WSC)*, Phoenix, AZ, USA, 2021, pp. 1-12, doi: 10.1109/WSC52266.2021.9715443.

[5] Platform Industrie 4.0, “Details of the asset administration shell - part 1”, version 3.0rc02, 2022. Available from: https://www.plattform-i40.de/IP/Redaktion/EN/Downloads/Publikation/Details_of_the_Asset_Administration_Shell_Part1_V3.pdf.

[6] H. A. NavarroI, L. C. Cabezas-Gómez. “Effectiveness-ntu computation with a mathematical model for cross-flow heat exchangers”, *Fluid Dynamics, Heat and Mass Transfer and Other Topics*, 2007, doi:https://doi.org/10.1590/S0104-66322007000400005.

# Decentralised Production Using Asset Administration Shell

Jakub Maslowski
*Department of Control and Instrumentation, FEEC*
*Brno University of Technology*
Brno, Czechia
221003@vut.cz

Tomáš Benešl
*Department of Control and Instrumentation, FEEC*
*Brno University of Technology*
Brno, Czechia
tomas.benesl@vut.cz

***Abstract*—This work gradually summarizes key areas in decentralised production and asset administration shell. The infrastructure stands on Siemens PLCs with OPC UA servers. Furthermore, this paper presents a practical solution together with testing results on applications of industry 4.0 concept of asset administration shell, serving as a proof of concept for future research in this area. The findings underscore the potential of decentralized production coupled with asset administration shell to revolutionize industrial processes, offering enhanced flexibility, efficiency, and autonomy.**

***Index Terms*—decentralised production, asset administration shell, digital twin, plc, opc ua**

## I. Introduction

Decentralised production has lately become one of the key ideas in industrial world. This concept, which brings more flexibility, efficiency and overall reduces dependence on central control, is getting more and more supporters. Yet there are not enough practical approaches due to lack of required new industrial background and control.

This work assess the possibility of giving products in the production line required tools, so that they would:

- Know what is supposed to happen with them, i.e. they would know their own recipes.
- Have abilities to communicate with other entities in the production line, so they could reserve actions by its recipes.
- Act appropriately, so assets with greater priority would be finished sooner, emphasising on optimal economic production.

It aims to present possible solution in form of newly developing concept Asset Administration Shell, which builds upon already existing infrastructure, thus reducing initial financial requirements. Yet at the same time it revolutionise the way of production by giving some form of "intelligence" to product itself.

## II. Asset Administration Shell

Concept of Asset Administration Shell (AAS in short) has been around for a few years, yet it remains relatively unknown to many. Therefore this section provides some form of illumination of this idea.

### A. Digital Twin as a Baseline of AAS

For the purposes of this article, the meaning of a digital twin is understood at its highest level of integration, as explained in [1], meaning the data flow between physical and digital objects are fully integrated.

The notion of the digital twin stands as a cornerstone in the realm of Industry 4.0 technologies, serving as a counterpart which mirrors physical assets. In the context of AAS, the digital twin emerges as a fundamental element, upon which AAS principles are built.

### B. Current Understanding of AAS

First major source of information in concept of AAS stemmed primarily from Platform Industrie 4.0, a prominent association supported by Germany. This organization played a pivotal role in the theoretical AAS development as a whole.

Later they merged into Industrial Digital Twin Association, where they adopted more practical approach. The IDTA boasts numerous companies, including major ones, which are the forefront of advancing Industry 4.0 concepts.

However, their open-source specifications of AAS primarily emphasize the commercial aspects, as can be observed from their AASX Package Explorer, available at https://github.com/eclipse-aaspe/aaspe. As mentioned earlier, this paper focuses more on the industrial aspects, so it can at most be inspired by their specifications and transform some of their concepts into practise.

Asset Administration Shell can be described as an advancement upon the concept of a fully developed digital twin. It got much more communication possibilities, meaning it is able to communicate with other AAS using predefined rules. It is worth a mention that the way of how AAS store its data is strict, where each type of data is stored inside predefined submodels, each containing data for the given group.

### C. Functionalities

Asset Administration Shell can be divided into two parts: Passive AAS and Active AAS. The passive AAS takes care of proper storing of its own data. That is their submodels and identifications. The active part on the other hand provides communication with other entities in the factory. The I4.0 language is defined in VDI 2193/1-2, from which this work draws required standards to ensure proper AAS communication possibilities. [2]

AAS communication participants are in service driven industry divided into two main groups. First of them are called Service Providers (SP). These entities are represented by machines in a factory, in this case PLCs, which offer their services to the second group. The second group is represented by Service Requesters (SR), e.g. assets. Their primary objective is to identify SPs capable of meeting specific requirements outlined in their individual recipes, thereby enabling them to allocate tasks, so they can reserve their actions there by predefined communication infrastructure mentioned earlier.

## III. Implementation

Data model presented in this work is not concerned about giving devices possibilities to either be both SP and SR at the same time or gradually change their status in their lifetime, although that is possible and even encouraged by several sources at IDTA. The reason is that additional work would not justify the few advantages resulting from their implementation, especially since this work aims to serve as a proof of concept, endeavouring to apply theoretical understandings of AAS into practise.

### A. Requirements

Implementation of AAS mentioned in this paper gradually follows up on authors previous work [3] and the infrastructure there created. Each device is represented by Siemens PLC S7-1500 due to their widespread and broad support, especially possibilities of their virtual simulation using PLCSIM Advanced V4.0, which enables testing of decentralised production at much larger scale.

OPC UA is the communication protocol of a choice, requiring OPC UA servers on each SP and OPC UA clients on SRs. The reason beying their object data representation and robust real-time communication even with higher network load. [3]

One of the final requirements is the diversified production itself. For decentralised production of this nature to be more profitable than standard hierarchical control, highly customisable products are necessary.

### B. Data Model

As outlined earlier, communication is performed by OPC UA methods on PLC servers together with shared variables via same protocol. Communication environment was made in form of XML file using SiOME application as showed on Figure 1. Each method has one input and one output parameter, always of a type string[1].

Priority queues are implemented in every PLC in order to ensure continuous production without unnecessary delays. Furthermore, the decision to utilize priority queues allows products to be handled based on their respective priorities, which may be for instance determined by factors as their urgency in manufacturing or their potential profitability.

[1]Data type in PLC is WString due to its compatibility.

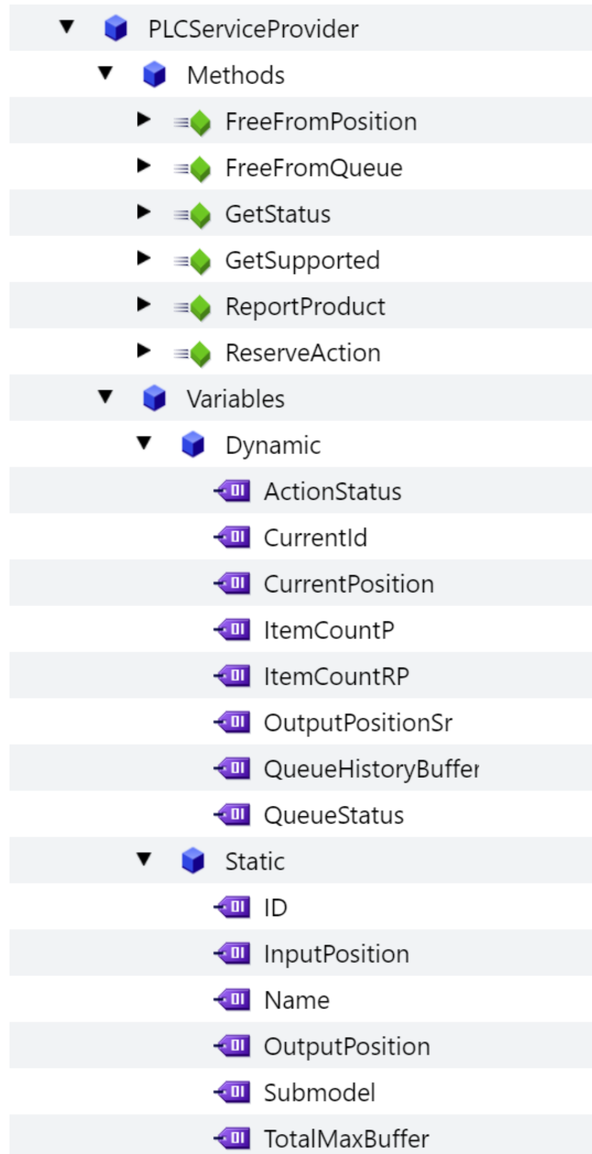



Fig. 1. OPC UA environment for PLC.

Moreover, each PLC maintains a queue of reported products, where products within the sphere of influence of the respective device are located. This implementation mechanism was devised to enable machines to identify which products from the priority queue could be worked upon, a capability that would not be possible in a virtual testing environment otherwise.

### C. Service Providers

There are three submodel types of service providers, where every PLC in presented data model is assigned to one specific submodel:

- **Storage** - giving and storing product materials
- **Manufacturing_Unit** - making changes on products
- **Manipulator** - transporting products between service providers

The unified architecture of OPC UA environment ensures that, although each type of submodel has different services to offer, the communication interface remains the same for all of them. This contrasts with previous research in this area [3], where this was not the case due to the early stage of a development.

Each PLC has its own data block for priority queue, as well as separate one for the queue of reported products.

Active AAS is ensured by OPC UA methods and several variables. Functionalities of those methods are as follows:

- **FreeFromPosition**: Product reports to device, that it is no longer occupying specified position.
- **FreeFromQueue**: Product removes itself from devices priority queue.
- **GetStatus**: Product asks device about his status.
- **GetSupported**: Product asks the price of requested service.
- **ReportProduct**: Product reports its position in devices sphere of influence.
- **ReserveAction**: Product reserves its service into service providers priority queue.

On the other hand, variables inform all products of their internal data flow, such as queues fulfillments, statuses of various actions, the type of submodel they represent, and so forth.

Priority queues are implemented in form of multilevel queue due to the fact that PLCs do not support dynamic memory allocation. They are stored in hierarchical structure using custom user defined types. There have been created low level functions for queue: clear, find, get, push, remove and high level functions for priority queue: clearAll, clearPriority, find, get, push, removeById and removeByPosition. These are used by respective OPC UA methods. PLC resource consumptions by priority queues are managable even for larger priority queue with a maximum of 10 priorities, each having 1000 reserved products.

Passive AAS is implemented in a unique specific form. The distinction between individual submodels is neither possible nor even requested in PLCs. For these reasons, only a basic form of passive data is available trough static shared variables. However, passive AAS also contains logic for internal data maintenance, such as the device itself and upkeep of priority queues.

### D. Service Requesters

Service Requesters as OPC UA clients were implemented in Python, despite the initial counterintuitive impression due to its overall slowness in these types of applications. The main reasons behind this choice could be summed up as follows:

- Source code for submodels by IDTA is in Python, which enables future cooperation.
- Python is great language for Artificial Intelligence (referred to as AI thereafter) in terms of libraries. That is important in possible future applications in SR with aim to optimise large masses of SRs.
- Speed of which python code is executed is insignificant in comparison with communication times between clients and several PLCs in a factory.

It has been opted to centralize service requesters into main AAS hub, which is connected to all OPC UA servers. This approach alleviates the strain on OPC UA servers, mitigating potential overload resulting from a high volume of clients represented by individual service requesters. They are stored in local sqlite3 database, as shown on Figure 2.

## IV. TESTING

The entire system was tested on 29 virtual PLCs: 3 Storages, 8 Manipulators and 18 Manufacturing units, all supporting specific materials and operations with different parameters. Every device had priority queue with 10 possible priorities, each having only 10 available spots for products, meaning 100 in total. This was done to overload queues, so service requesters would have to reserve their actions as ones with lower[2] priorities into available spaces, so effects on decreased product prioritisation could be better assessed even in these critical conditions.

AAS hub contained at the start 20 products with priority 2, 10 products with priority 5 and 30 products with priority 7. Further products were added during runtime, so already reserved products would have the option to re-register at different devices according to respective queue fulfillments. Each SR had on average 7 different required operations in their recipe.

There have been no problems with PLCs, mostly due to the fact that main AAS hub approached PLC servers gradually, so they were not overwhelmed. Product prioritisations were lagging due to the fact, that SRs were not unregistering themselves gradually in case of full queues. This did not happen with repeated testing in situations where each priority queue had more available space. However, even that did not help in situations, when several PLCs were in runtime turned into failure to simulate how products would react in case of broken devices. There have been no significant lagging in terms of AAS hub having too many SRs, since times required for devices to perform their operations were between 2,5 and 10 seconds, depending on device type and their ability to perform specific operation. In a scenario where this could become a problem, local sqlite3 database enables working with multiple AAS hubs, which reduces their individual complexities and thus decreasing delay between service requester actions.

## V. FUTURE WORK

One of the crucial considerations for the widespread adoption of AAS across diverse industrial environments are interoperability and scalability. Future work ought to focus on enhancing the scalability of AAS to support interconnected networks of assets. Additionally, efforts to improve the interoperability between different AAS implementations will be

[2]Priorities were in range of [1, 2, ... 10], where lower number = higher priority in the production.

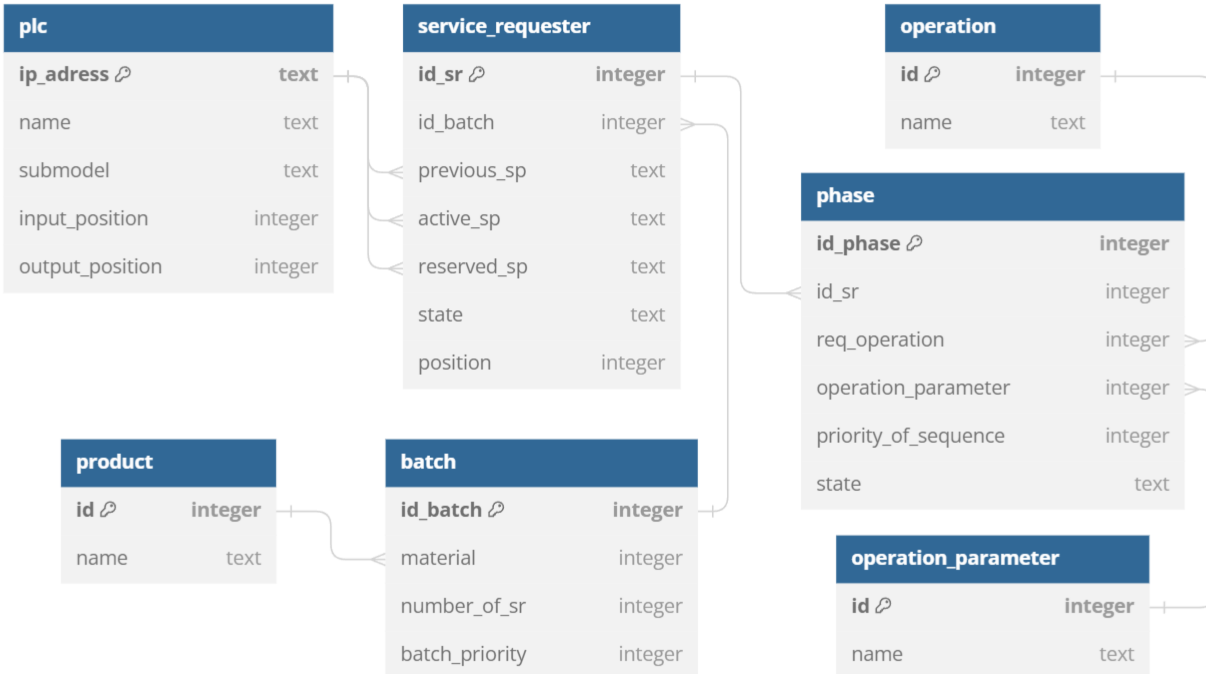



Fig. 2. Service Requesters database.

essential to enable future integration and compatibility across heterogeneous industrial ecosystems. For these reasons giving each AAS the option to be both SP and SR at the same time will be necessary.

Second major topic of future work is the enhanced autonomy and decision-making capabilities of AAS. Future research could explore the integration of AI, such as reinforcement learning and predictive analytics, to enable autonomous decision-making. This could further optimize production processes in real-time.

## VI. Conclusion

In this study, a novel method of decentralised production has been introduced, yielding results that have surpassed initial expectations. Trough the exploration of AAS as an advancement upon the concept of digital twins, the potential to revolutionize production processes has been vividly showcased. This framework empowers products with the capability to understand their recipes, communicate with other entities in the production line and act appropriately to optimize production processes.

Decentralised production with implementation of asset administration shell has proven to be a viable concept with potential applications in industry. by providing products with inherent ”intelligence”, AAS facilitates a level of adaptability and efficiency previously unseen in traditional production setups. Furthermore, the identification of certain deficiencies trough the implementation process presents valuable opportunities for future research in this area.

Overall, this study underscores the potential of decentralized production coupled with Asset Administration Shell. As industries continue to evolve towards more flexible, efficient and autonomous production systems, the integration of AAS stands as a pivotal step towards its realisation. Moving forward, continued research and innovation in this domain promise to unlock even greater efficiencies and capabilities.

## References

[1] KRITZINGER, Werner; KARNER, Matthias; TRAAR, Georg; HENJES, Jan and SIHN, Wilfried. Digital Twin in manufacturing: a categorical literature review and classification. Online, PDF. IFAC-PapersOnLine. 2018, vol. 51, no, 11, p. 1016-1022. ISSN 2405-8963. Available at: https://doi.org/10.1016/j.ifacol.2018.08.474. [cit.2024-03-05].

[2] BADER, Sebastian; Barnstedt, Erich; Bedenbender, Heinz; Berres, Bernd, Billmann Meik et al. Details of the Asset Administration Shell. Part 1 - the exchange of information between partners in the value chain of Industrie 4.0 (Version 3.0RC02). Online, PDF. Federal Ministry for Economic Affairs and Climate Action (BMWK). 2022. Available at: https://industrialdigitaltwin.org/en/wp-content/uploads/sites/2/2022/06/DetailsOfTheAssetAdministrationShell_Part1_V3.0RC02_Final1.pdf. [cit.2024-03-05].

[3] MASLOWSKI, Jakub. Asset Administration Shell pro PLC. Online, bachelor’s thesis. Brno University of Technology, Faulty of Electrical Engineering and Communication Technologies, 2022. Available at: https://www.vut.cz/www_base/zav_prace_soubor_verejne.php?file_id=241426. [cit. 2024-03-05].


thermoscientific

# SCARA robot as a 3D printer

1st Pavel Cviček
*Department of Control and Instrumentation*
*Brno University of Technology*
Brno, Czech Republic
xcvice02@vutbr.cz

2nd Tomáš Benešl
*Department of Control and Instrumentation*
*Brno University of Technology*
Brno, Czech Republic
xbenes23@vut.cz

***Abstract*—This article focuses on the possibility of using a SCARA robot with an alternative Beckhoff control system for 3D printing. The technologies and components required for implementation will be listed. Furthermore, the detailed design of the control program to execute the g-code will be described, including shortcomings and possible solutions. Also included is a tutorial that will guide the user through the setup of the slicer and will specify the necessary subsequent modifications to the g-code in order to run it. Finally, the achieved results and further development possibilities will be presented.**

***Keywords—SCARA, 3D printing, Beckhoff, Manipulators***

## I. Introduction

The SCARA (Selective Compliance Assembly Robot Arm) robot has a wide range of applications in the industry. This type of robot is most commonly used for applications such as Pick & Place or assembly applications due to their high accuracy and speed [1].

This work will focus on a SCARA robot from EPSON, markings H554BN, which however is no longer controlled by the original system but by a new system from the manufacturer Beckhoff.

The aim is to show that the SCARA robot with its new control system can perform other kinds of activities than just material transfer and the implementation of a G-code interpreter for 3D printing will be shown in detail.

## II. SCARA robot kinematics

Important terms in kinematics are the direct and inverse kinematics. They make it possible to plan the trajectories of robots. To solve the direct kinematics, it is necessary to know the angle of rotation or extension of the individual joints, which will provide getting the end position. On the other hand, the inverse kinematics solves how to set the angles of the joints to bring the manipulator to a specific point. Unlike the direct, the inverse is ambiguous, meaning that there are different ways to get to the endpoint [2].

The computation of the direct kinematic is in each cycle, for that is used the FB_KinCalcTrafo function. This makes it possible to detect the current position of the robot.

The program also needs to create a kinematic group that converts the motion into X, Y, and Z cartesian coordinates. This can be achieved using the FB_KinConfigGroup function, which creates a kinematic group from the original machine axes and outputs virtual axes representing the individual axes in the Cartesian system. In order to create this group, all motors must first be made operational using the MC_Power function. Without creating the kinematic transformation, it would not be possible to further implement the G-Code interpreter.

## III. Control system

The whole system is controlled by an industrial computer C6015-0010, to which the cards necessary for controlling and monitoring the robot including the extruder are connected. ”Tab. 1” shows the list of connected components over the EtherCAT bus. In addition, a PWM MOSFET module HA210N06 is used to control the temperature of the HotEnd.

TABLE I. Used Beckhoff terminals

| Terminal | Type |
|---|---|
| 2 x EL1008 | Digital input |
| 2 x EL2008 | Digital output |
| AX5203 | Servo driver |
| AX5201 | Servo driver |
| EL7031 | Stepper motor driver |
| EL3068 | Analog input 0…10 V |
| EL2502 | PWM output |

## IV. 3D printing

3D printing or additive manufacturing is the process of making three-dimensional objects from digital files. Objects in this process are created by successively laying down layers of melted material. The most suitable technology for additive manufacturing is FGF (Fused Granular Fabrication), which allows direct extrusion onto the printing surface using a pellet extruder. This technology is very similar to FDM (Fused Deposition Modeling) which uses a print string (filament) instead of a granulate. FGF has an advantage over more common FDM printers in manipulation as the print do not have string that could become tangled when the manipulator moves. The biggest advantage is the cost of the granulate, which is significantly lower due to the simpler production and storage options [3][4].

### A. *Required components*

The Mahor V4 pellet extruder from MAHOR XYZ can be selected for this system. It can print directly from plastic granulate, which can be of different types. Thanks to the bipolar stepper motor Nema 17 and a gearbox with a gear ratio of 5.18:1, the extruder is very powerful. It can print up to 200 g of plastic per hour with a maximum print speed of 60 mm/s. Thanks to the heating element it can reach temperatures of up to 300 °C [5].

An effector has been designed for the extruder which allows it to be mounted at different heights as required. At the same time, the gripper is designed so that if the plate with the extruder is fixed to the hollow cylinder at the highest possible height, there can not be collision of the movement with another axis of the manipulator.

Fig. 1. Effector with pellet extruder

## B. G-Code

G-code is a programming language for numerical control and it is used to control automated machines. The instructions it contains are used by the control unit, which commands the motors where to move, at what speed, and along which trajectory [6].

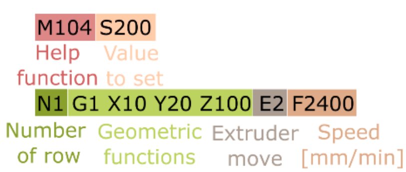



Fig. 2. G-Code example with description

There are many types of G-codes. In this case, instructions from the Sailfish (MakerBot) type will be used. This type was chosen based on the best compatibility with the implemented instructions in the Beckhoff library. The following table lists the most common commands for 3D printing, including whether this feature is implemented in the Beckhoff library. If any of the functions are not implemented, these instructions need to be subsequently replaced or removed in the required G-code.

TABLE II. COMMON USED G-CODE COMMANDS IN 3D PRINTING

| G-Code | Description | Implemented |
|---|---|---|
| G1 | Linear interpolation with programmed feed rate | Yes |
| G21 | Set units to millimeters | No |
| G28 | Home axes | No |
| G90 | Use absolute coordinates | Yes |
| G91 | Use Relative coordinates | Yes |

M-functions are used as auxiliary and preparation commands that control various machine functions. Typical commands include, for example, turning the cooler on/off, changing the tool or setting the print temperature.

TABLE III. COMMON USED M-FUNCTIONS IN 3D PRINTING

| M-Function | Description |
|---|---|
| M30 | Program end |
| M104 | Start heating up the nozzle most of the way |
| M106 | Turn On layer fan with required speed |
| M107 | Turn Off layer fan |
| M109 | Set temperature and wait for it to be reached |

# V. G-CODE INTERPRETER IMPLEMENTATION

Function blocks from the Tc2_NCI library are used to implement the interpreter. These blocks allow configuration of the interpolation group and control of the interpreter. Using this library, it is therefore possible, for example, to create a 3D group or to load and execute an NC (Numerically Controlled) program. The G-Code interpreter is implemented according to DIN 66025, which specifies the structure and content.

Before execution, the created .nc file must be loaded into the correct folder, from where it is then read out during execution.

After the creation of the kinematic group, interpolation is enabled and at the moment when the request to start the G-Code interpreter arrives, the state machine is started and the interpolation group is built. This is followed by loading the uploaded G-Code and checking if the interpeter is in the "Ready" state. At this point, the program can be started and in our case the fan for the extruder body and heater will be activated. This is followed by another check to check that we are not in the "Ready" state for at least a moment. This should ensure that the program does not end before it starts, as the next condition that watches for the end of the program, again checks if the interpreter is in the "Ready" state. When the execution of the loaded G-Code is finished, the interpolation group is destroyed and the program completes. During runtime, the currently executing G-Code can be paused or turned off completely. In case of an error, the state machine enters the Error state and it is necessary to cancel the kinematic and interpolation groups and then call the ItpResetEx2 function.

If a request to execute an M-function occurs in the code during program execution, it is evaluated based on the state of the HskMFuncReq variable in the NCI to PLC structure that provides data exchange. In the same structure can be found by the variable HskMFuncNo, which reads the specific M-function requested. In the state machine, the individual M-functions are then implemented and when they are completed, a bit is set to "True" which triggers the ItpConfirmHsk function to confirm the correct execution of the request. It is necessary to know which M-functions can occur in the generated G-code, otherwise it would be necessary to implement a state that will determine what should happen in the case that M-function is not implemented. M-functions are triggered outside the main loop of the G-Code interpeter.

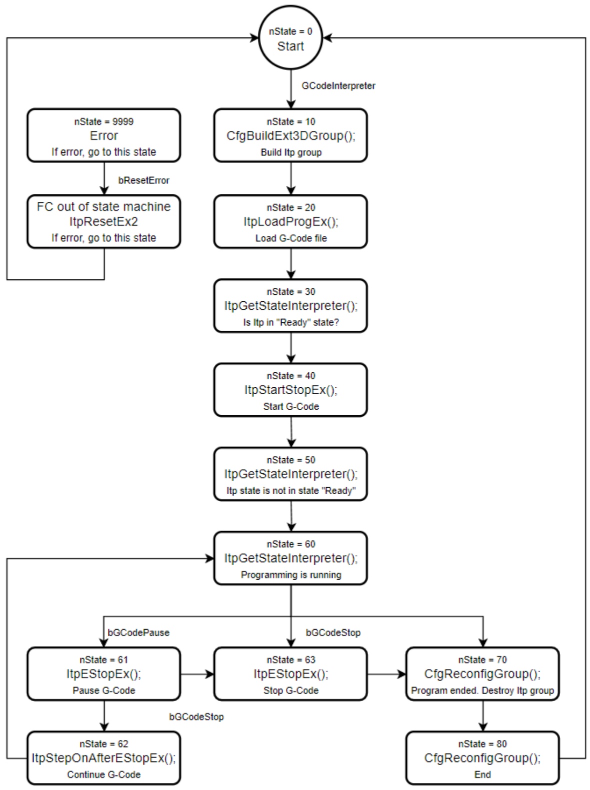



Fig. 3. Main state machine of a G-Code interpreter

## VI. PREPARING THE 3D MODEL

PrusaSlicer was used to create the G-code from STL (Stereolithography) format, in which it is possible to precisely define the parameters for your own printer according to the real configuration. This makes it possible to precisely define the workspace in which the object can be printed, this workspace is based on the limiting parameters of the robot.

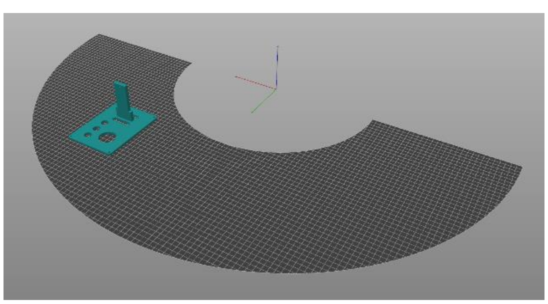


Fig. 4. Work area in PrusaSlicer

The most important parameters that need to be set are:

- Z axis offset
- Selecting the correct G-code type (default is RepRap/Sprinter type)
- Retraction distance
- Printer motion speeds (according to the maximum allowed motor speeds)
- Nozzle diameter

The exported G-code must be saved as .nc type, which can already be read from the program. Before running the program, it is still necessary to make the aforementioned modifications and edit the G-code so that it does not contain unsolicited instructions that are not implemented. It is also necessary to change the name for the extruder axis from E to Qx=, depending on which auxiliary axis the extruder motor was assigned to when the interpolation group was created.

## VII. TESTING

After uploading the G-code and running the code, it can be seen that the SCARA robot and implemented functionality works as required. As can be seen from the picture, it will be very difficult to tune the print parameters to achieve high quality prints.

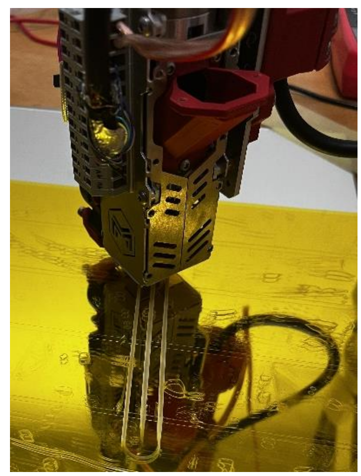


Fig. 5. First printing test

## VIII. FURTHER IMPROVEMENTS

In the follow-up activity, it will be necessary to treat error conditions that may occur during the process and could cause an error in the system. This could be, for example, the omission of a condition that disables the execution of the G-code when the heater temperature starts to drop while the program is running. At this point, irreversible damage to the extruder stepper motor could occur because the motor would no longer be able to push the unmelted material out through the nozzle.

Using the function blocks BlockSearch, ItpGetBlockNumber and similar, it should be possible to keep track of the currently executing line of G-code, so that in case of a system failure (e.g. feedback error), it is possible to jump from the error state to the last position after resetting and continue executing the program. However, it will be necessary to see if this situation is possible, mainly because we need to know the last set speed at which the entire interpolation group should move. It is also important to remember that the last position needs to be reached from a higher altitude to avoid collision between the pellet extruder and the print.

Furthermore, the reading of the values for the individual M-functions (Sxx value) will be completed and therefore it will be possible to set the individual print parameters correctly.

## IX. Conclusion

In this article, an unconventional way of using the SCARA robot for 3D printing was described. The procedure for building a G-Code interpreter in the new Beckhoff control system was shown. According to the results obtained so far, it can be concluded that after debugging the printing parameters and minor changes in the generated G-Code, the system can be used for 3D printing purposes.

## References


[1] SCARA Industrial Automation Robot. Online. In: *2018 International Conference on Power Energy, Environment and Intelligent Control (PEEIC)*. Greater Noida, India: IEEE, 2018, s. 173-177. ISBN 978-1-5386-2341-1.

[2] JAZAR, Reza N., c2010. Theory of applied robotics: kinematics, dynamics, and control. 2nd ed. New York: Springer. ISBN 14-419-1749-7

[3] FAHAD, Jabbar Qasim a QASIM, Alaa Jabbar. The Processes and Technologies of 3D Printing. Online. *International Journal of Advances in Computer Science and Technology*. 2015, roč. 4, č. 10, s. 6. ISSN 2320 - 2602.

[4] *Jak funguje 3D tiskárna: Základní technologie 3D tisku*. Online. Toner partner. 2022.

[5] *Mahor V4 pellet extruder*. Online. Filament2Print. 2023.

[6] *G-code*. Online. In: Wikipedia: the free encyclopedia. San Francisco (CA): Wikimedia Foundation, 2001-.

# Innovative control system for water supply management at a farm

Aleš Doležal
*Master Degree Programe FEEC*
*Brno University of Technology*
Brno, Czech Republic
xdolez89@vutbr.cz

***Abstract*— The thesis deals with the implementation of an innovative control system for the management and supply of water to the farm. The innovation focuses on all areas of the control system such as hardware part as well as software part. In the hardware part, it is mainly about the communication between the PLC and the pumps in the boreholes Where contact switching has been replaced by fiber optic communication. Furthermore, the non-compliant components were replaced by components that are more suitable for this placation. In the software area, tools have been used to simplify both the programming and the actual operation of the control system. The use of structures, the use of equations, and, more efficient sampling of measured data. A PLC from Unitronics was selected to control the entire application. The program for the PLC was written in the UniLogic programming environment from Unitronics.**

***Keywords*— Fiber optic, Ethernet IP, PLC, remote inputs and outputs, ladder**

## I. Introduction

### *A. Description of the old control system*

The earlier control system Fig. 1 for the water supply to the farm was already among the more modern control systems, but some of its components were already obsolete and in certain cases it had reached its physical limits. This was mainly the communication between the main switchboard and the secondary switchboards at the well pumps. In certain cases, when the distance between the main switchboard and the secondary switchboard exceeded 500 (m), there was a problem with the induced voltage from the secondary cable that carried the power for the well pump. In some cases, the induced voltage reached values of up to 130 (V), which was already reaching the switching voltage of the relays located in the secondary switchboard and causing it to trip. It was therefore necessary to solve this problem. Another problem was the ergonomics of the control and the versatility of the control system. The operator of the control system could not set different desired values, such as the level in the water tank or the outlet pressure from the water plant. Furthermore, there is no fault logging history for this system.

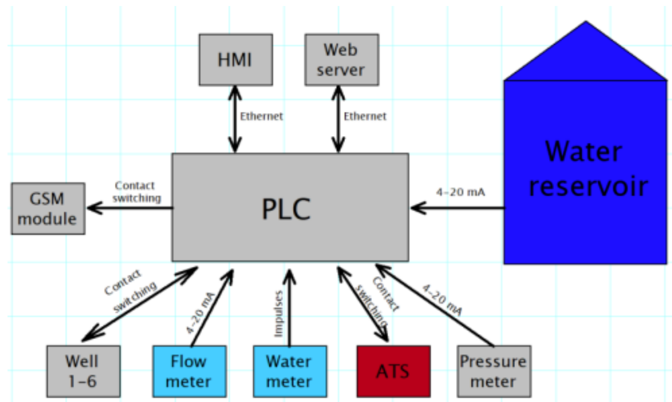



Fig. 1. Old control system [1]

### *B. Concept of the new system*

The control system Fig. 2 described in this paper is installed on a pig fattening farm that has undergone a complete renovation. The existing system, which was operated on the farm, no longer met the conditions of modern farming. It was therefore necessary to design, implement and operate a new control system for the management and supply of water. The control system described here is very similar to previous control systems that have been implemented on similar farms. The implemented water supply system is described by a PID diagram Fig. 3. The control system has several innovative solutions compared to previous control systems, the biggest innovation being in the area of communication and data transmission between the main switchboard and the wellhead columns. The previous communication using contact switching was already outdated, faulty and very prone to interference, so it was replaced by fibre optics, which is much more resistant to interference. With this technology, it is possible to add additional measurement and control elements to the borehole posts. Another innovative feature, not so visible anymore, is the unification of component suppliers. The main cabinet is divided into three parts - power, control and communication, each part supplied by a different supplier, rather than the previous applications where each component was from a different manufacturer. The last innovation is the structure of the program, in which more attention is paid to clarity and simplicity, structures, pop-up windows, etc. are used.

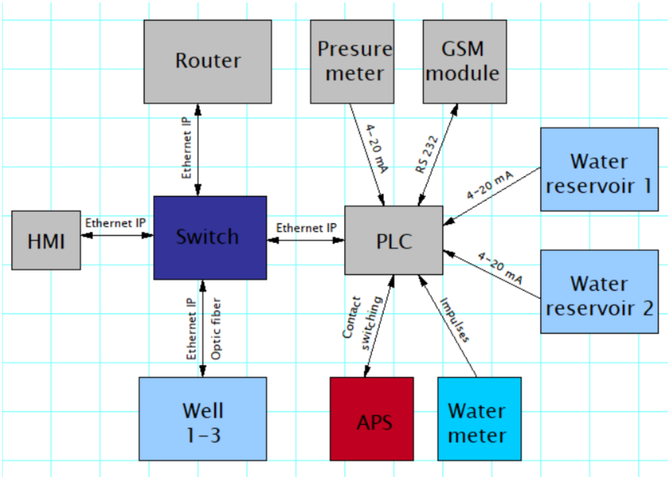



Fig. 2. Concept of new control system

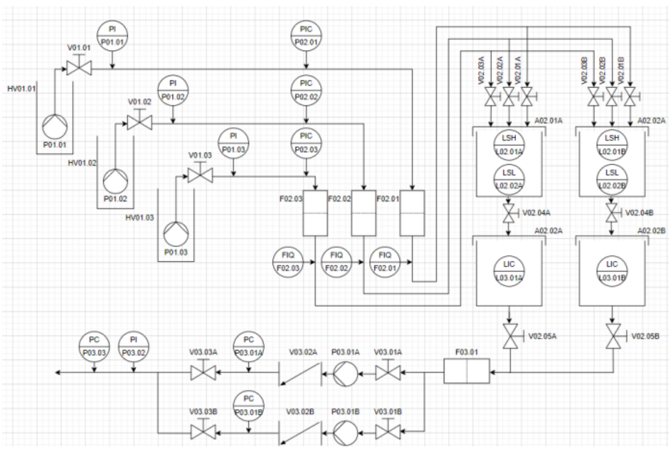


Fig. 3. PID diagram of the system

## II. Selection of Components

### A. PLC

The PLC from Unitronics was chosen to control the entire water supply technology on the farm, specifically the U-USC-B10-T42 model. This model is the same as in the previous applications. This solution is for the reason that the service does not have to stock multiple PLCs of different manufacturers and types. This PLC has 24 digital inputs, 16 transistor outputs of which two can be operated in PWM mode, it also has 2 analogue inputs supporting ranges (0-10 (V), 0-20 (mA) and 4-20 (mA)). Furthermore, the PLC contains a WebServer so that it can be connected to remotely, this feature is used in this application. The PLC is powered by 24 (VDC). Since the application contains more than two analog sensors, it was necessary to add an expansion module. Specifically, this is the UIA-0800N module, which contains 8 analog inputs.

### B. Switch

The switch used in this application is from Planet, model IGS-10080MFT. The switch is fully manageable. The switch has three RJ45 ports, the first port is used to manage the switch and the other two ports are used to connect the switch to other devices and support 10/100/1000BASE-T. In addition, it contains eight SFP trunks with 100/1000BASE-X mini-GBIC support. Three of the SFPs are plugged into MGB-TLX SFP modules with 1000BASE-LX transmission standard, which supports single mode, LC connector, and a maximum transmission distance of 20 [km]. The switch is configured using a web application. The switch is powered by 24 (VDC).

### C. Converter

For the conversion from optical fiber to metallic line, the IGT-805AT converter from Planet is installed. This converter is manufactured in an industrial design. The converter is equipped with one RJ45 connector and one SFP shaft with automatic detection of modules inserted into it. The converter does not offer a management option. The SFP shaft is complemented by an MGB-TLX SFP module with a 1000BASE-LX transmission standard that supports single mode, LC connector and a maximum transmission distance of 20 (km).

### D. Remote I/O

The remote I/Os are selected from Unitronics, from which the PLC is also selected. The remote I/Os communicate with the PLC using Ethetrnet IP. The remote I/O assembly consists of a URB-TCP2 communication module to which up to 6 modules can be connected. The input module selected was model URD-1600-8 which contains 16 digital inputs with a maximum voltage of 15 - 32 (VDC) and the output module selected was model URD-0004RH which contains four relay outputs with a maximum load of 240 (VAC) and 2 (A).

### E. Sensors

The sensors used in this application measure various non-electrical quantities. Analog hydrostatic probes are used to measure the water level in the reservoirs and mechanical switching floats are installed as a backup measurement. For the measurement of the outlet pressure, a pressure gauge based on increasing resistance as the pressure rises is chosen. A pulse water meter is used to measure the amount of water pumped.

## III. Network Configuration

### A. PLC

The configuration of the IP address for the PLC is done in the UniLogic programming environment from Unistronics. First, the IP address of the router to which the PLC is connected must be set, which is marked Default gateway. Next, the IP address of the PLC itself is set, under which it will communicate with the router. Last but not least, the subnet address or Subnet mask must be set. For this network, the subnet address 255.255.255.0 was chosen since it is a network of up to 254 devices.

### B. Remote I/O

Configuration is started by moving the switch in the ninth position on the side of the module to the top ON position, this enables the IP address setting in software. The next step is to open the UniLogic BOOTP Server, which will be offered to us when we click on the adapter we want to configure. This will open the BOOTP Server dialog box, in which we click on Start Bootp and start to see the devices on the network. We then power on the adapter, once the MAC address of the adapter appears in the dialog we double click on it, enter the desired IP address, select the appropriate network card of the computer and confirm with OK. Now we turn the adapter off again and turn it back on. Now, using the command line, we will perform a Ping test to see if the configuration went fine. If the test was OK, turn the adapter off again and switch the switch in the ninth position to the lower OFF position. If the test fails, repeat the whole procedure.

### C. Router

All router configurations were performed in the router's web application, which can be accessed by entering the router's IP address into the search engine and then entering the access password. Since remote access to control the technology is required, a public static IP address was set up. In order to connect to the PLC via the public static IP address, NAT routing was set up from the public static IP address to the PLC IP address.

## IV. Hardware configuration

### A. PLC

Hardware configuration of the PLC consisted of setting analogue and digital inputs. The PLC is equipped with four HS (high speed) inputs, but these are not required in this application, so they are set as regular input with a 6 (ms) input filter. The inputs are wired in sink logic which consists of the inputs having a common 0 (V) and switching 24 (V) to the inputs. The PLC is also equipped with two analog inputs. These analogue inputs had to be set up for the needs of the application. That is, they were set up as current analog inputs with a range of 4-20 (mA). The A/D converter ranges for both inputs were set from 0 to 4095.

### B. Expansion module

As this is an expansion module that contains analogue inputs, it was necessary to perform a similar configuration as for the PLC. The inputs were again set as current inputs and the A/D converter was again set to a range of 0 to 4095.

### C. Remote I/O

The hardware configuration of the remote inputs was again similar to that of the PLC. The inputs are connected to the sink logic and have the input filter set to 6 milliseconds. The outputs are relay contacts, so there is no need to deal with sink or source logic. Fault action and Fault value are set for the outputs. Fault action is the state when some error happens, like a power failure, and tells us what position the contact should stay in. There are two choices of Fault value or Hold last state. Fault value was selected for this application. Fault value can be set whether the contact should be ON or OFF during a fault.

## V. Programing

The following section describes the different parts of the programme that are part of the system upgrade. These parts of the program include, for example, the creation of a Ping function for checking the connection of remote inputs and outputs, switching between two storage tanks, and a pop-up window for setting the delayed switching of pumps in the well. The Fig. 4 shows a brief flowchart of the system.

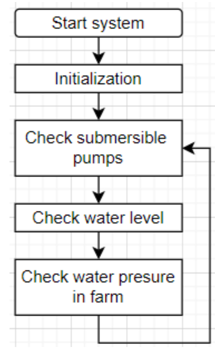



Fig. 4. Flowchart of the system

### A. Structs

To facilitate programming, one structure is created in the program called "Well". The structure contains several variables of different data types for easier and faster programming. These variables include, for example, a variable for counting pump motor hours, counting the amount of water pumped. Furthermore, these variables include triggering automatic or manual pump operation. As one of the last variables, there are variables needed for the implementation of filter flushing.

### B. Formulas.

Since the application contains analog sensors that are connected to the PLC, it was necessary to create a linearization of the measured data from these sensors in the program. The equation function was used for the linearization. The equation has the form (1). The variable range of the sensor is entered into input A. The current measured value coming from the sensor is entered into input B. Output C is a variable that corresponds to the current water level in the reservoir in meters.

$$\frac{(A * B)}{4095} = C \tag{1}$$

### C. Data sampling

The customer wanted the measured data to be displayed in graphs from which they could extract information for further planning and management of the business. In order to display the data in graphs, a sampler was needed. A total of three samplers were created. The first sampler was created to sample the measured data from the water level measurements in the reservoirs with a sampling period of 30 (s), the second sampler was created to sample the data measured from the outlet pressure with a sampling period of 10[s] and the last sampler was created to measure the pump on times in the well with a sampling period of 1 (s).

### D. Pop-up windows

Pop-up windows are created for quick access to set certain values in the application (pump delay) and for faster programming. In these pop-up windows, it is possible to set the delay time Fig. 5 for the pump to switch on if the well runs out of water and reaches the desired value again for a possible pump start. This measure was put in place to avoid breaking the pump by frequent switching, as some wells have a strong spring and would cause the probe contacts to connect when the well was being filled, which would then assess that there was enough water in the well and switch the pump on.

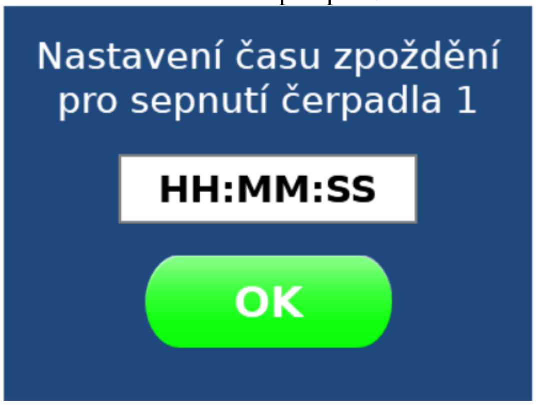



Fig. 5. Pop-up window for setting the delay time for switching on the pump in the well

## E. Ping function

The application contains devices that communicate with each other via Ethernet IP. A ping check function had to be programmed to monitor whether a given device at a given IP address is active Fig. 6. This function was written because of a communication problem with the converter and remote I/O. If they disconnected from power the PLC reported a power failure once and did not report a power failure the second time. This posed a high risk to the operator and to the water supply to the water tank itself. The function is triggered once per minute and tests all three wells simultaneously. If a well is not active the PLC evaluates this as a fault and sends a text message.

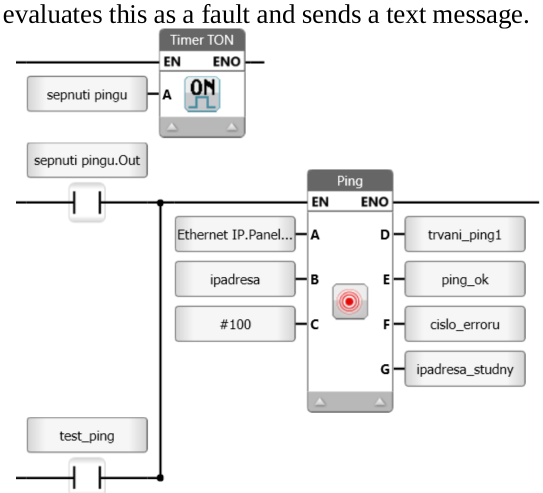



Fig. 6. Ping function

## F. Rinsing function

As it is an autonomous system, an autonomous filter rinsing system has been set up to prevent clogging and blockage. The rinsing of the filters is triggered after a certain amount of water has been pumped in. This quantity is set by the operator. After the specified quantity has been pumped, the rinsing valve is opened for a certain period of time, which is again set by the operator, the pump is started, and the rinsing takes place. After the time has elapsed, the pump switches off, the valve closes and the pumped water counter resets.

## G. Double storage tank

The application includes two pairs of storage tanks. Each pair contains an analogue hydrostatic probe that evaluates the current level and floats for emergency control of the water supply in case of failure of the hydrostatic probe. This solution was chosen for two reasons. The first reason was the indisposition of the building in which the storage tanks are located and the second reason was the cleaning and disinfection of the tanks themselves. This was to avoid interruption of the water supply to the farm and to be able to clean at the same time. When the tanks are cleaned they behave like one big storage tank. In the program, this meant that it could not evaluate from both tanks at the same time so a function had to be programmed to switch between the data from the hydrostatic probe and the floats. If Aku1 is selected on the HMI display the data from the sensors located in Aku1 is measured and if switched to Aku2 the data from Aku2 is evaluated.

# VI. Conclusion

The previous control system was successfully upgraded to a more modern, reliable, and easier-to-operate control system. The communication between the main control cabinet and the secondary control cabinets is now more reliable and can be expanded in the future with additional pump control and monitoring options. Just add another SFP module to the switch and add the well in the PLC. The operator is very happy with the simpler control and is happy with the autonomous filter cleaning. Programmatically, the application is very clear and neat. Another extension is to send monthly reports on the amount of water used, amount of faults encountered, etc.

[1] DOLEŽAL, Aleš. Control system for management and supply of water to the farm. Bachelor thesis, supervisor Miroslav Jirgl. Brno: Brno University of Technology, Faculty of Electrical Engineering and Communication Technology, Department of Control and Instrumentation, 2022..

# Design of 436.5 MHz Square Patch Antenna Fed by Coaxial Probe

V. Lukáč
*Faculty of Electrical Engeneering and Communication*
*Brno University of Technology*
Brno, Czechia
xlukac14@vutbr.cz

T. Urbanec
*Faculty of Electrical Engeneering and Communication*
*Brno University of Technology*
Brno, Czechia
ubranec@vutbr.cz

***Abstract*— This letter describes the design of a square patch antenna used for reception of satellite signals in the UHF band. This antenna is fed by coaxial probe with extended ground plane. Used dielectric substrate is Ertalon 6 SA with height of 15 mm. The designed antenna is simulated in the HFSS Ansys program. The simulation takes into account the maximum gain, the reflection coefficient and the bandwidth of the antenna.**

***Keywords—square patch antenna, dielectric substrate, gain, reflection coefficient, radiation pattern***

## I. Introduction

Microstrip patch antenna is a low profile antenna, usually printed on PCBs, which consists of ground plane or the reflector, dielectric substrate and the patch itself. Patch usually comes in different variations such as square, rectangular, triangular, circular, spiral or any kind of geometrical sheet [1]. When it comes to feeding techniques, patch antenna can be fed by microstrip line, coaxial probe, aperture couple feed or proximity couple. Patch antennas can be designed for various purposes, for example wireless portable devices (WLAN applications).

This particular square patch antenna is designed and used for antenna array for receiving signals in UHF band. Feeding of this antenna is realized by coaxial probe placed in right position for matching impedance. Lastly, the antenna design is simulated in HFSS Ansys and its reflection coefficient $s_{11}$ and radiation pattern is displayed.

## II. Antenna design

### A. Dimensions of the antenna

The structure of the square patch antenna is shown in Fig. 1. The parameters of the proposed antenna are $L_{sub}$, L and h, where $L_{sub}$ is the length of the substrate, L is the length of the square patch and h is the height of the substrate. First of all, the height of the substrate must be found. According to (2), it is possible to find in what range the height is located. The specified height must be much smaller than the center wavelength $\lambda_0$ ($h \ll \lambda_0$). The wavelength is calculated in (1), where the result is used in (2) [2].

$$\lambda_0 = \frac{c_0}{f_0} \tag{1}$$

$$0.003 \cdot \lambda_0 < h < 0.05 \cdot \lambda_0 \tag{2}$$

The height of the dielectric substrate has been chosen to h = 15 mm. According to (2), the chosen height is suitable, range of available heights is from 0.206 mm to 34.365 mm.

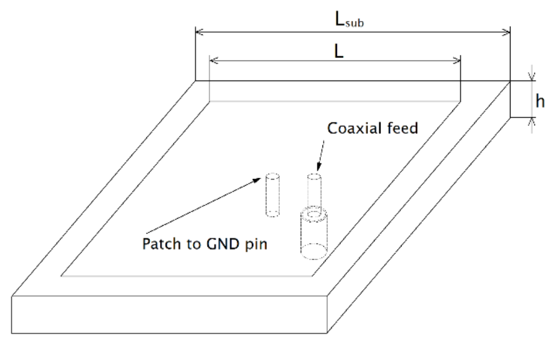



Fig. 1. Square patch antenna with coaxial feeding

After verifying the height of the dielectric substrate, it is possible to start calculating the dimensions of the square patch. The width of the patch W and the length of the patch L are the same, because the patch is a square (W = L). In order to obtain an accurate result, two iterations are used, where for example the resulted $L_0 = W_1$, $L_1 = W_2$ and so. First of all, the patch width $W_0$ is calculated in (3), then after determining the patch width, the relative effective permittivity of the substrate $\varepsilon_{reff0}$ is determined. Then the extension $\Delta L_0$ and the length of the patch $L_0$ are calculated. This entire procedure is repeated from zeroth to the second iteration [2].

*0. iteration*

$$W_0 = \frac{c_0}{2 \cdot f_r} \cdot \sqrt{\frac{2}{\varepsilon_r + 1}} \tag{3}$$

$$\varepsilon_{reff0} = \frac{\varepsilon_r + 1}{2} + \frac{\varepsilon_r - 1}{2}\left(1 + 12\frac{h}{W}\right)^{\frac{-1}{2}} \tag{4}$$

$$\Delta L_0 = 0.412 \cdot h \cdot \frac{\left(\varepsilon_{reff0} + 0.3\right)\left(\frac{W}{h} + 0.264\right)}{\left(\varepsilon_{reff0} - 0.258\right)\left(\frac{W}{h} + 0.8\right)} \tag{5}$$

$$L_0 = \frac{c_0}{2 \cdot f_r \cdot \sqrt{\varepsilon_{reff0}}} - 2 \cdot \Delta L_0 \tag{6}$$

*1. iteration*

$$L_1 = \frac{c_0}{2 \cdot f_r \cdot \sqrt{\varepsilon_{reff1}}} - 2 \cdot \Delta L_1 \tag{7}$$

*2. iteration*

$$L_2 = \frac{c_0}{2 \cdot f_r \cdot \sqrt{\varepsilon_{reff2}}} - 2 \cdot \Delta L_2 \tag{8}$$

TABLE I. CALCULATED PARAMETERS (DIMENSIONS)

| | |
|---|---|
| h | 15 mm |
| $W_0$ | 222 mm |
| $\varepsilon_{reff0}$ | 3.53 |
| $\Delta L_0$ | 4.71 mm |
| $L_0 = W_1$ | 173.54 mm |
| $L_1 = W_2$ | 174.92 mm |
| $L_2$ | 174.88 mm |

In table I, calculation results are displayed. Because the patch is square shaped, at the end of iteration, the length $L_x$ becomes $W_{x+1}$. After the second iteration, no further changes occur, and therefore it is not necessary to iterate further. All antenna dimensions are ready for simulation in the HFSS Ansys program.

## B. *Coaxial feed and impedance on the edge*

The antenna is fed by coaxial cable (Fig. 2). The inner conductor of the coaxial cable is connected to a square patch (it does not touch the ground plane). However, the coaxial feed cannot be plugged into a patch at any point. The reason is the different impedance at the point of antenna feed. For example, the impedance of the feed (coaxial cable) near the center of the patch (really close to the center pin) would be very low, while at the edge it would be maximal. To match the impedance to 50 Ω, the formulas for calculating the patch edge impedance (11) and calculating the distance $x_0$ at 50 Ω input impedance (12) must be used [2]. Lastly distance $x_1$ (13) is needed for simulation.

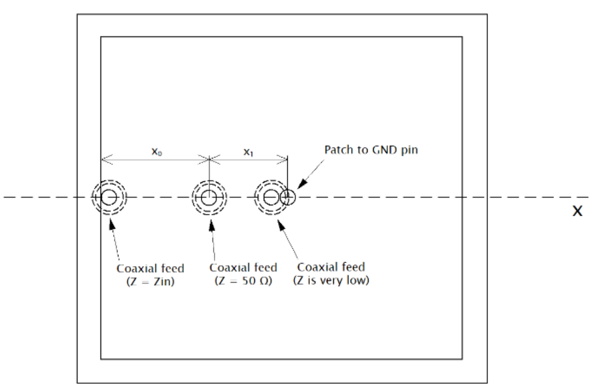



Fig. 2. Square Patch antenna with coaxial feeding

To find the required distance, the conductance and impedance at the edge of the patch must first be calculated. Conductivity $G_1$ is calculated according to (10), where W = L and $k_0$ is calculated according to (9). After determining the conductivity, the impedance at the edge of the patch (11) is calculated, the value of which is 240 Ω. Finally, the value of the distance $x_0$ is calculated at an input impedance of 50 Ω (12).

$$k_0 = \frac{2 \cdot \pi}{\lambda_0} \tag{9}$$

$$G_1 = \frac{W}{120 \cdot \lambda_0}\left[1 - \frac{1}{24}(k_0 \cdot h)^2\right] \tag{10}$$

$$Z_{IN}(x_0 = 0) = \frac{1}{2 \cdot G_1} \tag{11}$$

$$x_0 = \frac{L}{\pi} \cdot \cos^{-1}\left(\sqrt{\frac{Z_{IN}(x = x_0)}{Z_{IN}(x = 0)}}\right) \tag{12}$$

$$x_1 = \frac{L}{2} - x_0 \tag{13}$$

TABLE II. CALCULATED PARAMETERS (COAXIAL PROBE PLACEMENT)

| | |
|---|---|
| $k_0$ | 9.14 $m^{-1}$ |
| $G_1$ | 2.09 µS |
| $Z_{IN}$ | 240 Ω |
| $x_0$ | 60.39 mm |
| $x_1$ | 26.1 mm |

The values of the distance calculations are in from table II, where $x_1$ = 26.1 mm is a distance from the center of the patch. This value will be used when modeling the antenna.

# III. SIMULATION AND RESULTS

A square patch antenna fed by a coaxial cable is drawn and simulated in the HFSS Ansys program. The antenna is simulated in free space, the calculated dimensions of the patch and coaxial probe feed are applied when drawing the antenna from the previous chapter. After drawing the antenna, the reflection factor $s_{11}$ is tuned to the specified frequency $f_0$ = 436.5 MHz, the maximum gain of the antenna in the desired direction (main lobe) must be determined and adjusted, and the directional characteristic of the antenna must be drawn and its side lobes and back lobe must be suppressed as much as possible.

## A. *Antenna modeling*

First, the substrate is modeled in the shape of a square rectangle. Below the substrate is drawn a ground plane with the same x and y dimensions as the substrate. Above it is modeled a square patch 172.98 mm long and wide. The patch and the ground plane are made out of copper with a height of 0.5 mm. The copper center pin that connects the ground and the patch has width of 3 mm. The coaxial cable, which feeds the antenna, consists of 3 cylinders with height of 20 mm but the inner conductor of the coaxial cable must be 15.5 mm longer, because it serves as probe. The width of the inner conductor is 1.3 mm (Copper), the width of the dielectric is 5 mm (Teflon) and the coaxial shield is 5.5 mm wide (Copper). The entire antenna is surrounded by the air box. This air box is for a wave, radiation and far field results. Without it, the model simulation is invalid. After drawing the air cube, the square patch antenna model displayed in Fig. 3 is ready for simulation and further testing.

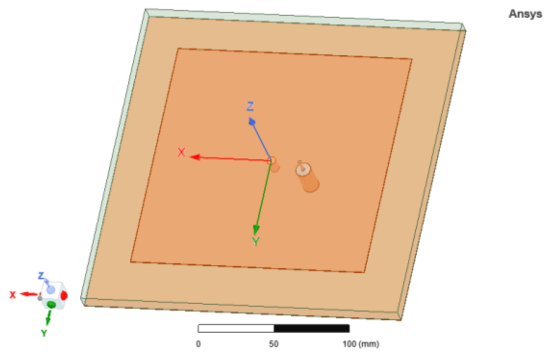



Fig. 3. Square patch antenna

## *B. Reflection coefficient*

Reflection coefficient $s_{11}$ tells how much power is reflected from the antenna. All the power from the antenna is reflected and nothing is radiated if the $s_{11}$ is approximately 0 dB [3]. Fig. 4 shows the characteristic of the reflection coefficient $s_{11}$ tuned at a frequency of 436.5 MHz. The calculated dimensions from the previous chapters do not guarantee an immediate result, but serve only as a guide. Changing the dimensions of the width of the substrate with the ground plane, changing the length of the patch and changing the position of coaxial feed will have effects on the tuning. For example, when changing the position of coaxial feed, the value of the reflection coefficient $s_{11}$ changes at the given frequency.

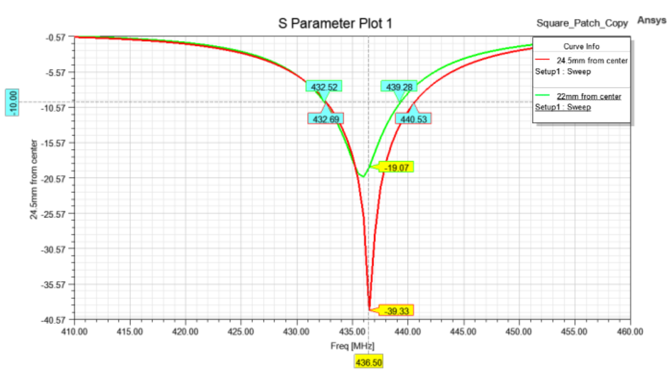



Fig. 4. Reflection coefficient $s_{11}$ with coaxial probe feed 24.5 mm (red) and 22 mm (green) from center

After changing the dimensions of the ground plane (Fig. 5) in order to achieve the maximum gain and bandwidth, the resonance at given frequency is achieved with bandwidth 7.84 MHz, from 432.69 MHz to 440.53 MHz at -10 dB. The length of a patch was changed as well from 172.98 mm to 170 mm to help achieve minimal reflection coefficient $s_{11}$. Current reflection coefficient at given resonating frequency is -39.33 dB.

## *C. Radiation pattern*

Radiation pattern is the angular dependence of the power of radiated waves from an antenna. It is usually shown in a polar radiation pattern (either in 2D or 3D) showing the antenna lobes. The main lobe shows the wanted direction of antenna radiation. The other lobes (side lobes and back lobe) are unwanted radiation that must be suppressed or ideally completely eliminated [4].

The gain of the antenna is adjusted by increasing the ground plane or decreasing the length of the patch. The larger the ground plane, the more the gain increases up to the maximum value. By increasing the length of the ground plane, the side lobes are completely eliminated and the back lobe is suppressed, but by adding length to the ground plane will lead to inefficient usage of space and undesired overall length.

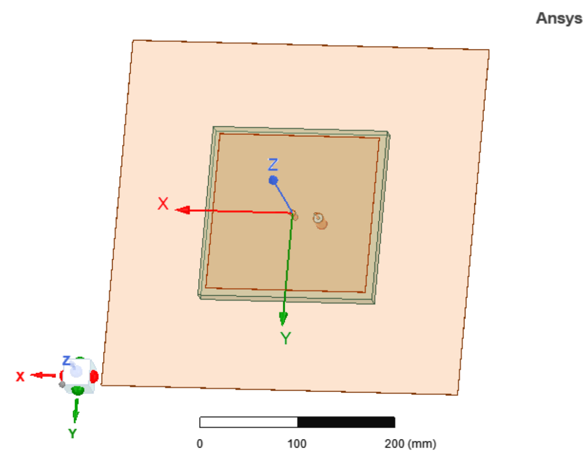



Fig. 5. Square patch antenna with expanded ground plane

The found length of the ground plane is 380 mm and the length of the patch is 170 mm. The maximum acceptable gain of the antenna is around 7.5 dB, displayed on 2D radiation pattern chart and 3D chart (Fig. 6 and Fig. 7). Unwanted side lobes are suppressed completely and the back lobe is suppressed as much as possible with regards to the acceptable maximal gain – main lobe.

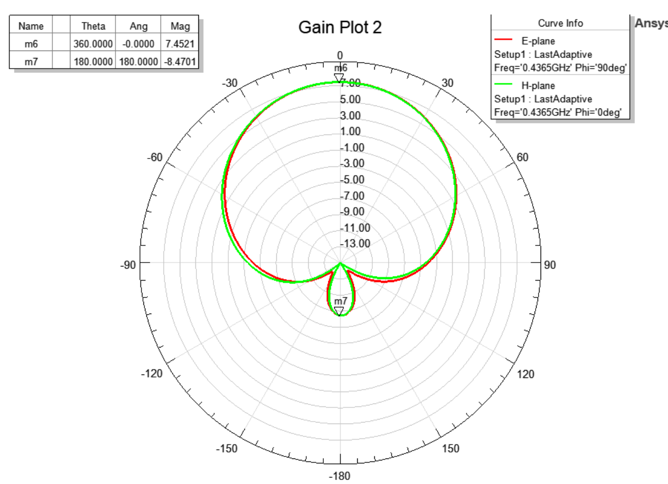



Fig. 6. Attuned radiation pattern of an antenna (E-plane and H-plane)

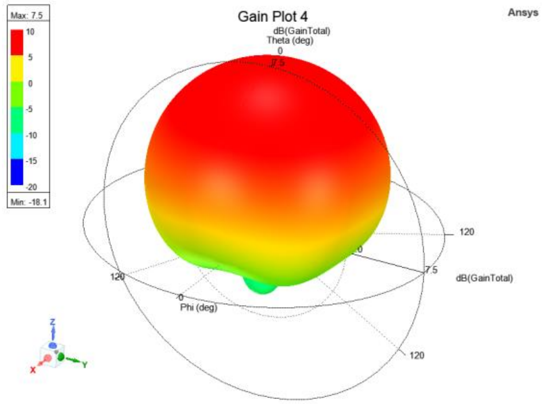



Fig. 7. 3D gain plot of the antenna

### *D. Design Extension*

The whole antenna design is possible to extend by adding another coaxial feed on x-axis with the same distance from the middle of the antenna as the first one. The results should be same or similar to a design with only one coaxial probe. If the reflection coefficient $s_{11}$ is the same as reflection coefficient of the newly created coaxial probe $s_{22}$ and the radiation pattern is similar to Fig. 6, then the whole antenna is designed correctly. The antenna can be used for reception of both polarizations.

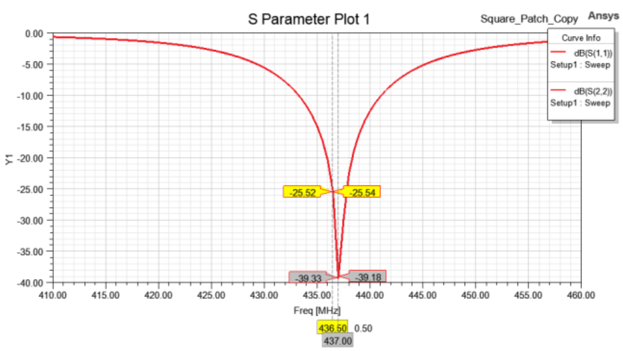



Fig. 8. Reflection coefficient $s_{11}$ and $s_{22}$ (2 coax. probes)

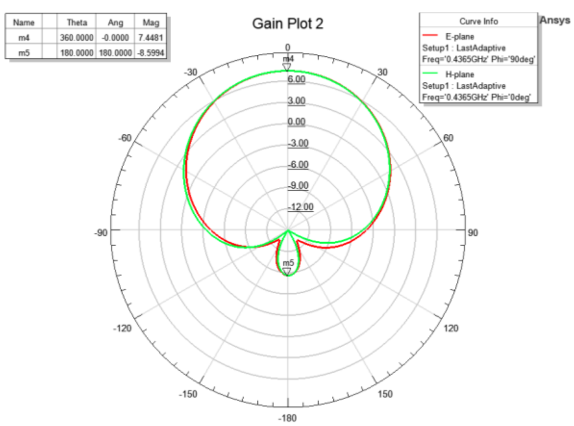



Fig. 9. Radiation pattern (2 coax. probes – E-plane and H-plane)

A small shift in resonant frequency occurred in Fig. 8, but the value of the $s_{11}$ is similar to the design with only one coaxial probe and reflection coefficient $s_{11}$ and $s_{22}$ are almost identical. The radiation pattern in Fig. 9 is almost identical with the one coaxial probe design as well. The same goes for 3D gain plot in Fig. 10. The whole antenna design can be considered as correctly designed.

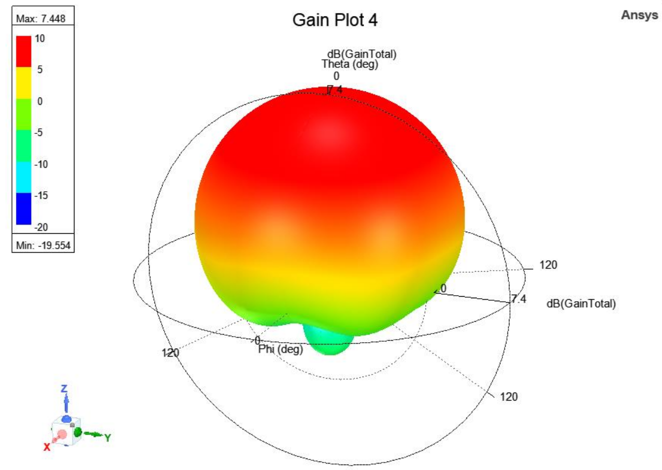



Fig. 10. 3D gain plot of the antenna (2 coax. probes)

## IV. Conclusion

In this paper, a square patch antenna was designed and simulated. Purpose for this particular antenna is to make it a part of an antenna array for satellite signal reception for UHF band. The designed antenna exhibits desired results, the antenna is well impedance matched with approximately 50 Ω and is resonating at desired frequency. The radiation pattern shows a decent maximal gain in the main lobe with a value of approximately 7.5 dB with usable and effective use of ground dimensions. The side lobes are completely suppressed and the back lobe is suppressed as much as possible. Overall, the antenna shows a good results and can be used for further development and design for antenna array.

## References


[1] *Mikropáskové antény* [online]. 39 [cit. 2023-12-31]. Available from: https://moodle.vut.cz/pluginfile.php/509804/mod_resource/content/1/P3s.pdf

[2] S-Parameters. *Antenna-Theory.com* [online]. 2015 [cit. 2023-12-26]. Available from: https://www.antenna-theory.com/definitions/sparameters.php#:~:text=S11%20represents%20how%20much%20power,antenna%20and%20nothing%20is%20radiated.

[3] Brno. In: Wikipedia: the free encyclopedia [online]. San Francisco (CA): Wikimedia Foundation, 2001- [cit. 2023-12-17]. Available from: https://en.wikipedia.org/wiki/Brno.

[4] Radiation Pattern. In: *Wikipedia: the free encyclopedia* [online]. San Francisco (CA): Wikimedia Foundation, 2001- [cit. 2023-12-27]. Available from: https://en.wikipedia.org/wiki/Radiation_pattern

# Optical ray propagation through turbulent underwater space

Bc. Jan Bartoň
*Department od Radioelectronics*
*Faculty of Electrical Engineering and Communication, Brno university of Technology*
Brno, Czechia
xbarto0c@vutbr.cz

***Abstract*— This paper is aimed at the description of an optical ray propagating through underwater turbulent environment. In the first part, the description of the properties of underwater turbulence and its impact on the optical beam is presented. In the second part of the paper, method based upon Snell's law is presented. In the third part, an approach using Ray tracing matrices is shown. The outcome of this thesis is a procedure for calculating the propagation of the optical beam through a turbulent underwater environment.**

***Keywords*— *Turbulence, underwater environment, optical beam, deflection, Snell's law, ABCD matrices, scintillation, lens.***

## I. Introduction

Turbulence is a crucial factor that needs to be considered when designing an underwater optical link. It can be modeled as a fluctuation of refraction index caused by variations in temperature, salinity and pressure along the propagation path which causes the optical ray to be deflected due to refraction and at the receiver only a fraction of the optical intensity is received by the receiving lens causing the received power to fluctuate. This phenomenon is called scintillation and degrades the system's performance [1].

Mathematically, turbulence is described by Navier Stoke's differential equations, whose exact solution is yet to be found and so they either need to be solved numerically or another approach needs to be taken when calculating it's influence on the propagation of the optical beam.

The turbid area can be divided into cells, that are considered to have a constant refractive index. Because the turbulence is by no means a stationary phenomenon, the refractive index of each cell varies with time and so do the values of respective refractive indexes. Statistical approach is once more utilised, where the value of a particular refractive index at a given time is estimated from its distribution function obtained from measurements. Depending on the dimensions of each cell, the cells can be approximated either as a volume with constant refractive index or as two lenses separated once again with a volume of constant refractive index.

## II. Snell's law

The simplest way to describe a turbulent area, is to consider it a homogenous volume with a constant refractive index $n_2$. In this case, beam deflection due to propagation through turbulence can be easily calculated using Snell's law (1)

$$n_1 \cdot \sin\alpha = n_2 \cdot \sin\beta\,, \tag{1}$$

where $n_1$ $[-]$ is the refractive index of the environment before the turbulent cell and $n_2$ $[-]$ is the refractive index of the turbulent cell itself.

Optical ray propagating from the environment with refractive index $n_1$ into the turbulent cell at an angle $\alpha$ is refracted at the interface of the cell to a lesser or a greater angle $\beta$ with respect to the primary optical axis. Whether the angle $\beta$ is greater or lesser than $\alpha$ depends on the ratio of refractive indexes $\frac{n_1}{n_2}$ (2). At the end of the turbulent cell, the optical ray is refracted once again at the second interface with respect to the secondary optical axis. If the two environments before and after the turbulent cell have the same refractive index, the angle at which the optical ray enters the turbulence is the same as the angle at which the optical ray leaves the turbulence. This principle can be seen in Fig. 1.

$$\beta \ldots \begin{cases} \beta > \alpha, \text{for } \dfrac{n_1}{n_2} > 1 \\ \beta = \alpha, \text{for } n_1 = n_2 \\ \beta < \alpha, \text{for } 0 < \dfrac{n_1}{n_2} < 1 \end{cases}. \tag{2}$$

Using Snell's law and trigonometric functions, formula for the deflection distance $y_a$ can be derived

$$y_a = L \cdot tg\left(\arcsin\left(\frac{n_1}{n_2}\sin(\alpha)\right)\right). \tag{3}$$

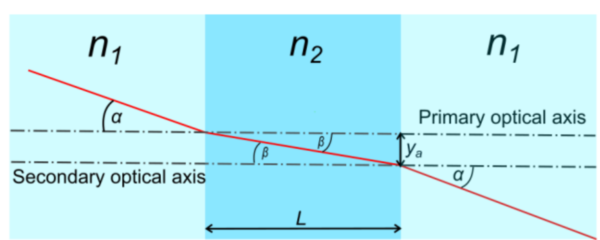



Fig. 1. Calculation of the turbulent beam deflection using Snell's law

In most practical situations, this approach cannot be used, as the turbulent cells also behave like lenses and this effect is neglected in this approach. It is however suitable for calculations, when the optical ray propagates through areas where laminar flow is present, which corresponds to the situation, where the optical link is oriented vertically, as the ocean's currents are usually laminar in this direction.

## III. RAY TRACING MATRICES

This method, also known as ABCD matrices, also relies on geometrical optics. It views turbulent cells as optical elements and links each cell to a certain matrix. Depending on how the turbulent cell is described, several types of matrices may be used [2].

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix} = \begin{pmatrix} 1 & t \\ 0 & 1 \end{pmatrix}. \tag{4}$$

Eq. (4) represents a matrix describing the same situation as is shown in Fig. 1, a medium with a constant refractive index, where $t$ is the thickness of the cell.

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ \frac{n_1 - n_2}{R_1 n_2} & \frac{n_1}{n_2} \end{pmatrix}. \tag{5}$$

Eq. (5) represents a matrix describing a lens, with a curvature radius $R_1$ and a refractive index $n_2$. It accounts for an incident wave passing through the interface between the lens and the medium before it with a refractive index $n_1$.

Those two matrices can be combined, to chain the optical elements together and hence easily include all the aforementioned effects of the turbulent cell together, as in Eq. (6)

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ \frac{n_2 - n_1}{R_2 n_1} & \frac{n_2}{n_1} \end{pmatrix} \cdot \begin{pmatrix} 1 & t \\ 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} 1 & 0 \\ \frac{n_1 - n_2}{R_1 n_2} & \frac{n_1}{n_2} \end{pmatrix}, \tag{6}$$

where the matrices represent, from left to right, lens at the input interface $n_1|n_2$, inside of the cell with a constant refractive index $n_2$ and thickness $t$ [mm], and the lens at the output interface $n_2|n_1$.

With the knowledge of the ABCD matrix, to determine the deflection of the optical ray, leaving the turbulent cell, it is enough to multiply ABCD matrix by a column vector, with a distance of the source from the primary optical axis at the first position and an angle at which the optical ray is passing the interface, which is called the deflection vector, just like in Eq. (7)

$$\begin{pmatrix} x_2 \\ \theta_2 \end{pmatrix} = \begin{pmatrix} A & B \\ C & D \end{pmatrix} \cdot \begin{pmatrix} x_1 \\ \theta_1 \end{pmatrix}, \tag{7}$$

where $x_1$ and $x_2$ are the distances from the optical axis of the system at the end of the turbulent cell ($x_2$) and at the beginning of the turbulent cell ($x_1$), $\theta_1$ and $\theta_2$ are the angles of divergence of the optical ray from the optical axis at the end of the turbulent cell ($\theta_2$) and at the beginning of the turbulent cell ($\theta_1$) and $A$, $B$, $C$ and $D$ are the parameters of the ABCD ray-tracing matrix. This situation is illustrated in Fig. 2.

As this method is based on paraxial approximation, the presence of paraxial rays is a requirement for using this method with suitable precision, which means that condition for the input angle $\theta_1$ [°] must be met [3], as is shown in Eq. (8)

$$\sin(\theta_1) \cong \theta_1. \tag{8}$$

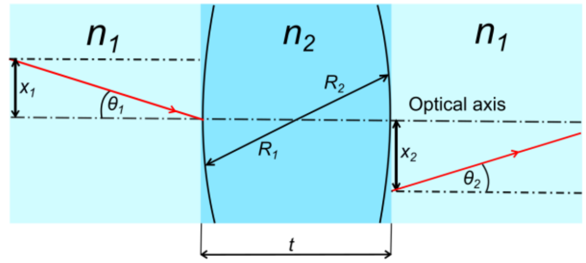



Fig. 2. Calculation of the turbulent beam deflection using Ray tracing matrices, in x-axis.

More optical elements could be chained together, simply by multiplying them, in a way that the matrix representing the first element in the chain must be at the rightmost side of the equation, because the order of operation in matrix multiplication is crucial. Turbulent underwater space can be divided into N turbulent cells, where each cell has its own refractive index $n_n$, thickness $t_n$, curvature radii $R_{n,1}$ [m], $R_{n,2}$ [m] and the individual cells are at the distance $l$ away from each other as can be seen in Fig. 3.

To calculate the beam deflection by the whole cascade of turbulent cells, it is first needed to determine the distance $x_{in}$, which is the initial deflection of the optical beam caused by the angle of divergence of the optical source. This distance can be calculated using trigonometric functions, as in Eq. (9)

$$x_{IN} = l_{TX} \cdot \mathrm{tg}(\theta_{IN}), \tag{9}$$

where $l_{TX}$ [m] is the distance between the transmitter and the first turbulent cell and $\theta_{IN}$ is the angle of divergence of the transmitter.

The total beam deflection after the passage through the whole turbulent space can be calculated according to the Eq. (10) [1]

$$\overrightarrow{X_{OUT}} = \boldsymbol{L_{RX}} \cdot \boldsymbol{T_n} \cdot \ldots \cdot \boldsymbol{L_2} \cdot \boldsymbol{T_2} \cdot \boldsymbol{L_1} \cdot \boldsymbol{T_1} \cdot \overrightarrow{X_{IN}} \tag{10}$$

where $\overrightarrow{X_{OUT}} = \begin{pmatrix} x_{out} \\ \theta_{out} \end{pmatrix}$ is the deflection vector at the output of the turbulent cell and $\overrightarrow{X_{IN}} = \begin{pmatrix} x_{in} \\ \theta_{in} \end{pmatrix}$ the deflection vector at the input of the turbulent cell, $L_n$ is the ABCD matrix of the inter-turbulent space, as in Eq. (4) and $T_n$ is the ABCD matrix of a turbulent cell, as in Eq. (6).

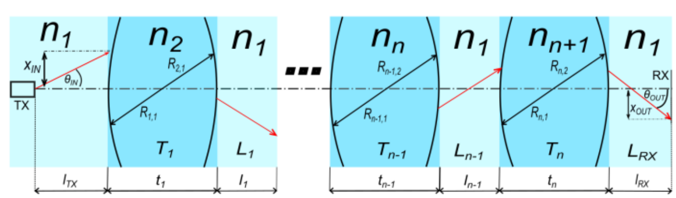



Fig. 3. Illustrative model of N turbulent cells chained together representing a turbulent space, in x-axis.

As water is an anisotropic medium, the refractive indexes of each turbulent cell are different in x and y axis, which means that to get the total spatial deflection of the optical beam, deflection vectors in both axes $\overrightarrow{x_{OUT}} = \begin{pmatrix} x_{out} \\ \theta_{out} \end{pmatrix}$ and $\overrightarrow{y_{OUT}} =$

$\begin{pmatrix} y_{out} \\ \Phi_{out} \end{pmatrix}$ need to be calculated separately and then their spatial coordinates need to be summed together to get the resulting deflection vector $\vec{v} = \begin{pmatrix} x_{out} \\ y_{out} \end{pmatrix}$ as is illustrated in Fig. 4. In this image, the beam is propagating from the black box (representing the transmitter) on the left side, through a turbulent space represented by four turbulent spherical cells, into the receiving lens, depicted by a black box on the right side. In this scenario, no power is received by the used receiver, as the beam is completely deflected by the turbulent space, the deflection vector is depicted by an orange arrow. The optical axis of the system is represented by the pink line connecting the transmitter and the receiver.

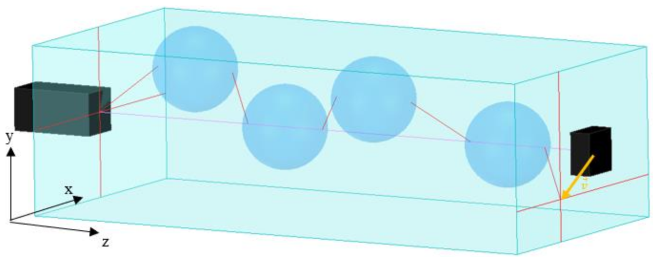



Fig. 4. Visualization of the beam deflection while propagating through a 3D turbulent space.

In applications, where duplex communication is desired, the question of reciprocity of the turbulent system comes into consideration. Unfortunately, the system of turbulent cells is not reciprocal, because the scattering properties of turbulence are asymmetric with respect to the reversal of the propagation direction. This is caused by the fact that turbulence is a random process, and the random phase shifts that are introduced by the turbulence are not reversible [4]. The reciprocity of the underwater turbulent space is further analysed in [5]. To analyse the deflection of the optical beam propagating in the opposite direction, one needs to reverse the order of the matrices in the matrix multiplication in Eq. (10), as in Eq. (11)

$$\overrightarrow{X_{OUT}} = \boldsymbol{L_{RX}} \cdot \boldsymbol{T_1} \cdot \ldots \cdot \boldsymbol{L_1} \cdot \boldsymbol{T_{n-1}} \cdot \boldsymbol{L_{n-1}} \cdot \boldsymbol{T_n} \cdot \overrightarrow{X_{IN}} . \qquad (11)$$

## IV. Calculating the Turbulent Attenuation

To calculate the power received by the receiving lens at the receiver it is necessary to first determine the distance by which the beam is deflected from the optical axis of the system, after propagating through the system of turbulent cells. The method used depends on the direction of communication. If the aim is to communicate vertically, for example between the ocean bed and a buoy on the surface of the sea, the water flow will be mostly laminar and so the easier method, based on Snell's law, Eq. (3), can be used or the matrix approach may be used as well, by representing the laminar flow by matrix from Eq. (4). If the communication is to take place in a horizontal direction, for example between two submarines, then the Ray-tracing approach has to be used, as the water flow will be mostly turbulent in this direction and the effect of the turbulent cells must be taken into account to obtain reliable results.

Once the resulting deflection vector is known, the received power by the receiving lens, if we account for a gaussian beam profile of the transmitted optical beam, can be calculated, as in Eq. (12)

$$P = \iint_D I_0 \left[ \frac{w_0}{w(z)} \right]^2 e^{-2\frac{x^2+y^2}{w^2(z)}} \, dD, \qquad (12)$$

where $I_0$ $[\mathrm{W \cdot m^{-2}}]$ is the amplitude of the beams' optical intensity, $w_0$ [m] is the waist radius of the beam, $w(z)$ [m] is the radius of the beam at the distance *z* from the transmitter, *x* and *y* are spatial coordinates of the deflection vector $\vec{v}$, see Fig. 4 and *D* is the surface area of the receiving lens. As the receiving lens is circular, the integral can be transformed into spherical coordinates, as in Eq. (13)

$$P = \int_0^{2\pi} \int_0^{a} A(z) \cdot r \cdot e^{-2\frac{(r\cdot\cos\theta - x_{out})^2 + (r\cdot\sin\theta - y_{out})^2}{w^2(z)}} \, dr d\theta \, , \qquad (13)$$

where *a* [m] is the radius of the receiving lens, $x_{out}$ and $y_{out}$ are the coordinates of the deflection vector $\vec{v}$ and $A(z)$ is a distance dependent coefficient, highlighted by blue colour in Eq. (12).

## Conclusion

The behaviour of an optical ray propagating through a turbulent underwater space has been investigated. Methods usable for calculation of the deflection of the optical beam by turbulence are presented, including the application of Snell's law and ABCD matrices. Procedure for calculating the received power by the receiving lens is shown. Attenuation by propagation through an underwater space is not accounted for here, nor is considered the power loss on the detector due to the non-optimal incident angle.

The turbulent space has a significant influence on the power received by the receiver. The factors that mainly decide how much the signal will be attenuated by the turbulent space are the length of the turbulent space, the angles in x and y axis at which the optical beam enters the turbulence and the surface area of the receiving lens. If the optical beam enters the turbulence along the optical axis of the system, it does not get deflected by the turbulence. To supress the turbulent attenuation, bigger receiving lens can be used.

## References


[1] LANZAGORTA, Marco. Underwater communications. Synthesis lectures on Communications. 2013. ISBN 978-1-60845-844-8.

[2] KOVALOVA, A.; HUDCOVA, L. a ROKA, R. Optical Ray Transfer Matrix Model of the Turbulent Cells Cascade. Online. Radioengineering. 2022, roč. 31, č. 4, s. 520-526. ISSN 1210-2512. Dostupné z: https://doi.org/10.13164/re.2022.0520. [cit. 2023-12-21].

[3] GERRARD, A. a Burch, J. M. Introduction to Matrix Methods in Optics- Dover Publication. 1975. ISBN 0-486-68044-4.

[4] KOVALOVA, Alzbeta a HUDCOVA, Lucie. Reciprocity of the Turbulent Cascade Model. Online. 2022 32nd International Conference Radioelektronika (RADIOELEKTRONIKA). 2022, s. 1-4. ISBN 978-1-7281-8686-3. Dostupné z: https://doi.org/10.1109/RADIOELEKTRONIKA54537.2022.9764902. [cit. 2023-12-22].

[5] GUO, Yujian; TRICHILI, Abderrahmen; ALKHAZRAGI, Omar; ASHRY, Islam; NG, Tien Khee et al. On the Reciprocity of Underwater Turbulent Channels. Online. IEEE Photonics Journal. 2019, roč. 11, č. 2, s. 1-9. ISSN 1943-0655. Dostupné z: https://doi.org/10.1109/JPHOT.2019.2898094. [cit. 2023-12-22].

# Implementation of LTE Cat-M Technology for Smart Measurement Scenarios

Viacheslav Tamentsev
*Department of Communications and Informatics*
*Brno University of Technology*
Brno, Czech Republic
236356@vutbr.cz

***Abstract*—The requirements introduced in Directive 2019/944 of the European Parliament for smart metering scenarios, which differ from the ones defined in ITU-R M.2410, accelerated the integration of cellular technologies for the Internet of Things (IoT). With the release of cellular IoT (CIoT) technologies in the Czech Republic, recently LTE Cat-M, the focus shifts towards evaluating the performance in massive scenarios where thousands of devices are connected to one base station. This paper examines the feasibility of implementing LTE Cat-M communication technology in Network Simulator 3 (NS-3) to cover massive Machine-Type Communication (mMTC). The paper analyzes the capabilities of cellular modules within the selected simulation tool Networks Simulator 3 (NS-3), including the LENA module. First, simulation scenarios are created to validate and benchmark the performance of legacy LTE/LTE-A networks with attention to (mMTC) scenarios. The key output is the determination that the 4G networks are unsuitable for these scenarios. Further, based on the obtained data, the initial discussion on the requested implementation to support the LTE Cat-M in the NS-3 is provided, pointing out the exact changes associated with the physical layer.**

***Index Terms*—Internet of Things, Low Power Wide Area, Sigfox, LoRaWAN, Narrowband IoT, LTE, LTE Cat-M, eMTC, Network Simulator 3, LENA, 5G-LENA, LENA-NB.**

## I. Introduction

Today's world of telecommunications is constantly evolving, and Long-Term Evolution for Machine-Type Communications (LTE Cat-M) technology is gaining attention as an effective means of enabling Internet of Things (IoT) communications. This technology, an evolution of the LTE standard, is specifically designed to support extensive machine-to-machine communication (mMTC) scenarios, making it ideal for intelligent analysis and monitoring scenarios [1].

With the development of IoT, the integration of all sectors of telecommunications is reaching new heights, enabling all aspects of our lives to be connected virtually. IoT is the virtualization of real-world objects to digitally process and exchange data between devices. The application of IoT in smart homes, including monitoring energy consumption and public spending, has great potential to reduce costs and improve resource management. Low-power Wide-Area Network (LPWAN) technologies, such as Long Range Wide Area Network (LoRaWAN), Sigfox, NB-IoT, and LTE-M, are promising means of wireless access to global networks that can further reduce the cost of services and increase their availability [2].

### A. Main Contributions

In this paper, the focus is to outline the major LPWA standards,, such as LoRaWAN and Sigfox, as well as LTE Cat-M and NB-IoT. Specifically, the attention is on the Cellular IoT (CIoT) technologies that are supposed to replace legacy cellular representatives, e.g., LTE, in the mMTC scenarios. This assumption is supported by the obtained simulation results, where the capabilities of the LTE cellular network (following the 3GPP Release 10 definition) were tested. The key focus of this paper is to verify the technical parameters of LTE Cat-M by analyzing classical LTE/LTE-A networks.

By briefly describing my research, we can highlight several key points:

- LPWA analysis of unlicensed and licensed spectrum technologies.
- Parsing available LTE modules and other network technologies (LENA, LENA-NB) in the NS-3 simulator.
- Performance evaluation of LTE technology for mMTC scenarios includes analysis of peak speeds, packet loss, and the maximum number of simultaneously connected devices to one base station.

The rest of this paper is organized as follows: Section II provides a comparative analysis of LPWA technologies together with LTE Cat-M, describing the main characteristics of the selected technology and their possible drawbacks. After describing the main differences between wireless access technologies, the focus is on LTE-Cat-M technology. Section III gives an overview of the available modules, their possible integration into the NS-3 simulator, and the results of modeling the LENA module for mMTC scenarios. The last section IV concludes, where we summarize that today's LTE technologies are not able to provide all the requirements for massive communications scenarios, as our tests in the NS-3 simulator show. The results of this research will be further utilized in the thesis.

## II. Comparative analysis of LPWA data exchange technologies

5G technology has standardized the requirements for IoT, including advanced mobile, critical, and mMTC, providing connectivity for a large number of devices with small data transfers. LPWA networks, including NB-IoT and LTE-M, as

TABLE I
COMPARISON OF THE TECHNICAL CHARACTERISTICS OF LPWA NETWORKS [3], [4]

| Technical specifications | SigFox | NB-IoT | LoRaWAN | LTE-M |
|---|---|---|---|---|
| Scope of licences | Unlicensed | Licensed | Unlicensed | Licensed |
| Throughput | 100 kbps | 1-200 kbps | 0,3-50 kbps | 1 Mb/s |
| Max.message size UP | 12 B | 1280 B | 243 B | 1280 B |
| Max.message size DL | 8 B | 1280 B | 243 B | 1280 B |
| Battery life, year | 10 (At consumption $<$ 2 uA) | 10 (At consumption $<$ 2 uA) | 8 (At consumption $<$ 3 uA) | 10 (At consumption $<$ 8 uA) |

well as LoRaWAN and Sigfox, are most suitable for such requirements, providing wide coverage and long battery life [3].

### *A. SigFox, LoRaWAN and NB-IoT*

SigFox, LoRaWAN, and NB-IoT technologies are different approaches to providing low-power, long-range connectivity for IoT. SigFox is based on a narrow bandwidth (868.8 MHz) and data transmission in unlicensed bands, enabling communication up to 10 kilometers. However, it is limited by the use of licensed sensors, a maximum message length (12 bytes), and a limit on the number of messages in the network (140 per day) [3].

LoRaWAN, on the other hand, uses broadband pulses and linear-frequency modulation for data transmission and operates in a variety of license-free frequency bands (e.g., 868 MHz, 915 MHz, and 433 MHz). It provides communication over distances of up to 5 kilometers in urban environments and up to 50 kilometers in open terrain, with up to 254 bytes of transmitted data [3].

NB-IoT integrates data from cellular users and IoT devices using the existing 4G network infrastructure. It provides data rates up to 200 kbps and requires a SIM card to operate, using different modulations for data transmission such as FDMA and OFDMA [3].

All three technologies have low power consumption and can provide long device lifetimes (up to 10 years), making them ideal for IoT applications. However, the choice between them depends on specific project requirements such as communication range, data rate, and infrastructure availability [3], [4].

### *B. LTE-M*

LTE-M technology, introduced by 3GPP with the eMTC (extended machine-to-machine communication type) protocol, has a wide frequency bandwidth of up to 1.4 MHz and supports voice messaging. Designed for the IoT, it strikes a balance between data rates (up to 1 Mbps) and efficient power consumption. LTE-M is a simplified version of the LTE standard optimized for long device battery life [5].

Unlike other technologies, LTE-M targets IoT applications with high data traffic requirements while providing low power consumption and long-distance data transmission capability. It utilizes data transmission schemes similar to LTE, using Orthogonal Frequency-Division Multiple Access (OFDMA) for

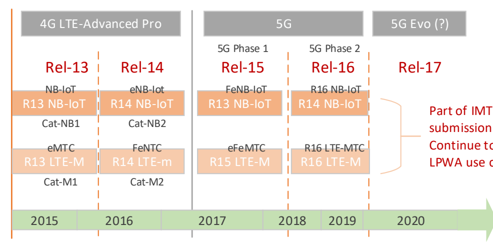



Fig. 1. Development of LTE Cat-M and NB-IoT technologies in 3GPP version [12].

inbound connections and Single-Carrier Frequency-Division Multiple Access (SC-FDMA) for outbound connections, allowing them to coexist in the same LTE cell [4].

TABLE II
MAIN PARAMETERS FOR LTE CAT-M TECHNOLOGY [2], [5], [6], [11]

| 3GPP Release | Features |
|---|---|
| **Release 13** | Bandwidth is narrowband (1.4 MHz), coverage extension (CE) using A/B modes. |
| **Release 14** | Support for higher data rates, multicast based on MBMS structure and improved positioning with PRS customizability. |
| **Release 15** | Focused on improving latency, spectral efficiency, and energy efficiency. Spectral efficiency was increased by using higher modulation and detailed resource allocation. New transmission power classes were added, and radio resource management and load control were improved. |
| **Release 16** | It brought improvements for LTE-MTC (or eMTC), such as better early data transmission and efficient resource utilization. Spectrum coexistence with NR and the ability to connect LTE-MTC devices for the 5G network were also important. |
| **Release 17** | Introduced support for NR or RedCap devices to address scenarios where LTE proves insufficient. RedCap includes a wide frequency range, energy efficiency, and air interface compatibility to optimize data transmission. |

Release 12 was an important moment for the IoT community as it facilitated research into methods to reduce the cost of devices and make them attractive to low-level MTS applications previously running on GSM/GPRS networks. This release allowed engineers to begin research into reducing peak speed and bandwidth, as well as converting the module to half-duplex operation [2].

Further versions of LTE-M introduced by 3GPP continue to focus on improving key parameters such as battery life, message latency, and overall performance. These improvements are aimed at supporting massive IoT applications and are reflected in Table II. The next step in LTE-M development includes an upgrade to Reduced Capability (RedCap) devices, which are expanding their functionality to become compatible with 5G networks. This will allow devices to utilize faster and more efficient networks for data transmission, opening up new opportunities for the development and deployment of IoT applications [7].

LTE Cat-M is based on the LTE/LTE Advanced network infrastructure and provides advanced connectivity for IoT devices. LTE Cat-M architecture extends LTE functionality [6].

The key components of LTE technology, including the network core and receiving stations, are described below [6]:

- Terminal Equipment (UE): A portable device that provides connectivity to a wireless network.
- Base Station (eUTRAN): A Radio Access Network (RAN) transmitter that provides radio access and control for the terminal equipment (UE).
- Evolved Packet Core (EPC): The core of LTE technology.

## III. Network Simulator 3

NS-3 is a network simulator with a discrete-event architecture that is mainly used for educational and research purposes. By wrapping NS-3 in Python, programs in that language can import the ”ns3” module using APIs. There are several basic modules in NS3 for simulation. We will focus on the LENA modules that provide support for mMTC [8].

### A. LENA and LENA-NB

The LENA module extends the NS-3 Network Simulator for more realistic modeling of LTE networks, including load balancing and Multi-RAT networks. It consists of LTE and EPC models that provide support for LTE and core network protocols and entities. The LENA module is used to develop and test LTE network management algorithms, as well as for research in load balancing and Multi-RAT scenarios. This tool is valuable for developers and researchers in LTE and its evolution [9].

LENA-NB includes Cellular-IoT C-plane optimization, Early Data Transfer feature, Radio Resource Control (RRC) resumption procedure, a new cross-subframe scheduler, adaptive modulation and coding, and a detailed energy state machine. This tool allows comparing the performance of different NB-IoT transmission modes, revealing the benefits of early data transmission [10].

### B. Communication scenario for mMTC

The LENA module was chosen for the basic communication of the UE with the eNodeB, which provides all the necessary communication requirements for LTE networks in the NS-3 simulator. As an implementation of the LENA module in the NS-3 simulator, the lte-full.cc scenario was chosen, which provides standard communication between devices with an EPC core. Its initial parameters include four devices and one eNodeB station. Also included in its network topology are a remote host and a Packet Data Network Gateway (PGW) gateway.

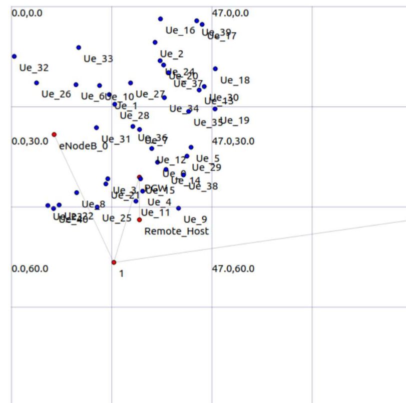



Fig. 2. Network topology for 40 UEs in NetAnim software.

In this scenario, the data flow is generated by the Bulk application on Transmission Control Protocol (TCP), which reduces the final data rate. Table III defines the basic parameters for the entire scenario, with a gradual increase in devices in the network. These parameters are not re-divided in the scenario anymore and are set by default.

TABLE III
Basic scenario parameters.

| Parameter | Value | | | | |
|---|---|---|---|---|---|
| Number of end devices (UE) | 4 | 14 | 21 | 32 | 40 |
| Number of base stations (eNodeB) | 1 | 1 | 1 | 1 | 1 |
| Number of remote users | 1 | 1 | 1 | 1 | 1 |
| Simulation time | 10 s | 10 s | 10 s | 10 s | 10 s |
| SRS signal periodicity | 40 ms | 40 ms | 40 ms | 40 ms | 40 ms |
| UE deployment area | 5.5 km^2 | 8.4 km^2 | 10 km^2 | 10 km^2 | 10 km^2 |
| Interval between packets | 100 ms | 100 ms | 100 ms | 100 ms | 100 ms |
| Maximum UE transmit power | 23 dBm | 23 dBm | 23 dBm | 23 dBm | 23 dBm |
| Delay UL | 21 ms | 60 ms | 80 ms | 120 ms | 140 ms |
| Average UL speed | 1300kbps | 840kbps | 570kbps | 380kbps | 300kbps |
| Bandwidth | 5 MHz | 5 MHz | 5 MHz | 5 MHz | 5 MHz |
| Transport layer protocol used | TCP | TCP | TCP | TCP | TCP |

As the number of devices in the network increases to 40, the speed will drop and the latency will increase, while maintaining a stable up-and-down channel for data transfer between devices. Fig. 2 shows the network topology for 40 devices in an area of 10 square kilometers. The location of these UEs relative to the base station is random and is determined by the RandomDiscPositionAllocator function in this scenario.

The problem with this scenario is that most of the network parameters are not defined in the scenario itself and are chosen as default parameters. So for the base scenario, the Sounding Reference Signal (SRS) parameter was set to 40. This means that when the number of devices in the network starts to increase, we will face the first network problems. Some of the packets will fall off at once, and the network will be overloaded before the whole data stream can be sent up or down the line.

This SRS parameter itself is a reference signal that is used for eNodeB in LTE to estimate the uplink channel quality for each subsector of the frequency domain. This signal helps to determine which part of the system bandwidth has the best channel quality for a particular piece of user equipment (UE) at a particular time.

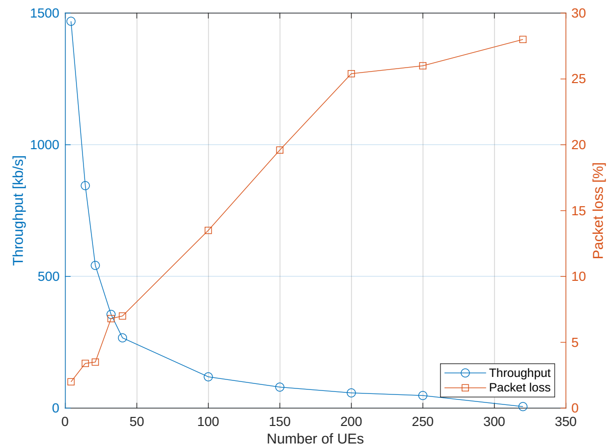



Fig. 3. Throughput and packet loss in relation to the number of UEs

The UE transmits SRS according to the configuration specified by signaling messages such as System Information Block 2 (SIB2), RRC Connection Setup, RRC Connection Reconfiguration, and others. The SRS transmission period can vary depending on the parameters set and can be configured to transmit every two subframes up to every 32 frames (320 subframes). It is not necessary to include it in the configuration of each UE, but in this way, it is possible to provide better resource allocation in the network by increasing the number of devices.

It is also necessary to extend the scenario to 320 devices with the SRS parameter, as LTE Cat-M technology assumes a much larger number of devices in the network at low bandwidth, as shown in Fig. 3. As a modification, the LENA module main file has been modified to control the SRS parameter. In the LENA module, the SRS states are defined in the lte-enb-rrc.cc file for eNodeB. Therefore, it was necessary to plug in this library and set the parameter to the desired value, as shown in Listing 1.

```
Config::SetDefault("ns3::LteEnbRrc::SrsPeriodicity",
    UintegerValue (320));
```

Listing 1. Module required to change the SRS parameter

In this scenario, the LENA module based on LTE Release 10 technology does not meet the network requirements for data transmission, latency, and packet loss. These data are directly related to the number of devices in the network in this scenario. The throughput in the network will decrease with each new device, which corresponds to the real situation in mMTC scenarios where a base station serves from one to several hundred UEs. Moreover, the biggest drop occurs in the first 40 devices. A further drop will occur more linearly.

## IV. Conclusions

The results of the paper indicate the limitations of LTE technology in mMTC scenarios, especially when using the 3GPP Rel. 10 standard. When a large number of devices are connected to a single base station, the data rate decreases, latency increases, and the level of data loss increases due to retransmissions. Even with a reference scenario of 40 devices, the system reaches its limits, which emphasizes the need for more efficient solutions for mMTC scenarios.

The scenario was later extended by increasing the SRS, raising the total number of devices, but introducing even higher packet losses and low throughput. A possible solution is the implementation of LTE Cat-M technology, which provides higher data rates and efficient utilization of LTE spectrum bandwidth. The expected data rates in the hundreds of Kbps while utilizing narrower bandwidth make LTE Cat-M an attractive option for IoT and Industry 4.0 scenarios.

Therefore, LTE Cat-M technology will be implemented in future thesis work. This will require modifications to the current LENA module. And also to implement the PSM and eDRX energy-saving features that are presented in the LENA-NB module. These changes will aim to integrate power-saving features into the LTE-M standard to ensure long-term battery life for IoT devices. LTE-M MIB and SIB message controllers need to be implemented in accordance with the Release 13 standard, and then appropriate timers need to be added. And make changes to the LteUeRrc and LteEnbRrc classes to handle new states and control timers.

These changes in NS3 of the LENA module will help achieve the necessary requirements for communication in mMTC scenarios, according to LTE Cat-M technology.

## References


[1] K. Benson et al., "SCALE: Safe community awareness and alerting leveraging the internet of things," in IEEE Communications Magazine, vol. 53, no. 12, pp. 27–34, Dec. 2015.
[2] O. Liberg, M. Sundberg, E. Wang, J. Bergman and J. Sachs, "Cellular Internet of things: technologies, standards, and performance," Academic Press, 2017.
[3] K. Mekki, E. Bajic, F. Chaxel and F. Meyer, "Overview of Cellular LPWAN Technologies for IoT Deployment: Sigfox, LoRaWAN, and NB-IoT," 2018 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops), Athens, Greece, pp. 197–202, 2018.
[4] S. Chalapati, "Comparison of LPWA technologies and realizable use cases," Fall Tech. Forum, Vol. 2018, pp. 1–12, 2018.
[5] 3GPP, "Study on provision of low-cost machine-type communications (MTC) user equipments (UEs) based on LTE." 3GPP Sophia Antipolis, France, 2013
[6] B. S. Chaudhari and M. Zennaro, eds., "LPWAN technologies for IoT and M2M applications," Academic Press, 2020.
[7] S. N. K. Veedu, M. Mozaffari, A. Höglund, S. Thangarasa, E. A. Yavuz and J. Bergman, "RedCap: Expanding the 5G Device Ecosystem - Ericsson," WHITE PAPER, 2023.
[8] NS-3 project, "NS-3 Manual Release NS-3.41," 2024.
[9] NS-3 project, "19.1. Design Documentation," 2024.
[10] P. Jörke, T. Gebauer and C. Wietfeld, "From LENA to LENA-NB: Implementation and Performance Evaluation of NB-IoT and Early Data Transmission in NS-3," In Proceedings of the 2022 Workshop on ns-3, pp. 73–80, 2022.
[11] 5G Americas, "L3GPP Releases 16, 17 & Beyond," WHITE PAPER, 2021.
[12] S. Lu, "5G Standards Overview & 3GPP Tech Evolution Trends," 2018.

# Application of Optimization Algorithms to Support Penetration Testing

1st Dominik Zacek
*Department of Telecommunications*
*FEEC, Brno University of Technology*
Brno, Czechia
xzacek08@vut.cz

2nd Willi Lazarov
*Department of Telecommunications*
*FEEC, Brno University of Technology*
Brno, Czechia
lazarov@vut.cz

***Abstract*—This paper presents a novel approach to support the pre-engagement phase of penetration testing, where testing tasks are assigned to penetration testers based on their knowledge and experience to ensure the most appropriate selection. To apply and verify our approach, we developed an automated tool that uses optimization algorithms for the task assignment process. Experimental testing shows that the application of algorithms based on optimization problems in the first phase of penetration testing could be a way to increase its effectiveness.**

***Index Terms*—penetration testing, cybersecurity, optimization, task assignment, linear programming**

## I. Introduction

In the rapidly evolving cybersecurity environment, a preventive technique to search for vulnerabilities in order to eliminate them, called penetration testing, has proven to be a good strategy for ensuring the security of information and communication technologies (ICT) [1]. Through penetration testing, organizations can proactively identify vulnerabilities and, based on all findings, significantly improve the security of their systems and applications. However, the effectiveness of these penetration tests depends on the assignment of tasks to penetration testers in relation to their knowledge and experience [2]. The penetration tester must be assigned a task that best matches her/his skills, while it is desirable that the tasks be assigned as evenly as possible among the testers.

The main contribution of this paper is the design and implementation of an advanced tool to streamline the mentioned pre-engagement phase of penetration testing during which tasks are assigned to individual testers. Using optimization algorithms, the developed tool automatically assigns penetration testers to tasks and optimizes the selection based on the testers' knowledge and the requirements of each task.

This paper is organized into four sections. Section II describes the current state-of-the-art in the domain of optimization problems in relation to penetration testing. Our approach is then described in Section III followed with comparison of selected methods and experimental testing. The last Section IV concludes the achieved results and discusses future work.

The research described in this paper was financially supported by the Ministry of the Interior of the Czech Republic, project No. VK01030019.

## II. Background and Related Work

The assignment problem in general can be viewed as a linear programming (optimization) problem. Linear programming is a mathematical model containing a linear function that ought to be minimalized or maximalized and several linear constraints on variables in the function. The downside of this general approach is that the model and constraints grow fast as the numbers of jobs and tasks grow. Because of that, a more narrow-case Hungarian algorithm, also known as the Kuhn-Munkres method, has been developed. The method works with balanced assignment problems, meaning that the number of jobs and tasks is equal. Thus, each task can be assigned to a single worker. Unbalanced assignment problems can be easily converted into a balanced assignment problem. The method also has other restrictions such that there cannot be a mandatory or forbidden assignment of tasks and workers [3]. Thanks to these restrictions, the method has complexity $\phi(n^3)$ and is computationally very efficient and fast.

Another general approach to solving assignment problems is to use machine learning (ML) methods. The most relevant methods related to ML are neural networks and Q-learning. Both rely on having a utility function that can be constructed using metrics that measure assignment uniformity and quality. These methods do not have restrictions like the Hungarian method and can be very fast when they are already trained. On the contrary, the solution, as opposed to linear programming, does not has to be a global optimum. This downfall has been studied and overcome. Using a dual neural network scheme based on a dynamic model has been proven to converge to a globally optimal solution [4].

In the context of this paper, other researchers have already examined the usage of optimization problems and artificial intelligence (AI) in the penetration testing domain. To support penetration testing using AI, some papers focused on mapping attack trees [5], finding an optimal attack policy [6], identifying attack graphs [7], and discovering exfiltration paths [8], all using the reinforcement learning method. However, most of these papers have explored the use of AI for penetration testing more on a theoretical level, and therefore more in-depth applied research is required, especially given that penetration testing is mainly a practical technique.

## III. Proposal of an Innovative Approach

Based on the state-of-the-art and the requirements of the pre-engagement phase, optimization problems were used as possible solutions to support this phase of penetration testing. Therefore, we propose an innovative approach that makes the preparation phase of penetration testing more efficient by applying optimization algorithms in order to minimize the overall penetration testing time. By the nature of the assignment problem we address, our approach is limited to penetration testing that is performed by two or more testers. Our solution automatically assigns penetration testers to subtasks based on their knowledge and experience to make the selection as appropriate as possible.

### *A. Comparison of Selected Methods*

To choose the best algorithm for our approach, 6 algorithms were compared on a generated dataset containing 50 penetration testers, 1000 tasks, and 10 task types (abilities). The comparison process is visualized in Figure 1 and the data format is as follows:

- Testers' skills are stored as type:skill level pairs and the total number of these pairs is chosen randomly. The skill type is given by an arbitrary number and, in practice, can be used as an identifier of different types of penetration testing skills and knowledge. The skill level is given as a decimal number from 0.1 to 1, which is randomly selected during generation within this range and then rounded to one decimal place.
- The tasks contain the required skills properties. The required skills property is an array type with numbers expressing the skill required by the task. The required skills are randomly chosen from 1 to the total number of skill types.

For comparison, we measured common skills metrics, missing skills, excess skills, uneven distribution, and duration. A comparison of the selected algorithms is given in Table I. In addition, the average duration of the assignment algorithm has been added to the table. The average was always calculated from 10 runs of the algorithm. The results are described in the following sections.

*1) Common Skills:* Table I shows that *Assignment based on common abilities* and *Linear programming (unbound)* have the best results in the common skills metric. However, neither of these methods works at all with the load of the tester. Both methods only maximize a given metric while adhering to the single rule of not assigning multiple testers to a single task. In practice, these methods would not work well because the fewest capable testers would be assigned to all tasks. Furthermore, both methods found essentially the same maximum. This phenomenon can be explained by the fact that the *Linear programming (unbound)* algorithm, due to the absence of any rule on uniformity of distribution, was actually looking for the same thing as the *Assignment based on common abilities* algorithm, just with a different procedure. Both algorithms are able to find the global maximum (of a single metric) in the dataset and therefore came up with the same result.

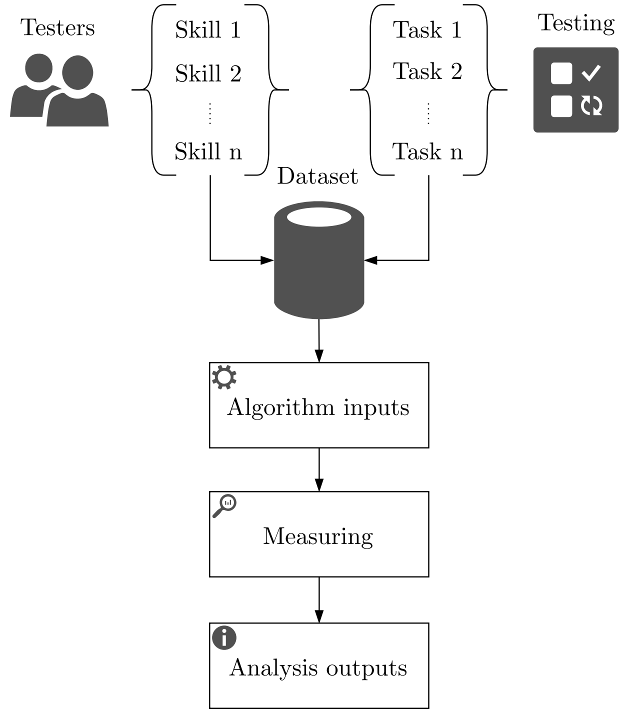



Fig. 1. Process of analyzing optimization algorithms

*2) Missing and Excess Skills:* The *Missing/Excess skills* algorithm has the best result (minimum value) in the *Assignment based on missing/excess abilities* column. This is because it has maximized this metric, without other constraints. The advantage over the *Common Skills* algorithm is that easier tasks are more likely to be assigned to less skilled penetration testers, making the overall assignment more balanced.

*3) Uneven Distribution:* The *Linear programming* and *Hungarian method* algorithms reached zero. Thus, each tester was assigned an average number of tasks. This result is given by the conditions of both algorithms. The algorithm *Assignment based on missing/excess skills* performed relatively well compared to the algorithms *Linear programming (unbound)* and *Assignment based on common abilities*. This is due to the nature of the metric used, which does not attempt to assign simple tasks to highly skilled penetration testers.

*4) Duration:* The *Hungarian method* algorithm achieved the shortest duration. This is due to the highly optimized `linear_sum_assignment` method from the `scipy` library, written in the C programming language. The *Random assignment* is slower than the *Hungarian method* due to the `for` loop, which is slower in Python than in C. The *Linear programming (unbound)* algorithm, also from the `scipy` library, is slower than the *Hungarian method*. The slowdown is due to the solution of a more general and therefore more complex linear programming problem. The *Linear Programming* algorithm is several times slower than the other algorithms due to additional constraint equations to determine the maximum number of assignments for a single tester.

TABLE I
COMPARISON OF SELECTED ALGORITHMS

| | Metrics | | | | |
|---|---|---|---|---|---|
| **Algorithm** | **Common skills [%]** | **Missing skills [%]** | **Excess skills [%]** | **Uneven distribution** | **Duration [s]** |
| Linear programming | 49,0340 | 50,9659 | 29,4599 | 0,0000 | 67,9903 |
| Linear programming (unbound) | 78,7158 | 21,2841 | 36,6550 | 75,6122 | 3,3667 |
| Hungarian method | 48,4780 | 51,5219 | 35,7939 | 0,0000 | 0,0060 |
| Assignment based on common abilities | 78,7158 | 21,2841 | 36,5689 | 75,4081 | 18,1674 |
| Assignment based on missing/excess skills | 60,2803 | 39,7196 | 25,2100 | 39,2857 | 18,2464 |
| Random assignment | 32,7766 | 67,2233 | 52,5266 | 9,0918 | 0,3633 |

Based on the metrics in Table I and the properties of the compared algorithms, we find the linear programming algorithm to be the best for several reasons. First, the algorithm is the most customizable. This means that the assignment constraints can be chosen as needed. For example, each penetration tester will be assigned an average number of tasks, or some penetration testers can have a limited number of tasks (e.g., part-time workers). Second, the assignment found under given constraints is guaranteed to be the globally optimal solution. Lastly, although the runtime is the longest as seen in Table I, it is still very reasonable (a little over a minute) for a mid-sized dataset.

### B. Application of Linnear Programming

To solve the linear programming problem of assigning tasks to penetration testers, it is necessary to first formulate it. The general equation for the assignment problem [4] without constraints can be formulated as follows:

$$\min \sum_{t \in T} \sum_{w \in W} c_{t,w} \cdot x_{t,w},$$

where $T$ is the set of tasks, $W$ is the set of workers, $c_{t,w}$ is the cost of assigning worker $w$ to task $t$, $x_{t,w}$ is the binary value ($x_{t,w} \in \{0, 1\}$) of whether task $t$ should be assigned to given worker $w$. Constraint to ensure that all tasks are assigned to one worker:

$$\forall t \in T, \sum_{w \in W} x_{t,w} = 1.$$

Optionally, it is also possible to construct a constraint on the number of tasks assigned to a single worker $w$:

$$\sum_{t \in T} x_{t,w} = n,$$

where $n$ is the desired number of tasks assigned to worker $w$. Lastly, as for every assignment problem solved using linear programming, it is necessary to set binary boundaries for all $x$ variables:

$$\forall t \in T, \forall w \in W, \quad 0 \leq x_{t,w} \leq 1.$$

By solving the equation as a linear programming problem we obtain all the values of $x$. If and only if $x_{t,w} = 1$ then task $t$ should be assigned to the worker $w$.

### C. Case Study

To validate our approach using linear programming for task assignment in the pre-engagement phase of penetration testing, we conducted a case study on a small dataset of penetration testing tasks. The scenario for experimental testing represents a real-life situation, where there are penetration testers with a different skill set. The first penetration tester should represent someone experienced with strong and wide knowledge. The second penetration tester is someone who is also experienced but is mainly focused on specific tasks. And the last one is a junior penetration tester with less experience.

For this scenario, we selected 12 tasks from the Application Security Verification Standard (ASVS) [9] and listed them in Table II, together with their required skills and the tester assigned after running the algorithm on this dataset (see Appendix A for detailed steps). Selected tasks require one or more of the following skills: *S1* – Input testing; *S2* – Parametrization; *S3* – Encoding; *S4* – Requests; *S5* – Validation; *S6* – Sanitation; *S7* – SQL injection.

These skills were associated with 3 penetration testers to whom the tasks will be assigned. Their skills and specific experience level are as follows: Tester 1 – *S1* (70%), *S2* (60%), *S3* (80%), *S4* (100%); Tester 2 – *S1* (100%), *S2* (80%), *S5* (80%), *S6* (90%); Tester 3 – *S1* (50%), *S3* (40%), *S4* (30%), *S5* (20%).

TABLE II
TASK ASSIGNMENT AMONG TESTERS

| No. | Task | Required skills | Assigned tester |
|---|---|---|---|
| Task 1 | ASVS-5-1-1 | S01, S02, S04 | Tester 1 |
| Task 2 | ASVS-5-1-2 | S01, S02, S04, S05, S06 | Tester 1 |
| Task 3 | ASVS-5-1-3 | S01, S05 | Tester 3 |
| Task 4 | ASVS-5-1-4 | S01, S05 | Tester 3 |
| Task 5 | ASVS-5-1-5 | S01, S04 | Tester 1 |
| Task 6 | ASVS-5-2-1 | S01, S06 | Tester 2 |
| Task 7 | ASVS-5-2-2 | S01, S06 | Tester 2 |
| Task 8 | ASVS-5-2-8 | S01, S06 | Tester 1 |
| Task 9 | ASVS-5-3-1 | S01, S03 | Tester 3 |
| Task 10 | ASVS-5-3-2 | S01, S03 | Tester 3 |
| Task 11 | ASVS-5-3-4 | S01, S02, S06, S07 | Tester 2 |
| Task 12 | ASVS-5-3-5 | S01, S06, S07 | Tester 2 |

Although the algorithm worked as expected, there are a few things that can be noted about this assignment. All testers were assigned the same number of tasks as defined by the linear programming constraints. Tester 1 has been assigned the hardest tasks 1 and 2, which require various skills. Tester 2 has been assigned tasks 6 and 7 that match the narrow skill set of the tester very well. In addition, tester 2 has been assigned tasks 11 and 12 that require skill *S07* which no tester has. These two tasks had to be assigned to someone, and in addition to the skill *S07*, they match the skills of tester 2 very well. Tester 3 has been assigned tasks 3, 4, 9, and 10 that he can somewhat perform, even though the tester's skill level is quite low. However, despite the limitations of experimental testing, we find linear programming to be useful in assigning tasks to testers based on their knowledge and experience.

## IV. Conclusions and Future Work

In conclusion, we propose a new approach for the pre-engagement phase of penetration testing based on linear programming. First, we have developed several metrics to measure the quality of assignments. Then we tested selected optimization algorithms to compare the resulting assignments with the developed metrics. From these algorithms, we found that linear programming is the most suitable for this purpose. To verify our approach, we conducted a case study on a small dataset with the selected algorithm, which demonstrates that applying optimization methods can increase the effectiveness of penetration testing, especially for better assignment of penetration testing tasks to individual testers.

Our further work will focus on the in-depth testing of other approaches that could be used to solve the assignment problem in the context of penetration testing, such as neural networks and reinforcement learning. After we compare all relevant algorithms, we will introduce feedback in the form of time spent on a task. This could lead to a more precise assignment of tasks to workers. In addition, we plan to develop a user interface built around this algorithm. If implemented in practice, this new approach can eliminate the need to manually assign tasks to workers and reduce the time spent on tasks.

## References


[1] S. Shah and B. M. Mehtre. "An overview of vulnerability assessment and penetration testing techniques," *Journal of Computer Virology and Hacking Techniques*, vol. 11, pp. 27-49, Nov. 2015.

[2] D. Garg and N. Bansal. "A Systematic Review on Penetration Testing," in *2021 2nd Global Conference for Advancement in Technology (GCAT)*, 2021, pp. 1-4.

[3] H. W. Kuhn, "The Hungarian method for the assignment problem," *Naval Research Logistics Quarterly*, vol. 2, no. 1, pp. 83-97, Mar. 1955.

[4] A. Nazemi and O. Ghezelsofla, "A dual neural network scheme for solving the assignment problem," *The Computer Journal*, vol. 60, no. 3, pp. 431-443, Mar. 2017.

[5] Z. Hu, R. Beuran, and Y. Tan, "Automated Penetration Testing Using Deep Reinforcement Learning," in *IEEE European Symposium on Security and Privacy Workshops (EuroS&PW)*, 2021, pp. 2-10.

[6] K. Tran, A. Akella, M. Standen, J. Kim, D. Bowman, T. Richer, and L. Chin-Teng, "Deep hierarchical reinforcement agents for automated penetration testing," 2021, arXiv:2109.06449.

[7] R. Gangupantulu, T. Cody, P. Park, A. Rahman, L. Eisenbeiser, D. Radke, R. Clark, and C. Redino, "Using Cyber Terrain in Reinforcement Learning for Penetration Testing," in *2022 IEEE International Conference on Omni-layer Intelligent Systems (COINS)*, 2022, pp. 1-8.

[8] T. Cody, A. Rahman, C. Redino, L. Huang, R. Clark, A. Kakkar, D. Kushwaha, P. Park, P. Beling, and E. Bowen, "Discovering Exfiltration Paths Using Reinforcement Learning with Attack Graphs," in *2022 IEEE Conference on Dependable and Secure Computing (DSC)*, 2022, pp. 1-8.

[9] "OWASP Application Security Verification Standard," owasp.org. https://owasp.org/www-project-application-security-verification-standard/ (accessed Feb. 27, 2024).


## Appendix

The procedure and calculations performed in Section III-C are described in detail in the following steps.

*Step 1*: Calculate the common skills of all tasks with respect to workers (penetration testers).

*Step 2*: Use the common skills values, that ought to be maximalized, to construct the maximized function:

$$\begin{aligned}\max\ & 2.3x_{1,1} + 1.8x_{1,2} + 0.8x_{1,3} + 3.8x_{2,1} + 3.5x_{2,2} + \\ & 1.0x_{2,3} + 1.4x_{3,1} + 1.8x_{3,2} + 0.7x_{3,3} + 1.4x_{4,1} + \\ & 1.8x_{4,2} + 0.7x_{4,3} + 1.7x_{5,1} + 1.0x_{5,2} + 0.8x_{5,3} + \\ & 1.5x_{6,1} + 1.9x_{6,2} + 0.5x_{6,3} + 1.5x_{7,1} + 1.9x_{7,2} + \\ & 0.5x_{7,3} + 1.5x_{8,1} + 1.9x_{8,2} + 0.5x_{8,3} + 1.5x_{9,1} + \\ & 1.0x_{9,2} + 0.9x_{9,3} + 1.5x_{10,1} + 1.0x_{10,2} + 0.9x_{10,3} + \\ & 2.1x_{11,1} + 2.7x_{11,2} + 0.5x_{11,3} + 1.5x_{12,1} + 1.9x_{12,2} + \\ & 0.5x_{12,3}\end{aligned}$$

*Step 3*: Make a constraint assuring that every single task is assigned only to one worker:

$$x_{i,1} + x_{i,2} + x_{i,3} = 1,\ i = 1..12$$

*Step 4*: Assure that every worker has assigned 4 tasks:

$$\begin{aligned}& x_{1,i} + x_{2,i} + x_{3,i} + x_{4,i} + x_{5,i} + x_{6,i} + x_{7,i} + x_{8,i} + \\ & x_{9,i} + x_{10,i} + x_{11,i} + x_{12,i} = 4,\ i = 1..3\end{aligned}$$

*Step 5*: Set the bounds on the $x$ variables:

$$0 \leq x_{i,1}, x_{i,2}, x_{i,3} \leq 1,\ i = 1..12$$

*Step 6*: After the utility function and constraints are set, the presented problem can be solved using regular linear programming solvers. For the purpose of this case study, Python linear programming solver method `linprog` from `scipy` library. Obtained result in format of $(x_{1,1}, x_{1,2}, x_{1,3}, x_{2,1} \ldots x_{12,3})$ is the following sequence of binary values:

$$\begin{aligned}&(1,0,0,1,0,0,0,0,1,0,0,1,1,0,0,0,1,0,0,1,0,1,0,0,0,\\ &0,1,0,0,1,0,1,0,0,1,0)\end{aligned}$$

The result is written out in more readable format in Table II.

# Lattice-based Threshold Signature Optimization for RAM Constrained Devices

1st Vladyslav Shapoval
*Department of Telecommunication*
*Brno University of Technology*
Brno, Czech Republic
236349@vutbr.cz

2nd Sara Ricci
*Department of Telecommunication*
*Brno University of Technology*
Brno, Czech Republic
0000-0003-0842-4951

***Abstract*—The DS2 scheme is a lattice-based $(n, n)$-threshold signature based on the standardized Dilithium signature. However, deploying DS2, as well as Dilithium, on microcontrollers is a challenge due to the memory limitations of these devices. While the decryption phase can be implemented relatively straightforwardly, the key generation and signing phases require the generation and manipulation of large matrices and vectors, which can quickly exhaust the available memory on the microcontroller. In this paper, we propose an optimization of the DS2 key generation and signing algorithms tailored for microcontrollers. Our approach focuses on minimizing memory consumption by generating large elements, such as the commitment key *ck* and the random commitment parameter *r*, on the fly from random and non-random seeds. This approach significantly reduces the overall size of the signature from 143 KB to less than 5 KB, depending on the number of signers involved. We also split the algorithms into two distinct components: a security-critical part and a non-security-critical part. The security-critical part contains operations that require secret knowledge and must be run on the microcontroller itself. Conversely, the non-critical part contains operations that do not require secret knowledge and can be performed on a connected, more powerful central host.**

***Index Terms*—Lattice-based cryptography, threshold signature, micro-controllers, optimization**

## I. Introduction

Threshold signatures allows us to enhance security in corporate environments. For instance, in blockchain networks, threshold signatures are used for multi-signature wallets. Specifically, multiple parties have to collaborate to sign a transaction, enhancing security and reducing the risk of single points of failure. With the threat posed by quantum computers, there is a need to shift to post-quantum secure signatures. Addressing this concern, Damgaard, Orlandi, Takahashi, and Tibouchi [1] proposed a two-round $n$-out-of-$n$ signature scheme called DS2 and built upon the Dilithium signature and, therefore, post-quantum secure. Notably, Dilithium is one of the signatures selected for standardization by the National Institute of Standards and Technology (NIST) 2022[1], underlining suitability for post-quantum security challenges. Furthermore, NIST has published a draft for their First Call for Multi-Party Threshold Schemes [2], which points out a growing interest in threshold schemes.

[1] https://csrc.nist.gov/Projects/post-quantum-cryptography/selected-algorithms-2022

Based on the work of Damgaard *et al.*, Dobias *et al.* [3] introduced a practical implementation of the DS2 algorithm on a Linux system running on a Raspberry Pi. Notably, this implementation requires at least 176,256 bytes of SRAM to be constantly allocated for storing data about parties and the resulting signature. Additionally, it requires at least 445,744 bytes of both heap and stack allocated memory for temporary parameters and data buffers. This implies that a microcontroller with at least 640 KB of SRAM is needed to run this implementation of DS2. Also, this implementation is heavily reliant on the transfer of large amounts of data over the network (hundreds of kilobytes). This could potentially create an additional bottleneck if the algorithm is used in conjunction with low-powered WPAN standards like 802.15.4. Only a small percentage of microcontrollers, currently available on the market, meet these requirements.

### A. Contribution

In this work, we present an optimization of DS2 $(n, n)$-threshold signature scheme implementation [3]. We have introduced several optimizations to the signature algorithm and added a central node and host PC to the original DS2 scheme. The host performs complex computation, thus relieving constrained secure device and reducing memory footprint by 90% and signature size by 96%. Moreover, we provide benchmarks of our implementation on constrained devices using ARM Cortex®-M4 core, compliant with 802.15.4 wireless communication standard.

## II. Architecture and Scheme Optimization

In this section, the architecture of our system and the proposed optimizations of DS2 scheme are described.

### A. Test bench description

Our system consists of two parts: a security-critical part and a non-security-critical part. The security-critical part contains operations that require secret knowledge and must be run on the microcontroller itself, namely the Secure Element (SE). Conversely, the non-critical part contains operations that do not require secret knowledge and can be performed either on the Central Element (CE) (i.e., another microcontroller) or on a host computer connected to the CE. The introduction of the

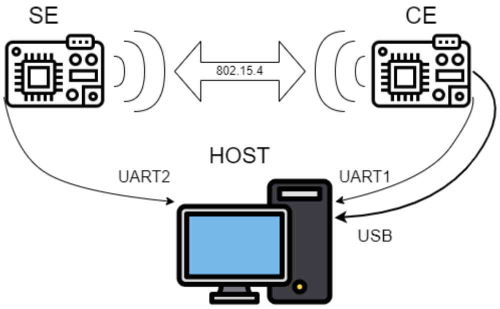



Fig. 1. Architecture of the test bench

central element is dictated by the chosen wireless standard, IEEE 802.15.4 [4]. This standard is widely used in home automation and industry and requires a central node called a Coordinator, which is responsible for the establishment of the wireless network. Also, the majority of PCs do not have 802.15.4 compliant hardware, so for communication with SE nodes, they need an 'adapter' that will act as the CE coordinator. The architecture of our test bench is shown on figure 1. CE is connected to the host PC via USB interface. Both SE and CE also send debug information to the host via UART interfaces.

The microcontroller chosen for the CE and SE is the STM32WB55 with an embedded low-power radio compliant with IEEE 802.15.4-2011 standard. An important point for our purposes is that the STM32WB55 has 256KB SRAM1, consisting of 192KB of actual RAM memory and 64KB of hardware memory parity check. Additionally, the STM32WB55 provides access to the NIST-compliant True Random number generator [5].

In our test bench, for testing purposes, we have only one SE node. In this way, the CE node, implemented on the same microcontroller as the SE, is able to perform signature computation itself without relaying data to the PC host. This approach allowed us to obtain execution time measurements, which we can compare in the future with measurements taken when the signature is calculated by the host PC.

### *B. Algorithms description*

We chose the same parameters for our DS2 implementation as in [3], namely $K = 2$, $L = 2$, $TC\_COL = 69$, $N = 256$, $Q = 8380417$. However, we set the number of parties only to 1 for testing purposes. It is noteworthy that each signer does the same computations.

Algorithm 1, that is for the key generation, was split into 7 parts (stages) permitting the communication between CE and SE and keeping the secure computations in the SE. Therefore, those parts were distributed between nodes SE and CE and enumerated as SE0, CE1, SE2, CE3, SE4, CE5, SE6. Algorithm 4 shows that the signature scheme was split into 4 parts (stages). Those parts were also distributed between nodes SE and CE and enumerated as SE0, CE1, SE2, CE3. Algorithms 1 and 4 depict which operations were assigned

**Algorithm 1** `KeyGen` $(1^\kappa)$

1: $\rho_n \leftarrow \{0,1\}^{256}$ ▷ `SE0`
2: Send out $\rho_n$ ▷ `SE0`
3: Receive $\rho_i; i \in [n-1]$ ▷ `CE1`
4: $\rho = \mathrm{H}(\rho_1, \rho_2, ..., \rho_n)$ ▷ `CE1`
5: Send out $\rho$ ▷ `CE1`
6: $\mathbf{A} \leftarrow R_q^{k\times\ell} := \mathtt{Sam}(\rho)$ ▷ `CE1/SE2`
7: $\mathbf{t}_n := \bar{\mathbf{A}}\mathbf{s}_n; \mathbf{s}_n \leftarrow S_\eta^{\ell+k}$ ▷ `SE2`
8: $(\mathbf{t}_{n_1}, \mathbf{t}_{n_0}) := \mathtt{Power2Round}(\mathbf{t_n}, d)$ ▷ `SE2`
9: $g_n := \mathrm{H}_2(\mathbf{t}_{n_1})$ ▷ `SE2`
10: Send out $g_n$ ▷ `SE2`
11: Receive $g_i; i \in [n-1]$ ▷ `CE3`
12: Send out $\mathbf{t}_{n_1}$ ▷ `SE4`
13: Receive $\mathbf{t}_{i_1}$ and check $g_i = \mathrm{H}_2(\mathbf{t}_{i_1}); i \in [n-1]$ ▷ `CE5`
14: $\mathbf{t}_1 := \sum_{i\in[n]} \mathbf{t}_{i_1}$ ▷ `CE5`
15: $tr \in \{0,1\}^{384} := \mathtt{CRH}(\rho||\mathbf{t}_1)$ ▷ `CE5`
16: Send out $tr$ ▷ `CE5`
17: Receive $tr$ ▷ `SE6`
18: **return** $pk = (\rho, tr, \mathbf{t}_1),\ sk = (\rho, tr, \mathbf{s}_n, \mathbf{t}_{n_0})$

to which stages in key generation and signature algorithms and how we distributed stages between CE and SE nodes. Algorithms 2 and 3 demonstrate how the commitment for each SE node is computed and verified. Algorithm 5 shows the verification process.

**Algorithm 2** `Commit` $(w_n, \mathbf{sr}_n, sck)$

1: **For** $k \in [0, 68]$ do
2: **For** $j \in [0, K-1]$ do
3: $\mathbf{r}_n[k][j] \leftarrow D(S_r, \mathbf{sr}_n, j, k)$
4: **For** $i \in [0, K-1]$ do
5: $ck[i][k] := \mathrm{H}(sck, i, k)$
6: $com_n[i][j] + = ck[i][k] \cdot \mathbf{r}_n[k][j]]$
7: $com_n + = w_n$
8: **return** $com_n$

A number of changes were introduced in each algorithm to comply with strict memory restrictions in microcontroller. Namely:

- **Algorithm 1 - Lines 16 and 17**: The resulting vector $t1$ is calculated by the CE but is not used directly by the SE nodes in the signature algorithm. Therefore, the CE needs only to calculate the parameter $tr$, which is 16 bytes long and send it to SE nodes instead of the full $t1$ vector.
- **Algorithm 4, Lines 1 and 2**: The commitment key $ck$ and the random parameter $r_n$ in the original DS2 scheme occupy at least 138KB each. Therefore, we represent these parameters by the seed values $sr_n$ and $sck$, which are 16 bytes each. The $sck$ is a short hash derived from the message and the parameter $tr$.
- **Algorithm 4, Line 4 and Algorithm 2**: The commitment computation function was modified. The Commitment key $ck$ and the random parameter $r_n$ are generated from

the respective seeds one block at a time. Each block is represented by a single polynomial. The computation of the commitment is also done block by block. This approach reduces the buffer sizes for both $ck$ and $r_n$ from 138KB to 1KB each. Notably, the computation is slower because matrix multiplication requires the regeneration of the same block several times.

- **Algorithm 4, Lines 8-11**: The challenge polynomial is generated from the seed $sc$, which is 16 bytes long. The seed is generated from the calculated commitment $com$ in the form of a matrix of polynomials. Instead of sending the whole matrix to the SE node, the seed is sent in a single packet. Accordingly, the SE node can regenerate the challenge polynomial $c$ from the seed.
- **Algorithm 4 - Lines 15-16**: The SE nodes send out seeds $sr_i$ instead of the whole matrix $r_i$, which allows for a significant reduction in traffic.

**Algorithm 3** `Open` $(sck, com_i, \mathbf{sr}_n, \mathbf{w}_n)$

1: **For** $k \in [0, 68]$ do
2:   **For** $j \in [0, K-1]$ do
3:     $\mathbf{r}_n[k][j] \leftarrow D(S_r, \mathbf{sr}_n, j, k)$
4:     **For** $i \in [0, K-1]$ do
5:       $ck[i][k] := \mathrm{H}(sck, i, k)$
6:       $com_n[i][j] += ck[i][k] \cdot \mathbf{r}_n[k][j]$
7: $com_n += w_n$
8: **If** $com_i == com_n$ **return** 1
9: **Else return** 0

- **Algorithm 4, Line 19 and Algorithm 3**: The Open algorithm uses the same optimization function as the Commit and follows the same logic of block-by-block computation from the given seeds.
- **Algorithm 4, Line 21**: The signature contains all received seeds $sr_i$ and $sc$, instead of the resulting parameter $r$. This allows for a reduction in the signature size from 143KB to 5KB.
- **Algorithm 5, Lines 1-4**: Verification of the signature will be done on the Host PC, and host must recreate parameter $r$ from the given seeds $sr_i$ and the challenge polynomial.

**Algorithm 5** `Verify` $(pk, \Sigma, \mu \in M)$

1: $c = H(sc)$
2: **For** $i \in [n-1]$ do
3:   $\mathbf{r}_i \leftarrow D(S_r, \mathbf{sr}_i)$
4:   $\mathbf{r} := \sum_{i \in [n]} \mathbf{r}_i$
5: $ck := \mathrm{H}_3(\mu, pk)$
6: $\mathbf{w}' := \bar{\mathbf{A}}\mathbf{z} - c\mathbf{t}_1 \cdot 2^d$
7: $c' := \mathrm{H}_0(\mathtt{Commit}_{ck'}(\mathbf{w}', \mathbf{r}), \mu, \mathtt{CRH}(pk))$
8: **If** $||\mathbf{z}||_2 \leq \sqrt{n}B$ and $c = c'$ **return** 1
9: **Else return** 0

## III. Experimental Results

In this section, our benchmarks on communication and memory consumption are presented.

**Algorithm 4** `Sign` $(pk, sk_n, \mu \in M)$

1: $sck := \mathrm{H}(\mu, tr)$ ▷ SE0/CE1
2: $\mathbf{sr}_n := rand()$ ▷ SE0
3: $\mathbf{w}_n := \bar{\mathbf{A}}\mathbf{y}_n; \mathbf{y}_n \leftarrow D_s^{\ell+k}$ ▷ SE0
4: $com_n := \mathtt{Commit}(\mathbf{w}, \mathbf{sr_n}, \mathbf{sck});$ ▷ SE0
5: Send out $com_n$ ▷ SE0
6: Receive $com_i; i \in [n-1]$ ▷ CE1
7: $com = \sum_{i \in [n]} com_i$ ▷ CE1
8: $sc := \mathrm{H}_0(com, \mu, tr);$ ▷ CE1
9: Send out $sc$ ▷ CE1
10: Receive $sc$ ▷ SE2
11: $c = H(sc)$ ▷ SE2/CE1
12: $\mathbf{z}_n := \begin{pmatrix} \mathbf{z}_{n_1} \\ \mathbf{z}_{n_2} \end{pmatrix} := c\mathbf{s}_n + \mathbf{y}_n$ ▷ SE2
13: Rejection sampling on $(c\mathbf{s}_n, \mathbf{z}_n)$, with probability $min(1, D_s^{l+k}(\mathbf{z}_n)/(M \cdot D_s^{l+k}(\mathbf{z}_{c\mathbf{s}_n,n})))$ continue, otherwise goto Line 3 ▷ SE2
14: $\mathbf{z}_{n_2} = \mathbf{z}_{n_2} - c\mathbf{t}_{n_0}$ ▷ SE2
15: Send out $(\mathbf{z}_n, \mathbf{sr}_n)$ ▷ SE2
16: Receive $(\mathbf{z}_i, \mathbf{sr}_i); i \in [n-1]$ ▷ CE3
17: **For** $i \in [n-1]$ do
18:   $\mathbf{w}_i := \bar{\mathbf{A}}\mathbf{z}_i - c\mathbf{t}_{i_1} \cdot 2^d$ ▷ CE3
19:   **If** $||z_i||_2 > B$ or $\mathtt{Open}(sck, com_i, \mathbf{sr}_i, \mathbf{w}_i) \neq 1$ **then** abort ▷ CE3
20:   $\mathbf{z} := \sum_{i \in [n]} \mathbf{z}_i$ ▷ CE3
21: **return** $\Sigma = (sc, \mathbf{z}, \mathbf{sr}_1, \mathbf{sr}_2, ..., \mathbf{sr}_n)$

### A. Communication transport protocol

The main drawback of the original DS2 implementation [3] is that each node must store in its memory all open parameters of all other nodes during the signature computation, mainly their respective shares of the open key and commitments, which leads to a large memory footprint. The simplest solution to this problem is to introduce a powerful central host. The host will store key shares and commitments from all parties and perform calculations on them itself. In this case, each party does not have to store information about other parties and waste resources on the computation of the signature.

For communication between nodes, a MAC layer of the 802.15.4 standard is used. Therefore, in order to correctly identify data and its source in the network, we created a simplified transportation layer protocol. The maximum 802.15.4 MAC frame payload length varies between 72 and 116 bytes, depending on the length of the MAC frame header. We reserved 8 bytes for the packet header and, from empirical experience, we limited data length to 100 bytes. Each SE node is given its own ID from 0 to 253. Node IDs 254 and 255 are reserved for CE and broadcast, respectively. The structure of the packet is shown in Figure 2.

| 1B | 1B | 1B | 1B | 4B (optional) | max 100B (optional) |
|---|---|---|---|---|---|
| packet_len | dst_node_id | src_node_id | msg_code | data_offset | data |

Fig. 2. Packet structure.

### B. Execution time measurement

A unique corresponding message code was assigned to each stage in each algorithm. The execution of those stages is triggered upon receiving a packet with a corresponding message code. The internal state machine is responsible for the correct execution order of stages.

TABLE I
TIME MEASUREMENTS IN CE AND SE NODES

| Algorithm | DS2 operations | Total | $\Delta_{DS2/Total}$ |
|---|---|---|---|
| KeyGen CE | 18,479,909 | 44,737,892 | 41% |
| Sign CE | 1,556,054,176 | 1,586,106,405 | 98% |
| KeyGen SE | 20,121,257 | 44,737,892 | 44% |
| Sign SE | 1,545,056,389 | 1,575,181,694 | 98% |

We measured the execution time in CPU cycles of the microcontroller using dedicated software timers to obtain a value independent of CPU frequency. We found that the time needed to receive and process a single packet can vary several times, from 220,000 CPU cycles up to 800,000, due to interrupts, internal task scheduling, and inter-processor communication inside the microcontroller. Because of this, the time spent on each stage, as described in the text above and in Algorithms 1 and 4, also varies significantly, and overall measurements would be deemed inaccurate. However, the average execution time of the entire algorithm is generally consistent enough, and therefore, we chose a different approach: we measured only the execution time needed for DS2 scheme-related calculations, such as commitment generation or polynomial operation. At the same time, a separate timer measures the total execution time of algorithms, including operations not directly related to the DS2 scheme itself, like sending and receiving a packet or reassembling packets. Execution time measurements are shown in Table I. It can be seen that in the case of key generation, cryptographic computations take only 41% of the total execution time. The rest is spent on send/receive operations. On the other hand, during signature generation, 98% of the execution time is occupied by the commitment generation only, which takes on average roughly 1,500,000,000 CPU cycles.

### C. Memory consumption

The memory footprint of DS2 implementation on microcontroller consists of several parts: static memory allocated for global variables ($M_{static}$), stack-allocated memory for function calls ($M_{stack}$), and static memory allocated by the main application ($M_{app}$)[1]. Memory allocation in the heap region is rarely used on microcontrollers because in general it is handled by the OS.

Using internal build analyzer tool from CubeIDE, we determined memory consumption of our implementation. As shown in Table II, we managed to reduce memory consumption to only 10% in comparison to the original DS2 scheme implementation[2].

TABLE II
MEMORY CONSUMPTION COMPARISON

| Node | $M_{static}$ | $M_{stack}$ | $M_{app}$ | Total | $\Delta_{Total/Origin}$ |
|---|---|---|---|---|---|
| CE | 20KB | 28KB | 22KB | 70KB | 11% |
| SE | 29KB | 14KB | 22KB | 65KB | 10% |

Additionally, we achieved a reduction of signature size. Original signature consisted of 3 parts: polynomial challenge $c$ (size 1KB), two vectors $z_n = (z_1, z_2)$ (2KB each), and random parameter $r$ (a polynomial matrix $2 \times 69$ = 138KB). We replaced $r = \sum_{i=0}^{n-1} r_i$ with the concatenated random seeds themselves $sr_0, sr_1, ..., sr_n$, each 16 bytes, obtained from each party. If we also replace $c$ by seed $sc$ 16 byte, we can obtain a signature size of less than 5KB for any feasible number of parties.

## IV. CONCLUSION

In this work, we propose several optimization techniques and modifications of the DS2 scheme for its deployment on microcontrollers and other memory-constrained devices. We were able to reduce memory consumption by 90% at the cost of execution time. Additionally, we reduced the signature size by 96%, so it can now be stored in constrained devices. The most computation-heavy operation is the commitment generation, which takes on average 1,500,000,000 CPU cycles. From this fact, it follows that in the future, the use of a more powerful host PC to open commitments from parties is justified, since the time to generate a signature will further increase linearly with the increase in the number of participants. However, centralizing the scheme can potentially introduce several vulnerabilities. An adversary can manipulate data sent as a response from the CE to the SE node. For example, an adversary can change the parameter $tr$ during key generation. Further investigation and development are required.

## REFERENCES


[1] DAMGÅRD, Ivan; ORLANDI, Claudio; TAKAHASHI, Akira and TIBOUCHI, Mehdi, 2021. Two-Round n-out-of-n and Multi-signatures and Trapdoor Commitment from Lattices. Online. Public-Key Cryptography – PKC 2021. Lecture Notes in Computer Science. S. 99-130. Available at: https://doi.org/10.1007/978-3-030-75245-3_5.

[2] BRANDÃO, Luís T. A. N. and PERALTA, Rene, 2023. NIST First Call for Multi-Party Threshold Schemes. Online. S. 1–61. Available at: https://doi.org/10.6028/NIST.IR.8214C.ipd.

[3] DOBIAS, Patrik; RICCI, Sara; DZURENDA, Petr; MALINA, Lukas and SNETKOV, Nikita, 2023. Lattice-Based Threshold Signature Implementation for Constrained Devices. Online. Proceedings of the 20th International Conference on Security and Cryptography. 2023-7-10, s. 724-730. Available at: https://doi.org/10.5220/0012112700003555.

[4] IEEE Standard for Low-Rate Wireless Networks, 2020. Online. IEEE Std 802.15.4-2020 (Revision of IEEE Std 802.15.4-2015). S. 1-800. Available at: https://doi.org/10.1109/IEEESTD.2020.9144691.

[5] STMICROELECTRONICS, 2022. Datasheet - STM32WB55xx STM32WB35xx. Online. .: STMicroelectronics. Available at: https://www.st.com/en/microcontrollers-microprocessors/stm32wb55rg.html.


[1]On average $M_{app} \approx 22KB$. This includes all memory buffers required for UART, USB and wireless communication and internal state variables.

[2]Memory consumption of DS2 scheme implementation from [3] $\approx 640KB$.

# ABB v Brně
## Lepší svět začíná s vámi

Jsme součástí globálního technologického lídra působícího v oblasti elektrotechniky, robotiky, automatizace a pohonů. Z Brna dodáváme rozváděče vysokého i nízkého napětí, modulární a digitální systémy, přístrojové transformátory a senzory. Tyto produkty míří do nejnáročnějších projektů ve více než 100 zemích světa. Najdete ke mimo jiné v nejvyšší budově světa Burdž Chalífa, v londýnském metru nebo v datovém centru Facebooku. Svou pobočku zde má i Operační centrum.

Přibližně 2100 zaměstnanců

Tržby 10 mld. Kč

Významné centrum výzkumu a vývoje

Nejnáročnější projekty ve 100 zemích světa

4 výrobobní závody, centrum digitálních řešení, R&D, servis a operační centrum

**Studentům a absolventům nabízíme:**
- Placené brigády
- Odborné praxe
- Stáže
- Trainee program
- Vedení balakářské nebo diplomové práce
- Hlavní pracovní poměr po ukončení studia

**Kontaktujte naši náborářku:**
Lucie Vašatová
lucie.vasatova@cz.abb.com
+420 704 997 141

Virtuální prohlídka
Vídeňská Heršpická

Webové stránky
Operační centrum

## Rozváděče vysokého napětí

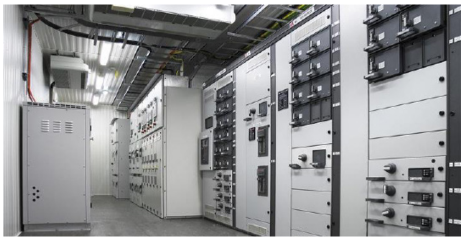
Výroba rozváděčů vysokého napětí se v našem závodě těší dlouhé tradici a pyšní se statusem největšího závodu svého druhu v Evropě. Co do počtu zaměstnanců, tak i plochy výroby, se jedná o největši jednotku v rámci brněnského závodu. Své produkty dodává do přibližně 100 zemí světa.

## Rozváděče nízkého napětí

Rozváděče nízkého napětí fungují v režimu tzv. virtuálního závodu a sdílí svou kapacitu se závodem v polské Bielsko-Białe. Opět se jedná o jeden z největších závodů svého druhu v Evropě. Jeho produkty jsou dodávány do přibližně 50 zemí světa.

## Přístrojové transformátory a senzory

Jednotka přístrojových transformátorů a senzorů je 2. největší v rámci brněnského závodu. Zároveň se jedná o největší závod svého druhu na celém světě. Vůbec první transformátor byl v závodě vyroben již v roce 1919 a jejich výroba je tak založena na více než stoletém know-how.

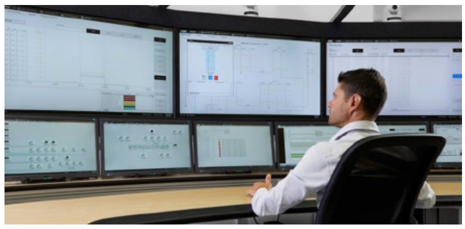


## Modulární systémy

Přestože se jedná o mladou jednotku, má už za sebou několik významných milníků. Své modulární rozvodny, dodávané obvykle v kontejnerovém řešení, dodává do projektů napříč odvětvími. Nejvýznamnější z nich jsou  data centra a projekty v oblasti těžby a distribuce ropy a zemního plynu.

## Centrum digitálních řešení

Cílem jednotky Digital Solution Center je realizace inteligentních řešení v oblasti monitoringu, sběru a archivace dat a hlavně vzdáleného řízení distribuce energie. Zákazníkům nabízí řešení na míru, flexibilitu a vysokou přidanou hodnotu.

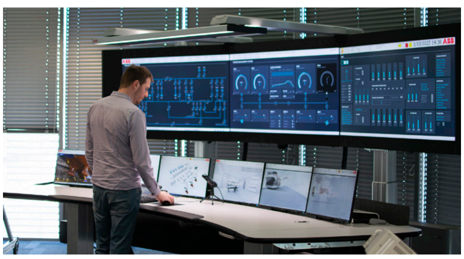


## Servis přístrojů vysokého a nízkého napětí

Jednotka servisu spolupracuje se zákazníky v rámci oprav, údržby a vylepšování jejich zařízení po celou dobu jejich životního cyklu. U svých zákazníků v přibližně 100 zemích světa stráví technici asi 40 tisíc hodin ročně. Jednotka využívá nejmodernější nástroje včetně tréninkového centra pro zákazníky nebo rozšířené reality.

## Technologické centrum

Technologické centrum se zabývá výzkumem a vývojem produktů. V rámci ABB se celosvětově jedná o velmi významnou výzkumnou jednotku. K dispozici má špičkově vybavenou laboratoř umožňující provádění nejrůznějších simulací a výrobu a testování prototypů.



## Operační centrum Evropa

Operační centrum Evropa (EUOPC) je globálním centrem ABB pro průmyslovou automatizaci a elektrifikaci. Zákazníky má asi v 50 zemích. Z EUOPC pocházejí řídicí systémy pro nové generace ropných plošin. Zajímavé jsou projekty přispívající ke stabilizaci el. sítě při přechodu na generování a využití zeleného vodíku nebo elektrifikaci lodní přepravy.

# Modeling the spread of computer viruses

Amaury de Kergorlay
*Department of information technology architecture*
*ISEP – Paris Institute of Digital Technology*
Paris, France
amaurydekergorlay@gmail.com

Josef Rebenda
*Department of Mathematics, Faculty of Electrical Engineering and Communication*
*Brno University of Technology*
Brno, Czech Republic
rebenda@vutbr.cz

Yuriy Rogovchenko
*Department of Mathematical Sciences, Faculty of Engineering and Science*
*University of Agder*
Kristiansand, Norway
yuriy.rogovchenko@uia.no

***Abstract*—The paper presents construction and implementation of a model of the spread of computer viruses in a small network. The model is based on probabilistic approach that describes interaction between an attacking virus and a computer. Simulation results and their interpretation suggest that the model is in a good concordance with reality.**

***Keywords*—mathematical modeling, simulation, spread of a computer virus, cybersecurity, probabilistic approach**

## I. Introduction

Nowadays, there are a lot of computers with internet access, and about 2200 of them get hacked every day ([1], [2]). It makes about 810 000 infected computers every year that we know of. Those infected computers may cause great damage to families, people or companies as their personal and private data may end up anywhere in the world being sold to the highest bidder or used to blackmail them.

During my internship at Brno University of Technology, I worked together with my advisors on the simulation of the spread of a computer virus to see how fast the contagion can be transmitted across a given number of computers. We started by investigating existing mathematical models of this phenomenon.

A lot of papers use the SIR and SEIR model to represent the spread in the same way as biological viruses spread in real life such as Covid-19. We found that it was not the most convenient model as it is not the physical proximity of a neighbor's infected computer that will increase the chance that a computer gets infected. In fact, all computers connected to the internet are exposed to an infection.

Instead of the continuous compartmental model we decided to choose a probabilistic approach where the probability of being hacked vary from one computer to another. In this paper we describe construction of the model, present simulation results, introduce and discuss possible explanations of the results and outline further improvements of the model.

## II. Construction of the Model

### A. *Level of attack and defence*

As mentioned in the Introduction, we implemented a model based on probability. However, having a single probability to be infected does not reflect our reality. In the modern society, people have different knowledge and skills in computer science and therefore do not take the same precaution while browsing the web. Assigning the same probability of a computer virus infection to a cybersecurity expert and a teenager discovering internet does not seem realistic and accurate.

That was the reason why we decided to implement three levels of defense, simulating cybersecurity knowledge of the users. Level one is the most vulnerable user while level three user has the highest degree of knowledge.

To complement this process of thinking, we also implemented three levels of attack. As the users do not have the same knowledge, also the attacks differ in the extent of stealth, impact, or procedures used. A suspicious mail can be easily spotted and recognized by most users whereas a trojan horse is not easily discoverable for some users. We decided to split the attacks into three levels using the same methodology as for the three levels of defense.

Combining the two components of the model we get 9 possible encounters between the different levels of attack and defense.

### B. *Probability of computer virus infection*

The probabilities that we use in the model are empiric and we will now explain the reasons for choosing those values.

We considered that the same level of attack and defense has a 50% probability to be infected except for the level three, as it represents experts in cybersecurity that were specifically trained in how to prevent those attacks and thus avoid them more easily.

Further, if the level of defense is increased by one, the probability to get infected drops by 20-35%. Then as the probability of getting infected can never be zero, there is a 5% probability for a level three expert's computer to get infected by a level one attack. The same reasoning is used for a level three attack and a level one defense, where there is a 95% chance to get infected but can never be 100%.

The table below, we will refer to it as the probability matrix, shows the probabilities of the computer being infected for every encounter:

| Level of defense\level of attack | Level one | Level two | Level three |
|---|---|---|---|
| Level one | 50% | 75% | 95% |
| Level two | 25% | 50% | 60% |
| Level three | 5% | 15% | 40% |

### *C. Stages of infection*

Next, we considered different stages that a computer can be at. As a basis we chose the compartmental model known as SEIR (Susceptible-Exposed-Infected-Recovered) which can be used for description of stages of an infection whether biological or cybernetic.

Stage 0 is the default (initial) stage for each computer in the experiment. The computer is sane and can be a target of an attack. We call this stage susceptible.

Stage 1 is the exposed stage. In this stage, the computer has been successfully attacked but the malicious code has not been activated yet. Therefore, it cannot infect other computers.

Stage 2 is the infected or infectious stage. In this stage, the computer has been successfully attacked and tries to attack other computers in the network.

Stage 3 is the recovered stage, when the computer has been cleaned of the virus. Then the computer will go back to stage 0.

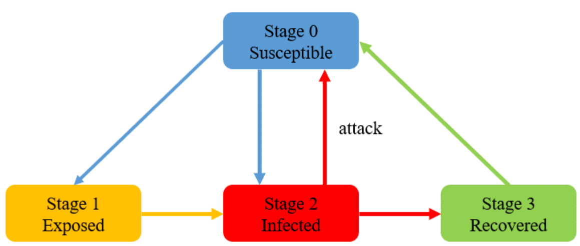



Fig. 1. Diagram of the model dynamics

Once we adopted the model of the spread dynamics (Fig. 1), we continued with other specific aspects. It is unlikely that a person who has lost data or money through clicking on a link in a suspicious email will ever click again on the same type of link. To include increased experience and awareness of such user against an attack in our model, a computer will gain one level of defense when it gets infected and reaches the stage 3 eventually. However, the level of defense can never exceed level three which is the level of cybersecurity experts.

Another parameter that we regarded as useful is incubation period: the number of days of incubation of the virus inside the computer. In the simulation application one can set the incubation period as a non-negative integer at the beginning of the simulation.

To keep the simplicity of the model, we considered the "population" – number of computers in the network – being constant in time. Therefore, no attack can be "lethal" to a computer and every infected computer always gets to the stage 3 after some time.

Another specification is that a computer can only have one virus at a time. In the reality when someone has a computer virus in their machine and is aware of that, such user will first try to resolve the problem before using the computer again. Practically, when a computer gets infected, we remove it from the list of susceptible computers.

With those specific parameters and effects taken into consideration, it was possible to define the model of a computer in a network. A computer is defined by an ID, a level of defense, a stage and the number of days spent in that stage. This last attribute was included to implement the recovery period once a computer got infected.

The ID is an integer from a given range, for example from 1 to 100. The range represents the number of computers in the network. We decided to keep the upper limit of the range rather low, not exceeding 100, to avoid overloading the software that runs simulations. The stage is an integer between 0 and 3 that defines the current stage (compartment) of the computer as discussed previously. The level of defense is an integer from 1 to 3 reflecting expertise of the user in cybersecurity. The number of days in the stage is an integer greater than or equal to 0.

## III. Implementation and Simulations

After completing the theoretical part by including specific aspects into the model, we needed to implement the model in a suitable programming language. We used Python due to its versatility and accessibility. It is easy to create figures in this language and therefore it is convenient for displaying simulation results. In the code we implemented several random actions that can have impact on the outcome. They concern the probability to get infected presented earlier, the choice of which computers get attacked, and the way the level of defense is assigned to a computer. For the last one we set a vector of probabilities to generate the level of defense. First, we considered that nowadays, most people are used to browsing the web and have heard about cybersecurity without being experts. Therefore, we set a 55% chance for a computer to be a level 2 computer. Second, as the digital world becomes more accessible to young, elderly, and unexperienced people, we considered that a computer has a 35% chance to be a level 1 computer. That left us with a 10% chance to become a level 3 computer, which seems to be a realistic as IT skills becomes very important and people get better training on cybersecurity problems at their workplace. Having set those probabilities, we used the random choice function and get similar distribution of the levels of defense in each simulation so the data can be compared. Further, every day of the simulation, computers which will be attacked that day are chosen randomly. Finally, a random number between 0 and 1 is assigned to each attacked computer, and if the number is lower than the corresponding probability in the probability matrix, the computer becomes infected. Including all those random effects can result in differences between two simulations having the same parameters.

To start a simulation, one must input two integers: the number of computers in the network and the number of days that represent the duration of the experiment. To allow for observing trends, there is a fixed number of attacks every day. If for example there is 6 attacks every day, we may expect that the number of successful attacks increases or decreases over time, which will help us understand the trends in the network behavior.

After each "day" of the simulation, successful attacks of the day and the total number of infected computers are counted. Those values are stored in a dataset and displayed in a figure at the end of the simulation. Both the dataset and the figure can then be used for data analysis.

We performed three types of simulations. The first type was a small-scale simulation with a population of 30 computers and a 20-day period, followed by a medium-scale simulation with a population of 100 computers and a 150-day period. The last type was the longest large-scale simulation with 100 computers and 650-day period. Results of simulations of the first type can be very different from each other whereas simulations of the two other types should be more similar within the type.

The purpose of the single simulation of the first type was to see the general behavior of the model and to have simple data to interpret. Next, we increased the number of computers and the duration of the experiment to see whether there is any scalable phenomenon: Does increasing the numbers emphasize a specific behavior? Addressing this question required larger number of simulations of the second type. The last type was also performed in a larger number of simulations to reveal whether there are any potential trends or equilibria by increasing the duration of the experiment significantly.

## IV. Interpretation of the Results

### A. The 20-day simulation

The first simulation was a convenient experiment to check functionality of the model. In this simulation we only had one attack on each level per day due to the small number of computers. An interesting phenomenon to start with is learning from mistakes and adaptability of the users, that is, which levels of attack infect the computers in which phase of the experiment. As we can see in Fig. 2 below, all attacks were having success in the beginning, shown by the spikes going from 0 to 1. However, close to the middle of the experiment we can see that the level one attack became less successful as we cannot see any

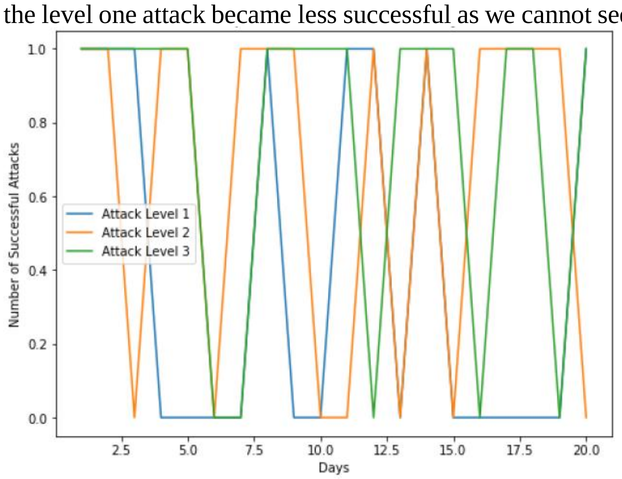



Fig. 2. Successful attacks day by day in the short-term simulation

blue spike after day 12. It might reflect the aspect of evolution of experience. However, the success rate of the two other levels remained the same, especially for the level three attack.

On the contrary, the other type of figure – number of infected computers over time - could not bring any interesting information as it only displayed a sharp rise in the number of the total computer infected. Therefore, we needed to increase the length of the simulation period to observe a potential trend.

### B. The medium scale simulation

In this type of simulation, we increased the number of computers to 100 in the network and the number of days to 150.

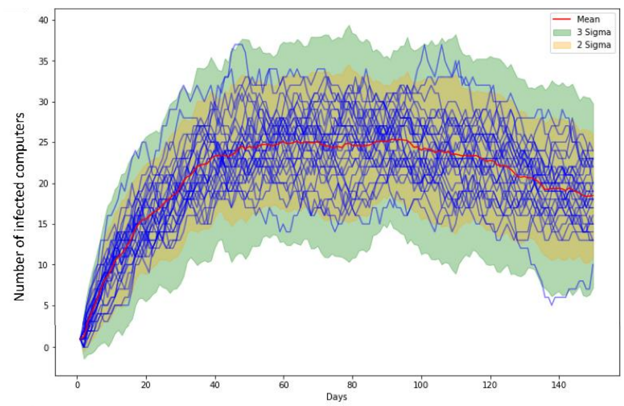



Fig. 3. Number of infected computers in the medium-scale simulations

Evolution of the total number of infected computers in the medium-scale simulation is displayed in Fig. 3 above.

In this figure we can see the results of 30 simulations run with the same parameters. From all this data we obtained more information about the model and the phenomenon we tried to simulate. The red curve represents the mean of the number of infected computers for each day for all the 30 simulations. It reached its maximum after 40 days with a total of 25 infected computers at a time. This result tends to validate the expected proportion of around 25-30% of the population infected at the peak of the infection. The "2 Sigma" stripe corresponds to the interval where 95% of the data should be found. In this "2 Sigma", we observe that the peak varied from 17 to 34. It represents an interval of nearly 20% of variation between the simulations, reflecting the probabilistic nature of the model: Random assigning of the level of defense to a computer, random choice of attacked computers and random chance of getting infected. It means that from one simulation to another, the result can be quite different. It corresponds to the reality where we also observe disparities. Some populations are better educated in digital literacy than others, which leads to different outcomes when a computer virus outbreak happens.

The first sharp rise is the consequence of the occurrence of the virus. As the population has never been exposed to this virus, the computers get infected easily even though the level of attack is low. After some time, the computers and the users start to learn about the attack and how to prevent the contagion.

As more and more people learned about the problem with the virus and how to avoid it, the number of infected computers decreased as a reaction to the preceding fast spread of the virus. The small increase at day 90 happened due to the computers that have not been exposed to the virus previously. This phenomenon can occur also in the spread of contagious diseases such as the waves of coronavirus that the world experienced recently.

The medium scale simulation was run in a longer time period to investigate in detail evolution of the number of infected computers, but the period was not sufficiently long to see any trends or equilibria. Will the number of infected computers decrease until every machine is immune or can a small number of computers remain infected over time?

### *C. The long-term simulation*

To look for potential trends and to answer the question of existence of an endemic equilibrium or perpetual oscillations that we experienced with Covid-19 we ran the long-term 650-day simulations. The number of computers in this type of simulation was also set to 100.

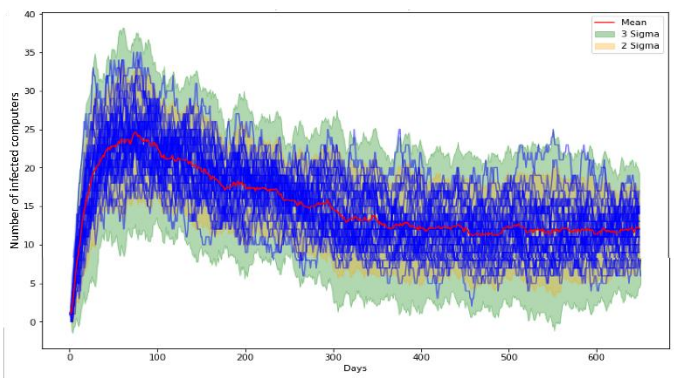



Fig. 4. Number of infected computers in the long-term simulations

The curve of the number of infected computers displayed in Fig. 4 suggests that network immunity is almost impossible in our model. Close to the end of the simulation period of this experiment, the number of infected computers oscillated around the endemic equilibrium of 13 infected computers which corresponds to 13% of the total population.

One interesting phenomenon we could observe is that the number of infected computers dropped over time. It happened because the computers gained a level of defense after they got recovered. However, the level two and level three attacks still were successful. This corresponds to the real situation in companies and society in general. Even though people learn from mistakes they may not resist new types of threats. If one thought about preventing the spread of computer viruses, it would cost money on training people, better computers, etc. Because the amount of money necessary for that would be huge, companies accept some risk of being hacked. This is also reflected in Fig. 4. Training and preventive measures help people resist low level attacks but not sophisticated attacks.

## V. Further Work

To improve the performance of the model and obtain better simulation results, we may think about including more aspects to the model. One of them could be a dynamic population where severe attacks could lead to the “death” of a computer and its removal from the network, and new computers can be added to the network as people buy them when the old ones are not powerful or fast enough.

Another improvement could be increasing the number of levels of defense and attack to have a more precise and complex model. As we could see, three levels are already a good representation but adding one more level could reveal more subtle behavior during the simulation.

We could also experiment with special scenarios, for example a sleeping virus waking up at a specific time on all machines. In a scenario like that we could simulate whether the population would have time to adapt to infection as we observed for the constant exposure to it. Would such an outbreak of contagion suppress the adaptation or will there still be some computers that could resist the attack and adapt themselves?

## VI. Conclusion

In this study, we developed a model of the spread of computer viruses in a network. By implementing the model, we were able to simulate the spread of infection through a certain number of computers that are in the network within a certain period of time. Based on the simulation results and the discussion, we can conclude that the developed model is in compliance with the reality. After a series of successful attacks, people start to learn and react accordingly to the contagion. This effect then makes the number of infected computers, after some fluctuations where the numbers rise a little, to oscillate closely around an endemic equilibrium. We may link the existence of this equilibrium to the fact that for economical and practical reasons, companies and the society accept a reasonable risk to get hacked. However, the best option always is to avoid being hacked. To prevent a computer virus infection, a user can follow the recommendations presented in the report [3].

## Acknowledgment

The work of the third author was supported from the project MeMoV II, No. CZ.02.2.69/0.0/0.0/18 053/0016962 funded by the European Union, the Ministry of Education, Youth and Sports of Czech Republic, and Brno University of Technology. This support is gratefully acknowledged.

## References

[1] Cobalt. (2023). Cybersecurity statistics for 2024. https://www.cobalt.io/blog/cybersecurity-statistics-2024

[2] Pew Research Center. (2017). Americans and cybersecurity. https://www.pewresearch.org/internet/2017/01/26/americans-and-cybersecurity/

[3] Verizon. (2021). Data breach investigations report (DBIR). https://www.verizon.com/business/resources/reports/dbir/

# Design of a device for measuring the electrical resistance of soldered joints during fatigue tests

Bc. Pavel Kopřiva
Author
*Brno University of Technology*
Brno, Czech Republic
220879@vut.cz

Ing. Alexand Otáhal Ph.D.
Supervisor
*Brno University of Technology*
Brno, Czech Republic
alexandr.otahal@vutbr.cz

***Abstract* — This document deals with design of device capable of measuring electrical resistance of solder joints during fatigue testing. It contains theory behind fatigue testing, chosen parameters, design, and construction of the measuring device.**

***Keywords* — *Fatigue tests, Resistance measurement, Solder joints, BGA package, Design, DPS***

## I. Introduction

Solder joints can be found in almost every single electronic device. They are facilitating the connections between electronic components and the printed circuit boards. Because of that, their state is critical to the correct function of these devices. In consumer grade products the solder joint reliability is not a big concern, because their function is not critical. Solder joint reliability is important for example in areas where there is a risk to the human life. Areas like medical devices, aircrafts, spacecrafts, or military equipment.

These devices must be compliant with certain types of standards. These standards set specific fatigue tests that these devices must withstand. The tests for solder joints consist of temperature cycling and mechanical stress. Temperature cycling targets the difference between thermal expansion of given component and printed circuit board. This difference will create sheer forces on the solder joints. These forces are changing with temperature and over time they lead to cracks in the solder joints which will in time lead to failed joint.

Mechanical stress tests are similar in the fact that they are also applying force to the solder joints until they break, for example shear strength testing. Additionally, there are vibration tests, where the sample is exposed to different kinds of vibrations for specified amount of time.

To determine if the solder joint has failed, its electrical resistance is measured. If the resistance rises above given threshold for given amount of time, the joint is considered faulty. The nominal resistance of typical solder joint is in order of hundreds of microohms. The faults definitions are set by standards. Every standard has different values, but they generally agree on the fact that failure state is achieved when the joint resistance momentarily rises above 1 000 Ω, or 20 % rise in value over several measuring cycles.

The measuring device must be able to measure resistance value from hundreds of microohms to kiloohms. The required fault duration changes based on the used standard.

## II. Standards

Thermal cycling is defined by cyclically rising and lowering of temperature in certain amount of time. The temperature change has a defined profile specified by each standard depending on its class. For example, standard ČSN EN 621377-4 requires that the temperature should swing between 125 and -40 °C [1]. The fault is defined as resistance rise above 1000 Ω for 100 ms [1]. Different standards like IPC-9701A or MIL-STD750 use different temperature ranges but they have the same definition of faulty joint.

Mechanical stress testing standards like ČSN EN 62137-3 are defining the theory behind these tests and set the standards of how these tests should be performed [2]. Failure is also determined by increase in resistance. The difference is that the time of that failure is shorter. For example, in drop tests, if there is a crack in solder joint it can be pushed together in normal state and have low resistance, but when the board is dropped, the joint can be disconnected for short amount of time. Strictest standard JESD22 B111 sets the time to 200 ns [3]. This means that the measuring device must be able to facilitate its measurements in that given amount of time.

## III. Design requirements

Firstly, we need to determine the parameters which we will try to achieve with our design. We will base our design around the ball grid array packages and solder joints. These so called BGA packages are more susceptible to fatigue tests because of the high density of solder joints. The measurements are being done on specially designed BGA packages which have their inner connections similar to the example shown in Fig. 1. This setup in combination with printed circuit board - PCB, will create different areas of the BGA where the solder joints will be connected in series. This will allow for checking multiple solder joints at once. This also means that the base resistance will be higher. We will need to create device that will be able to measure resistance in range of 1 mΩ to 1 kΩ with ability to detect changes in hundreds of microohms.

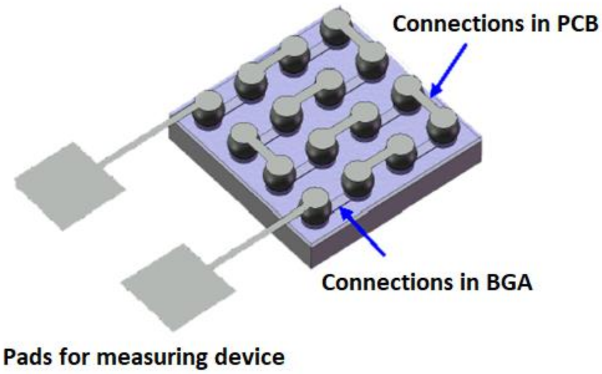



Fig. 1. Example of the test setup for measuring electric continuity [4]

Currently available device was constructed as a master thesis by Ing. Tomáš Vejmola and is only meant to be used for temperature cycling. The device is capable of checking 40 channels at once but is only able to determine if the connection is continuous or not. Therefore, we will try to create a device capable of measuring the same number of channels and provide actual value of electrical resistance. Furthermore, the original device is able to check connection every 50 ms, this is not suitable for mechanical vibration and drop tests. The strictest standard specifies that sample must be taken every 100 ns. That is the target that we will be trying to achieve.

## IV. RESISTANCE MEASUREMENT METHODS

There are multiple ways of measuring electrical resistance. We need to measure resistance changes of around one hundred microohms. This resistance is similar to the resistance of the connecting wires. This means that we cannot use basic methods like two wire method.

First suitable method for measuring resistance in this range is the Kelvin measurement method, which is also known as four wire method. This method uses separate wires for measuring voltage and delivering current, which eliminates the resistance of these wires. This is the most widely used method for measuring resistance in similar ranges.

Alternatively, we can use methods which are using measuring bridges. These methods are usually more accurate, but they are also more complicated. Because we do not need to know the exact accurate value of the resistor, we can use the four wire method.

## V. DESIGN OF RESISTANCE METER

First step in starting the design of our device was to figure out its basic topology. This topology is shown in Fig. 2. Because we decided to use four wire method for measuring resistance, we need a way to supply current to the measured joints and then a way to measure voltage on these joints. Voltage will be measured using analog to digital convertors, aka. ADCs. Because ADCs are expensive devices, we will have to limit their numbers. This mean that we will have to switch between inputs with the use of multiplexors (MUX). The input voltage from the solder joints will be too low for the ADCs. Solution to that problem is to have an input amplifier that will amplify the signal to the correct range.

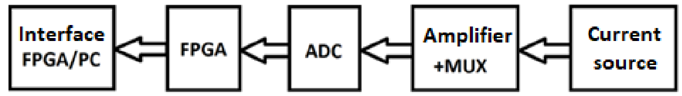



Fig. 2. Topology of the propposed device

After the ADCs, there will be need to process the digital signal. This action will be done using field programmable gate array - FPGA. The data will then have to be transmitted to a computer for analysis.

### A. *Current source*

The easiest way to achieve this function is to use some type of driver for light emitting diodes (LEDs). Value of the current was chosen to be 20 mA. LED drivers are ideal for our purpose because they are current sources, which is exactly what we need. We have picked MIC2843AYMT-TR. It is a 6 channel, 20 mA current source with standard current deviation of ±1,5 % [5]. This deviation could have some sway in the accuracy of measurement, but we have decided that this will not be causing many issues, because we mainly need to measure if the value has risen above preset level. The actual accurate value of resistance is less important.

### B. *Amplifier and multiplexor*

The input signal value varies in the order of microvolts. Input ranges of ADCs are commonly somewhere around one volt. Because of that we will need to amplify this signal. The other thing is that ADCs with higher sampling rates, which we will need, have fully differential inputs. Our input signal is not fully differential. We are using current sources which are meant to be used with LEDs. This means that the solder joints will be connected between open drain input of the current source and power supply voltage. This means that we have to not only amplify it, but also convert it.

One of the demands is the ability to capture changes in input signal that only last for 100 ns. This means that our input amplifier must have high enough gain bandwidth to be able to amplify this signal. We chose to use instrumental amplifiers as input amplifiers. They are easily implementable because the input signal can be routed directly into the instrumental amplifier. In the end we chose INA849DR. It has 8 MHz gain bandwidth at gain setting of 100, input offset voltage of 35 μV [6].

However, the output of the chosen instrumental amplifier is not fully differential and therefore we need to add another stage. This gives us an opportunity to select amplifier with a feature called voltage clamp. This feature allows us to limit the output voltage of this amplifier to protect the inputs of our AD convertors. This function is needed in case of total failure of the solder joint. When this happens the voltage spikes up to the supply voltage of the amplifiers which the ADC input pins are not able to withstand. We chose amplifier LMH6553SD. This amplifier has 900 MHz of gain bandwidth and more importantly has only 10 ns settling time [7].

There are two main ways how to facilitate the input signal multiplexing. First way is to put first stage amplifiers directly at the input and put the multiplexors behind them as shown in Fig. 3. Alternatively, the multiplexors can be placed in front of the first stage amplifiers. Example of this is shown in Fig. 4.

The second option reduces the amount of amplifiers needed which also drastically reduces the cost of the device. Our input signal is not a stiff source. When we will try to multiplex it, it will take a longer time to settle. That means that the multiplexors cannot be switched as fast as they could be switched in the first configuration.

Theoretically it might not be an issue but there is a possibility that we will not be able to reliably multiplex these signals at the maximum required frequency. However, it provides us with lower price for the complete device. Booth options were considered, and the second option was chosen in the end. This was done because we require the high number of channels mainly in temperature cycling tests. These tests do not require fast measurements. Meaning that for thermal cycling, these variants are the same. The high frequency measurements are required only for vibration and drop tests. In worst case scenario we will not be able to reliably multiplex the input signals and we will be left with only direct channels. Meaning that the number of channels will correspond to the number of AD convertors. We chose to install enough of direct channels to fulfil our needs.

We picked multiplexor SN74CBT3251PWR. It is eight channel multiplexor with 7 Ω resistance in open state and 7 ns switching time [8]. This means that every one AD convertor will be able to sample from eight different channels.

### C. Analog to digital convertors

Analog to digital convertors are usually the most expensive parts of any design. That means that the price will be one of the main deciding factors. We found ADC3244IRGZT which is fulfilling all our requirements. It is 14 bit, 10 to 125 MSPS (mega samples per second), pipeline ADC which has two channels in one package [9]. This ADC has serial LVDS data outputs (low voltage differential signal), which are ideal for direct use with FPGA devices. We will use four of these ADCs which means that we will have 64 input channels and 8 direct input channels.

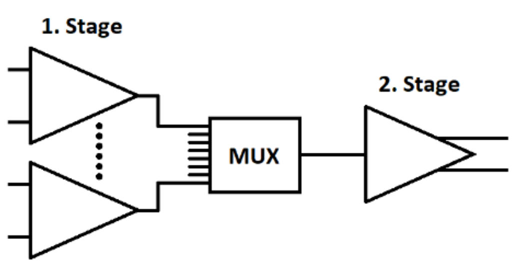



Fig. 3. First way of organizing amplifiers and multiplexors

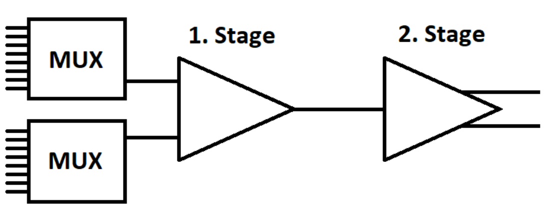



Fig. 4. Second way of organizing amplifiers and multiplexors

### D. Field programable gate array

We will be using FPGA as the main controller for the whole device. We chose Intel MAX 10 (10M08DCU324C8G) FPGA. This FPGA fulfils all our requirements. It is capable of interfacing with the chosen AD convertors. It also has internal program memory, so it can be used as standalone device without any other components which is perfect for our application.

### E. Itrerface between FPGA and PC

We need a way to transmit digital data signals from the FPGA into a PC. We chose the USB (Universal serial bus) interface supported by the EZ-USB FX3 chip from company Infineon (CYUSB3014-BZXC). This chip contains a microprocessor core which is used for converting data incoming from 32-bit parallel data bus (from FPGA) into the USB 3.0 standard [10].

Because we need to transfer the data at high data rate, when there are more channels detecting faults at the same time, we will designate this one USB port only for transmitting data. We still need to transmit data from the PC to the FPGA. For example, information about the start and stop of the measurement and configuration. Luckily the EZ-USB FX3 supports UART (Universal asynchronous receiver-transmitter) for communication and the company Infineon provides chip that works as a bridge between USB and UART. It is named CY7C65215 32LTXIT.

In the end our device will have two USB C ports. One will be routed for USB 3.0 speeds and will be used only for streaming data to the PC. The other one will be at USB 2.0 speeds and will be used for programming and control of the measurements.

## VI. Printed Circuit Board

Final design of the PCB can be seen in Fig. 5. The board is 224.4 mm by 178.8 mm and has six layers.

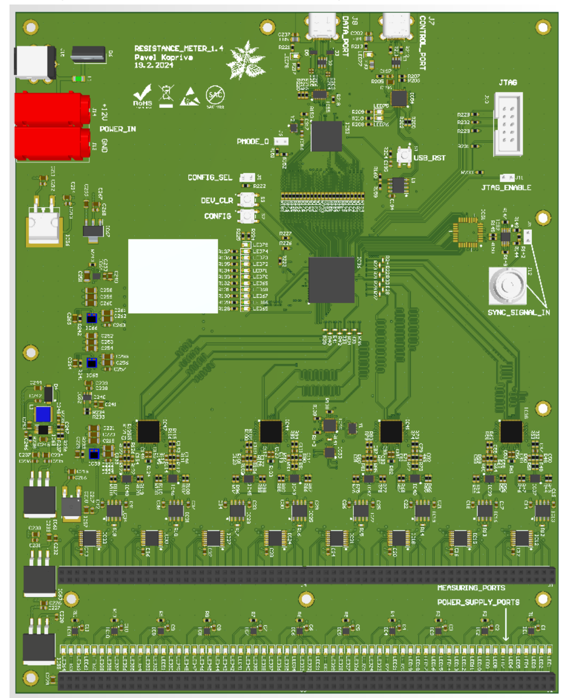


Fig. 5. Final printed circuit board design

## VII. Testing

As of the time of writing, the device is currently under testing. So far, we have been able to confirm the functions of the analog frontend. This means that the current sources, multiplexors, first and second stage amplifiers are working correctly.

We found that the first stage amplifiers are letting thru a small amount of current thru their inputs, which is enough for the LEDs to turn on and shine with reduced intensity. This means that in the event of total solder joint failure the LED will not go completely dark when the amplifier is connected to it via the multiplexors. This is undesired effect, however it is not interfering with the function of our device.

What is interfering with the function of our device is the FPGA to PC interface. Currently we were not able to make the EZ-USB FX3 chip to answer via the USB bus. There are multiple reasons why this might be happening, but we do not know exactly at this moment in time.

The FPGA and ADCs are working correctly to the best of our knowledge but because we are not able to transmit the data from the FPGA to the PC, we cannot confirm it for sure yet.

## VIII. Conclusions

This work deals with the design of a device for measuring the electrical resistance of solder joints during fatigue testing. Designed printed circuit board is shown in Fig. 3. The start of this work deals with the motivation behind construction of this device and provides basic information about the fatigue tests with their corresponding standards. Next chapters follow with the design of our device. First, we design the main topology and then we dive into each and every design block. The last chapter deals with the testing of our device.

The device is currently under testing. We have identified an issue with the FPGA to PC interface which is limiting out ability to thoroughly assess the entire system. We can confidently say that the whole analog part of our design is working correctly as designed. The ADCs and FPGA are also working correctly as far as we can test. The main thing now is to debug the PC interface and figure out if the device as a whole works as expected.

The material price of the components and printed circuit boards is somewhere around 750 euro. This price is not including the countless hours of time that went into the design of this device. If we look at commercially available solutions with comparable parameters, they start from around 3 000 euro. Even thou we are unable to state the exact price of our device, it will be considerably cheaper.


## Acknowledgment

I would like to thank my supervisor Ing. Alexand Otáhal Ph.D. for effective, professional, and methodical help and for providing valuable advice.

I would also like to appreciate Ing. Pavel Tomíček for his willingness and professional help in the design of the final solution.


## References


[1] ÚŘAD PRO TECHNICKOU NORMALIZACI, METROLOGII A STÁTNÍ ZKUŠEBNICTVÍ. ČSN EN 62137-4, Technologie montáže elektroniky - Část 4: Metody zkoušek trvanlivosti pro pájené spoje povrchově montovaných pouzder s vývody typu plošné pole. [online]. Česká agentura pro standardizaci 2015. Availible from: ČAS, https://csnonlinefirmy.agentura-cas.cz/ [cit. 2023-11-05].

[2] JEDEC SOLID STATE TECHNOLOGY ASSOCIATION. JESD22-B111A, Board Level Drop Test Method of Components for Handheld Electronic Products. 2016.

[3] ÚŘAD PRO TECHNICKOU NORMALIZACI, METROLOGII A STÁTNÍ ZKUŠEBNICTVÍ. ČSN EN 62137-3, Technologie montáže elektroniky - Část 3: Směrnice pro volbu metod zkoušek vlivu prostředí a zkoušek trvanlivosti pro pájené spoje. [online]. Česká agentura pro standardizaci 2012. Availible from: ČAS, https://csnonlinefirmy.agentura-cas.cz/ [cit. 2023-11-05].

[4] Understanding Benefits of Daisy Chain. Topline [online]. 2023 [cit. 2023-11-10]. Availible from: https://www.topline.tv/daisychain.html

[5] MIC2843AYMT-TR. MOUSER ELECTRONICS [online]. 2023 [cit. 2023-12-03]. Availible from: https://cz.mouser.com/ProductDetail/Microchip-Technology/MIC2843AYMT-TR?qs=U6T8BxXiZAVfYEbmymh%252Bgw%3D%3D

[6] INA849DR. MOUSER ELECTRONICS [online]. 2023 [cit. 2023-12-01]. Availible from: https://cz.mouser.com/ProductDetail/Texas-Instruments/INA849DR?qs=eP2BKZSCXI4qgDpsz1tg9g%3D%3D

[7] LMH6553SD/NOPB. MOUSER ELECTRONICS [online]. 2023 [cit. 2023-12-02]. Availible from: https://cz.mouser.com/ProductDetail/Texas-Instruments/LMH6553SD-NOPB?qs=7lkVKPoqpbamUxr1qWCjfg%3D%3D

[8] SN74CBT3251PWR. MOUSER ELECTRONICS [online]. 2023 [cit. 2023-12-01]. Availible from: https://cz.mouser.com/ProductDetail/Texas-Instruments/SN74CBT3251PWR?qs=JS6RUWRH9DXd%2F%252BuwaezudQ%3D%3D

[9] ADC3244IRGZT. MOUSER ELECTRONICS [online]. 2023 [cit. 2023-12-02]. Availible from: https://cz.mouser.com/ProductDetail/Texas-Instruments/ADC3244IRGZT?qs=7z%252BmIopC6%2FJv2ZdRPfnA9w%3D%3D

[10] CYUSB3014-BZXC. *MOUSER ELECTRONICS* [online]. 2023 [cit. 2023-12-03]. Availible from: https://cz.mouser.com/ProductDetail/Infineon-Technologies/CYUSB3014-BZXC?qs=xP2cAaN2A3Wt%252BQariFLwRg%3D%3D

# Capacitance Measurement of Building Materials


Autor: Ondřej Skalský
*Department of Control and Instrumentation*
*BUT FEEC*
Brno, Czech Republic
xskals03@vut.cz

Supervisor: Petr Beneš
*Department of Control and Instrumentation*
*BUT FEEC*
Brno, Czech Republic
benes@vut.cz



***Abstract*—This paper presents the basic principles of nondestructive indirect moisture measurement of masonry from dielectric property analysis. For the selected parallel plate method, three methods that have been experimentally tested are presented. The aim of the work is to provide a comprehensive body of theoretical information complemented by an unambiguous procedure supported by experiments for the measurement of permittivity of building materials at different humidities and temperatures, which will serve as a basis for subsequent experimental acquisition of the moisture and temperature dependence of permittivity for selected building materials. The purpose of obtaining these characteristics is to develop and calibrate a low-cost embedded prototype system for continuous moisture measurement of building materials.**



***Index Terms*—Humidity measurement, permittivity measurement, dielectric polarization, paraller plate method, building materials.**


## I. Introduction

High humidity in masonry and buildings brings a number of negative phenomena. Mould thrives on damp masonry, which is neither good for the masonry nor for people. The high water content also impairs the insulating properties. It is therefore advisable to analyze the moisture content. The most accurate moisture measurements are made using direct measurement methods. These evaluate the water content of a given sample. The disadvantage is that the analysis takes place in a laboratory and the sample has to be removed from the site, which is often undesirable. Therefore, there are also indirect moisture measurement methods, which are mostly based on the measurement of the dielectric properties of the material, that are dependent on moisture. Dielectric properties can be measured over a wide frequency range ($10^{-6}$ Hz to $3 \cdot 10^{15}$ Hz [1], [2] ) and there are a variety of methods for measurement.

Methods for measurements in the radio and especially microwave frequency range are based on measurements of the wave properties of the dielectric such as the speed of the propagating electromagnetic wave, s-parameters describing the reflection at the input, output or attenuation of the conductor formed by the dielectric under test [3]–[5].

Some methods require extracting separated test sample to measure or have high error for porous materials. Certain measurement methods are not suitable for materials with large dielectric losses, others are suitable only for liquids or semi-solids. The cost of the instrumentation is also an important factor. An overview of chosen methods is shown in the figure 1 [3]–[5].

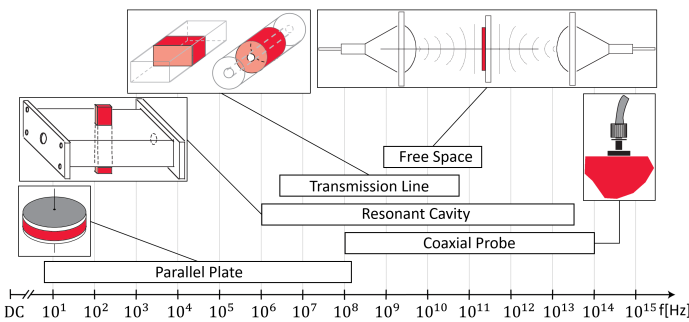



Fig. 1. Example of a figure caption [1], [3], [6]–[8]

We can conclude that for our purposes the paralel plate method will be the best choice. It allows the measurement of solid porous materials and from the measurements it is relatively easy to calculate the complex permittivity of the dielectric. The ability of using this method at low frequencies implies lower demands on the measurement circuits, which makes the measuring instrumentation quite affordable [9].

It should be mentioned that moisture meters for solids in the desired economy price range are mostly resistance moisture meters that evaluate the moisture content of water based on conductivity. However, these suffer from gradual degradation of the electrodes due to the necessity of their contact with the material to be measured. This results in a change of parameters over time and makes them unsuitable for long-term measurements.

## II. Theory

In the following part, dielectric property permittivity will be described as well as polarization principles, which are strictly related to this quantity.

### A. Permittivity

Permittivity $\epsilon$ $[\mathrm{F/m}]$ is a complex quantity describing dielectric properties. For convenience we will work with the relative permittivity $\epsilon_r$ related to the absolute permitivity of vacuum $\epsilon_0 = 8.854 \cdot 10^{-12}\,\mathrm{F/m}$ via $\epsilon_r = \epsilon/\epsilon_0$ $[-]$, which could be decomposed into real and imaginary part

$$\epsilon_r = \epsilon_r' - j\epsilon_r''. \tag{1}$$

The real part $\epsilon_r'$ describes the ability to store the energy of the electric field as a result of polarization, and the imaginary part $\epsilon_r''$ specifies the dielectric loss. Ration of those parts are defined as a dissipation factor $D$ $[-]$

$$D = \epsilon_r'/\epsilon_r''. \tag{2}$$

Note that for a given substance permittivity depends on a number of phenomena, such as the frequency of the excitation field, its intensity, temperature, etc. For the sake of simplicity, however, we will not write permittivity as a function of these quantities. Furthermore, we assume dielectric isotropicity and thus a simplified scalar description of the permittivity is used instead of the general tensor one [1], [9], [10].

### *B. Polarization principles*

Different ways of arranging the electrically charged particles of a dielectric inserted into an external electric field are associated with different types of polarizations. The finite speed of rearrangement of electrically charged particles (i.e. the polarization) causes the permittivity to be frequency dependent [1], [11].

We can notice from the figure 2 that the frequency dependencies of the real and imaginary components are related. At the so-called relaxation frequencies $f_i$ [Hz], local maxima of the imaginary component of permittivity occur and the real component decreases around these frequencies with increasing frequency.

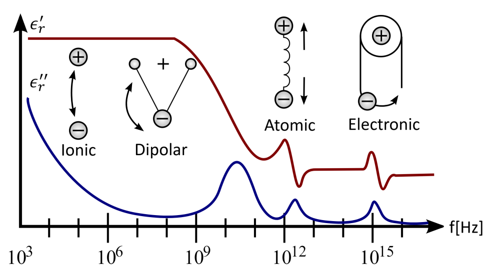



Fig. 2. Frequency dependence of permittivity [11].

The origin of this behaviour is, in simple terms, the same for all polarization mechanisms. If $f \gg f_i$ is valid, where $f$ is the frequency of the external electric field and $f_i$ is the given relaxation frequency, then the given type of polarization takes no effect because the time required to create the given type of polarization is significantly longer than the period of the excitation signal. Conversely, for $f \ll f_i$, the dielectric manages to polarize with a given polarization type and the polarization mechanism contributes to the total real part of the permittivity. For the case $f \approx f_i$, the dielectric is being constantly repolarized. Since this process is lossy due to internal friction, the complex permitivitty increases in the region of these frequencies [6].

**Ionic polarization** - It is the polarization of the dielectric by the accumulation of charges at the boundaries of the sample. This polarization thus occurs only in substances with free charges, i.e. in conductive substances [1], [12].

**Dipole polarization** - In dipole, or sometimes called orientation polarization, the force effects of the external electric field cause the permanent dipoles to align with the field. Permanent dipole is an approximation of electrical effect of a non-changing mutual arrangement of charged particles in a molecule or small group of atoms. Dipoles approximate macroscopic behaviour of those arrangements using a pairs of charges of the same size, opposite polarity and different position generating the same electric field from an external point of view [6].

**Electron and Atomic polarization** - These types of polarization can be imagined as dipole polarization, where the dipoles are induced due to the force effects of an external electric field on charged particles.

If the dipole is induced by the displacement of the electron cloud of an atom or molecule, we speak about electron polarization. In atomic polarization atoms or ions of polyatomic molecules are displaced by an electric field. In both cases, the inner forces of the particles are elastic, which are responsible for a resonant overshoot of the real relative permittivity, as seen in the figure 2 [6], [9].

**Waxwell-Wagner polarization** - This polarization can be observed in materials composed of several substances whose real or imaginary permittivity differs from each other. Individual parts of a given material with a different permittivity than their surroundings behave in this surroundings similarly as dipoles, due to the accumulation of charges at the edges caused by diferent polarizations in neighbouring regions. Hence, this type of polarization is sometimes called as interfacial polarization. The Maxwell-Wagner polarization depends on the size of the elements of the mixture, their shapes and geometrical arrangement and its effect is difficult to model even for binary substances [6], [12], [13].

## III. Measurement theory

Unfortunately, the effect of Maxwell-Wagner polarization on the relative permittivity of $\epsilon_r$ is significant and the required detailed knowledge of the composition for analytical modelling of dielectric behaviour is unachievable in the field of building materials. Besides the strong dependence of this polarization on temperature and humidity, a significant role is played by the dependence of the electrical conductivity of water on the salt content as well. The influence of the sample folding sequence is also a complication in dielectric moisture measurements. The permittivity depends on the previous water content, i.e., whether the material is wetting or drying, which is related to the way water binds to the material structure [1], [6], [12], [13].

It follows from the above that the determination of the moisture content of building materials will have to be based exclusively on the experimentally determined frequency characteristics of the permittivities of the desired materials for the specified temperature ranges and mass fractions of water.

### *A. Parallel plate method*

The Agilent 16451b electrode system with the Agilent 4294A LCR analyzer was used to measure the permittivity using the parallel plate method. The measurement principle is shown in the figure 3. As can be seen, the electrode system

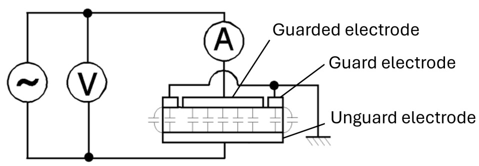



Fig. 3. Measuring principle of the Agilent 16451b electrode system [14].

includes a guard electrode to reduce errors caused by the strain of the electric field at the edges of electrodes. The sample excitation is performed by a harmonic signal. The amplitude of the measured current corresponds to the absolute value of the permittivity of the sample, the currrent phase shift $\pi/2-\delta$ is determined by the dissipation factor $\mathrm{tg}\delta = D$. The LCR meter expresses these measured values in terms of parameters of an equivalent circuit. For our purposes, a parallel equivalent circuit consisting of capacitance $C_p$ and resistance $R_p$ and a serial one consisting of $C_s$ [F] and $R_s$ [$\Omega$] were used [14], [15].

The manufacturer Agilent offers a total of 3 methods for measuring the real relative permeability, which chosen as a quantity to evaluate the humidity. Out of these, Contact and Contactless methods were used. The Contact method was modified to theoretically compensate the error caused by unevenness of the sample while preserving the requirement of performing only one measurement of capacity. We will refer to this modified version as the Contact modified method [14].

### B. Contact method

This is the simplest of the methods. The method assumes that the entire space between the electrodes is filled with the material to be measured. The permittivity is given by

$$\epsilon_r' = \frac{t_a C_p}{\pi (d/2)^2 \epsilon_0 \alpha}, \tag{3}$$

where $t_a$ [m] is the thickness of the sample, $d = 0.038$ m is the diameter of the measuring electrode and $\alpha$ $[-]$ an effective electrode area constant, which values can be found in [14]. The advantage of this method is that there is no need to know the actual electrode distance $t_g$ [m] and thus no need to calibrate the electrode system [14].

In reality, however, there is always an air gap between the electrodes and the sample causing the calculated permittivity to be lower than the actual permittivity. The method error is [14]

$$\delta\epsilon_r' = \frac{1-\epsilon_r'}{\epsilon_r' + t_a/(t_g - t_a)}. \tag{4}$$

### C. Contactless method

The idea of the method is to limit the error caused by the presence of an air gap. The method requires measuring the capacitance of $C_{s1}$ without the presence of the sample and $C_{s2}$ with the sample inserted, keeping the electrode spacing $t_g$ the same for both measurements. From the known electrode distance $t_g$ and sample thickness $t_a$, the permittivity of the material is then calculated

$$\epsilon_r' = \left(1 - \left(1 - \frac{C_{s1}}{C_{s2}}\frac{\alpha_2}{\alpha_1}\right)\frac{t_g}{t_a}\right)^{-1}. \tag{5}$$

The coefficients $\alpha_1$ and $\alpha_2$ correspond to the measurements of the capacities $C_{s1}$ and $C_{s2}$ [14].

### D. Contact modified method

This method, like the non-contact method, associates the measurement with an equivalent circuit consisting of a series combination of a pair of fictive capacitors, where one has a dielectric consisting only of the sample under test and a thickness $t_a$ and the other fictive capacitor has an air dielectric with a thickness $t_g - t_a$.

Compared to the contactless method, only one measurement of the series combination $C_s$ of these fictitious capacitors needs to be performed. Based on the known dimensions, the capacitance of the fictive air capacitor can be calculated and from the measured capacitance the capacitance of the fictive capacitor consisting of the sample under test can be calculated. From the known capacitance and dimensions of this fictitious capacitor, it is then easy to calculate the real relative permittivity of the sample

$$\epsilon_r' = \frac{C_s t_a}{\epsilon_0 \epsilon_{air}' (\pi d/2)^2 \alpha_a - C_s(t_g - t_a)} \frac{\alpha_a \epsilon_{air}'}{\alpha_b}. \tag{6}$$

## IV. Experiments

The aim of the experiments was mainly to verify the theoretical assumptions and to provide a basis for future more extensive measurements. The experiments were carried out on three types of materials: cetris, plasterboard, gypsum. On the basis of the findings, a manual [16] was written describing the procedure for measuring the real relative permittivity of the sample, and measurement software was also developed to automate and speed up the measurements. These materials together with the theoretical assumption provide the basis for upcoming extensive measurements of conditioned samples in a humidifier at a given temperature to obtain detailed dependencies by a defined procedure.

### A. Comparison of measurement methods

Both the non-contact method and the contact modified method should compensate completely the air-gap error. However, permittivity calculations according to these methods sometimes fail. The singularity occurring in the equations 5 and 6 causes that for higher values of permittivity ($\epsilon_r' \gtrsim 50$) the results are strongly affected by the accuracy of the measured values. Measurement error can even cause the results of these methods to be negative.

Due to this inconvenience, these methods were abandoned and for further measurements it is recommended to use only the contact method. The error of the contact method has to be eliminated using samples with the best possible planparallelism and smoothness of the surfaces to minimize the air gap and therefore the error it causes.

An illustrative example of the frequency response of the real relative permittivity with the dissipation factor of humid gypsum obtained using the contact method together is given in figure 4 below.

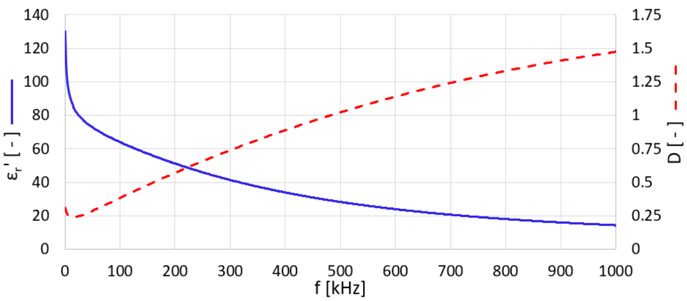



Fig. 4. Measured frequency dependence of real relative permittivity and dissipation factor for humid gypsum.

During the experiments, the Analog Devices AD7746 capacitance meter was tested as a candidate for a low-cost capacitance meter. However, as it turned out, this capacitance meter in fact measures the absolute value of impedance and converts it to capacitance according to the used frequency, assuming that the measured impedance is purely capacitive. The measured capacitance with this circuit comes out $D + 1$ times higher than the actual one, which according to the figure 4 is a non-negligible error for any chosen frequency in our desired span.

### B. Procedure for upcoming measurements

The measurement procedure presented by the manufacturer in [14] has been slightly modified and supplemented with additional steps, such as sample thickness measurement methodology, electrode system preparation, sample weight measurement, sample conditioning, and more. The purpose of making such a manual [16] is to ensure repeatability and traceability of the experiments performed in future measurements.

### C. Measuring software

To measure the data efficiently, measurement software in NI LabVIEW was developed. Using the Agilent 4294A NI library the Agilent 4294A is remotely operated trough NI-VISA via an Ethernet interface. The sotfware allows the user to remotely set the analyzer parameters, perform measurements and calculate permittivity values according to the specified methods. Afterwards, all measured quantities including calculated permittivities are stored in the resulting file.

## V. Conclusion

The theory associated with parallel plate permittivity measurements were summarized in this paper. Experiments excluded both methods that should eliminate the error caused by the air gap due to the inconsistent and unreliable results and thus only the contact method will be used for further measurements.

A procedure in [16] has been created with a detailed description of the individual measurement steps, which will serve as a reliable basis for subsequent large-scale permittivity measurements to obtain the dependence of permittivity on moisture and temperature of the desired materials.

The goal is to obtain the dependencies of permittivity on humidity and temperature for the given materials followed by selecting a suitable and affordable measuring circuit that can measure the permittivity of the material with the required accuracy. As mentioned, it is necessary that this meter actually measures capacitance, not absolute impedance.

The final step will be to create a custom electrode system, complement the capacitance meter with control circuits and create an embedded device for moisture measurement based on the obtained dependencies and test this device in use.

## References


[1] KREMER, F. a A. SCHÖNHALS. "Broadband Dielectric Spectroscopy," Springer, 2003. ISBN 978-3-642-56120-7.

[2] VATANSEVER, Fatma a Michael R. HAMBLIN. "Far infrared radiation (FIR): Its biological effects and medical applications", 2012. Available: doi:10.1515/plm-2012-0034

[3] "Basics of Measuring the Dielectric Properties of Materials: Application Note," USA: Keysight Technologies, 2017. Available: https://www.cmc.ca/wp-content/uploads/2019/08/Basics_Of_MeasuringDielectrics_5989-2589EN.pdf.

[4] VENKATESH, M.S. a G.S.V. RAGHAVAN. "An overview of dielectric properties measuring techniques," CANADIAN BIOSYSTEMS ENGINEERING, 2005. Available: https://library.csbe-scgab.ca/docs/journal/47/c0231.pdf.

[5] NELSON, Stuart O. "Dielectric Properties of Agricultural Materials and Their Applications," 1st U.S. Department of Agriculture, USA: Academic Press, 2015. ISBN 978-0-12-802305-1.

[6] "Agilent Basics of Measuring the Dielectric Properties of Materials: Application Note," USA: Agilent Technologies, 2006. Available: https://academy.cba.mit.edu/classes/input_devices/meas.pdf.

[7] TERESHCHENKO, O. V., F. J. K. BUESINK a F. B. J. LEFERINK. "An overview of the techniques for measuring the dielectric properties of materials," 2011 XXXth URSI General Assembly and Scientific Symposium [online]. IEEE, 2011, 2011, 1-4. ISBN 978-1-4244-5117-3. Available: doi:10.1109/URSIGASS.2011.6050287

[8] EL KHALED, Dalia, Nuria CASTELLANO, Jose GÁZQUEZ, Alberto-Jesus PEREA-MORENO a Francisco MANZANO-AGUGLIARO. "Dielectric Spectroscopy in Biomaterials," 2016, 9(5). ISSN 1996-1944. Available: doi:10.3390/ma9050310

[9] KWAN, Chi Kao. "Dielectric Phenomena in Solids: 2 - Electric Polarization and Relaxation ," San Diego: Academic Press, 2004. ISBN 9780123965615. Available: https://www.sciencedirect.com/science/article/pii/B9780123965615500128.

[10] "Standard Test Methods for AC Loss Characteristics and Permittivity (Dielectric Constant) of Solid Electrical Insulation1: Designation: D 150 – 98", ASTM International, 1998.

[11] XIAOFAN, Yang, Liu XIAOMING, Yu SHUO, Gan LU, Zhou JUN a Zeng YONGHU. "Permittivity of Undoped Silicon in the Millimeter Wave Range."

[12] CHEN, Yongping a Dani OR. "Water Resources Research," 2006, 42(6). ISSN 00431397. Available: doi:10.1029/2005WR004744

[13] CHEN, Yongping a Dani OR. "Effects of Maxwell-Wagner polarization on soil complex dielectric permittivity under variable temperature and electrical conductivity," Water Resources Research. 2006, 42(6). ISSN 00431397. Available: doi:10.1029/2005WR004590

[14] "Keysight 16451B Dielectric Test Fixture: Operation and Service Manual," Keysight Technologies, 2008.

[15] "Agilent 4294A Precision Impedance Analyzer: Operation Manual," Japan: Agilent Technologies, Kobe Instrument Division, 2001.

[16] SKALSKÝ, Ondřej. "Kapacitní měření vlhkosti stavebních materiálů," Unpublished.

[17] MANCL, Vlastimil. "Systém Hefaistos," Brno, 2019 [cit. 2023-08-14]. Available: https://www.vut.cz/www_base/zav_prace_soubor_verejne.php?file_id=189915. VUT Brno FEKT.

# Experimental audio effect based on dynamic signal filtering

David Leitgeb
*Department of Telecommunications*
*Brno University of Technology, Technicka 12*
616 00 Brno, Czech Republic
xleitg03@vutbr.cz

***Abstract*—This paper deals with an audio effect that utilizes multiple types of digital signal processing to creatively produce various sound colors with musical signal as its input. Used signal processing techniques include: frequency filters, delay line with signal interpolation, low-frequency oscillators. The paper includes description of the structure of the proposed audio effect, approaches used during the implementation process and several examples of the functionality of its individual features. The proposed effect was developed using Matlab and its Audio Toolbox extension.**

***Index Terms*—digital audio effect, frequency filtering, delay line, low-frequency oscillators, Matlab**

## I. Introduction

The increasing amount of available digital audio effects and technological progress in general has paved the way to music production for many people, especially in the last few years. A few decades ago, in order to make music that sounded professional, the only way to achieve this goal usually was to visit a studio with a lot of expensive hardware. Today all one needs is a laptop with the right software, at least for some music genres. For the most part this is possible thanks to the introduction of Virtual Studio Instrument (VST) by Steinberg in 1996 [1]. This popular audio plug-in interface makes it really simple to use large amounts of various audio effects from different manufacturers in the users favorite host application, also called Digital Audio Workstation (DAW). Thanks to this technology (and of course other plug-in formats) it is possible to fit almost the whole studio in just one computer.

The main idea behind the development of the proposed plug-in device was to build an audio effect that could be useful when trying to create interesting colors of sound using a combination of several types of digital signal processing. The main inspiration for the proposed device was Spektral Delay by Native Instruments introduced in 2001, although there are some fundamental differences. Spektral Delay used Short Time Fourier Transformation (STFT) to divide the input signal into up to 1024 frequency bands [2]. The designed effect uses regular infinite impulse response (IIR) filters to divide the input signal into 5 separate bands (plus one direct band with no filtering).

## II. Effect structure

The audio effect plug-in presented in this paper was developed using MathWorks' Matlab (version R2023b) and its Audio Toolbox extension (version 23.2). The effect consists of multiple types of signal processing: frequency filters, delay line with signal interpolation and low-frequency oscillators. The order of processing performed on the input signal is shown in Fig. 1. After splitting the input signal into each band, the first performed processing is frequency filtering. Proposed implementation allows the user to process the same input signal with arbitrarily chosen filter – the input signal is not split into fixed frequency bands, therefore the outputs of the individual bands can overlap in the frequency domain. Some details about the implementation of filters is described in III-A. After the signal is filtered, it is processed by a delay line. Closer description of its implementation is further explained in III-B. These two operations are only done for bands 1–5. The last two parameters that can be altered for all of the individual bands (including the direct band) are panorama and gain. After this processing, the signals from all of the bands are summed together and sent to the output of the effect with the option to control the final output gain. The last component left to mention are the two low-frequency oscillators (typically referred to as LFOs) located at the bottom of Fig. 1. As seen here, the output signal of these oscillators can modulate

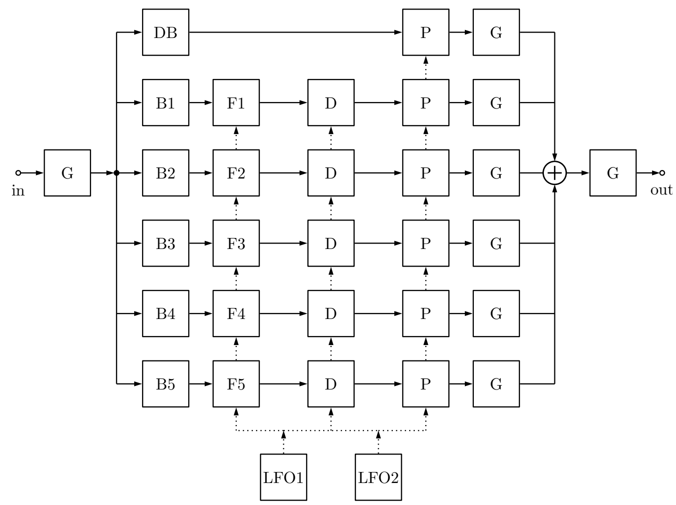



Fig. 1. Signal flow of the designed audio effect, where G = Gain, F1–5 = frequency filters, D = delay line, P = panorama, LFO1 & LFO2 = low-frequency oscillators, DB = direct band and B1–5 = bands 1–5 .

specific parameters such as cutoff frequency of the filters, delay length for bands 1–5 and panorama for all of the bands. The user can define the type of signal generated by the LFOs by adjusting parameters such as the signal shape (sine wave, triangle wave, square wave, saw wave, sample&hold), the rate (expressed in both Hz and note lengths) and the amplitude of the generated signal (sometimes referred to as depth in this context).

## III. Implementation

This section briefly describes the approaches used to implement the most crucial parts of the presented audio effect plug-in: filters and delay line with linear interpolation. Source code and the VST3 version of the audio effect plug-in generated by Matlab can be found in [3].

### A. Filters

One of the key components of the designed effect are frequency filters. In terms of digital frequency filter implementation, there are several types to choose from, each with their own advantages and disadvantages. The kind of filter used for the purpose of the proposed audio effect plug-in is classified as infinite impulse response (IIR) filter, which is one of the properties of linear time-invariant (LTI) systems. Digital audio effects often use IIR filters in their implementations, mainly for lower required memory used to store coefficients. Compared to finite impulse response (FIR) filters, IIR filters also need less computing power, which makes them more suitable to use for the purpose of the proposed effect [4]. Choosing the appropriate kind of filters can have a major impact on the overall computing latency when combining higher number of filters and other components together.

The actual implementation of the filters is done by calculating a set of coefficients that correspond to the desired filter type. Equations used to calculate these coefficients can be found in [4]. They vary depending on the filter type and order. The results of these calculations are used in the following transfer function [4]:

$$H(z) = \frac{b_0 + b_1 z^{-1} + b_2 z^{-2}}{1 + a_1 z^{-1} + a_2 z^{-2}}, \tag{1}$$

where $b_0$, $b_1$, $b_2$, $a_1$ and $a_2$ are the calculated coefficients. This transfer function is used for both first-order and second-order filters, the only difference is $a_2$ and $b_2$ is set to zero for first-order filters. All of the coefficients are used as one of the inputs in Matlab's `filter` function, which accepts them alongside with the input signal and initial conditions of the given filter. Received output is the filtered signal and final conditions of the filter. Transfer of the initial and final filter states is crucial when the signal is processed sample frame by sample frame in real-time processing.

The audio effect plug-in offers seven different types of filters the user can choose from (see Fig. 2), all of them resulting in different spectral characteristics of the output signal:

- highpass (HP),
- low-shelving (LS),

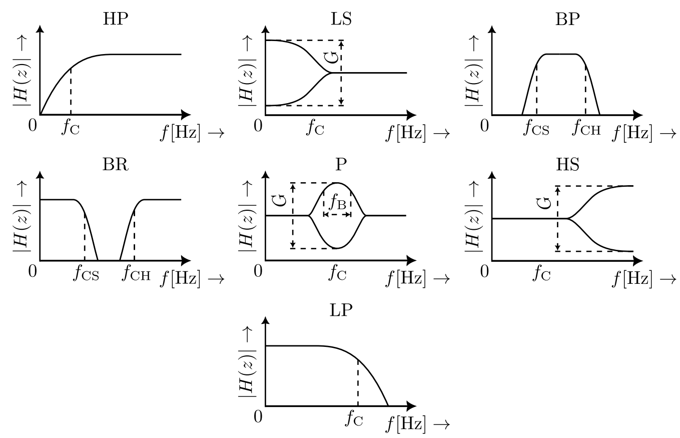



Fig. 2. Filter types available in the audio effect plug-in together with their parameters such as cutoff frequency ($f_C$), lower ($f_{CS}$) and upper ($f_{CH}$) cutoff frequency, gain ($G$), bandwidth ($f_B$).

- bandpass (BP),
- bandreject (BR),
- peak (P),
- high-shelving (HS),
- lowpass (LP).

All of the listed above are implemented as second-order filters. In the case of the highpass and the lowpass filters, the user can also switch to the first-order implementation, which results in gentler slope of the frequency curve. As already mentioned, the cutoff frequency of the individual filters can be modulated by low-frequency oscillators which allows for dynamic changes in the frequency spectrum of the output signal. The modulation is performed on a sample-by-sample basis, while the cutoff frequency set by the user can be changed between individual audio frames.

### B. Delay line

Another crucial part of the audio effect plug-in is the delay line. In general, audio effects that use delay lines can be divided to effects with either static or dynamic delay times [5]. Structure of the implemented delay line is shown in Fig. 3. It consists of three routes the processed signal can take: flow directly to the output (feedforward), get delayed by *M* samples and then go to the output or after getting delayed get sent back before the delay operator (feedback). There are three

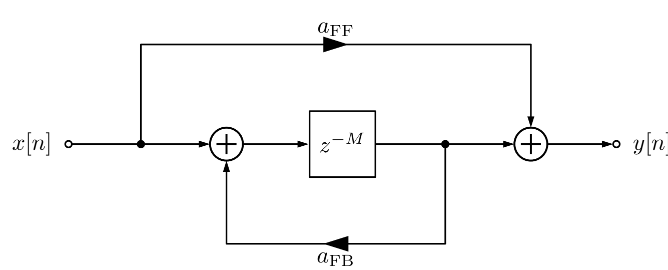



Fig. 3. Structure of the implemented delay line.

parameters that can be controlled by the user: the gain of both the feedforward and the feedback route ($a_{\mathrm{FF}}$ and $a_{\mathrm{FB}}$) and the length of the delay in milliseconds that is afterwards converted to the corresponding amount of samples $M$. As already mentioned, the signals generated by the two LFOs can be used to modulate the delay time – apart from the LFO configuration, the user can also alter parameters such as modulation range (in milliseconds), which is shared for all of the delay lines, but can be further adjusted for the individual bands using the modulation depth parameter (in percent). Depending on the settings of these parameters, it is possible to achieve a number of different effects widely used in all sorts of music. Typical examples with the following settings are [5]:

- **effects with static delay length**:
  - *phase shifter* (*resonator*): delay time $< 25$ ms,
  - *slapback* (*doubling*): delay time 25–50 ms,
  - *echo*: delay time $> 50$ ms,
- **effects with modulated delay length**:
  - *vibrato*: $a_{\mathrm{FF}}$, $a_{\mathrm{FB}} = 0$, delay time modulation range 0–3 ms, typically sine wave LFO,
  - *chorus*: $a_{\mathrm{FF}} \neq 0$, $a_{\mathrm{FB}} = 0$, delay time modulation range 1–30 ms, various LFO waveforms possible,
  - *flanger*: $a_{\mathrm{FF}} \neq 0$, $a_{\mathrm{FB}} \neq 0$, delay time modulation range 1–2 ms, typically sine wave LFO.

In order to delay a digitally stored signal, delay lines use a technique called circular buffer. Its principle is really simple, the incoming input signal is written in the so called circular buffer and after the desired time has elapsed, the same part of the signal is sent to the output. Both operations – writing in and reading out of the buffer are being done at the same time. In order to have a functioning delay line, the circular buffer has to be longer than a regular buffer would be. The length is usually determined by the maximum delay time that is possible to achieve in the given effect – e.g. if the highest value the delay can be set to is two seconds, the length of the circular buffer has to be at least twice the sampling frequency [6]. The term circular buffer comes from the fact that after reaching its last element with either of the two operations, the position of given operation is moved back to the first element, creating an imaginary connection between both ends of the buffer.

When implementing a digital effect that features modulation of the delay parameter, one major issue has to be taken care of. Digital signals are stored as discrete samples, meaning the information of what the signal looks like between these samples is missing. When converting the user set delay parameter from milliseconds to the corresponding number of samples, the result is often not integer. While this is not necessary a problem in the case of static delay lengths which can be simply just rounded to the nearest integer, doing so with dynamic delay lengths will introduce audible artifacts. The solution for this problem is some sort of interpolation. The current version of the presented effect uses linear interpolation which was simple to implement, yet it made a significant improvement in quality of the output signal. The principle of linear interpolation (also known as first-order interpolation) is to blend two consecutive samples in the right proportion [6]:

$$x(t) = (n+1-t)x[n] + (t-n)x[n+1], \quad n \leq t < n+1, \tag{2}$$

where $x[n]$ and $x[n+1]$ are the two interpolated samples and the expressions in round brackets set the correct ratio. Using this equation to calculate the value of the delayed signal results in smoother signal waveform – sample interpolation acts as a lowpass filter suppressing the otherwise audible artifacts caused by abrupt changes in the signal waveform.

## IV. Results

After finishing the development of the audio effect plug-in, several examples showcasing the functionality of its individual features were prepared. Selected ones are included in this section of the paper (see Fig. 4, 5, 6, 7, 8), the rest of the examples with corresponding audio can be found in [3].

The first example is the vibrato effect created by modulating the delay time. Sine wave with the frequency of 1 kHz was used as the input signal. Both the feedforward and feedback routes had their gain set to zero, range of the modulation was set to 3 ms and the rate of the low-frequency oscillator was set to a quarter note. While preparing the examples, the tempo was set to 140 beats per minute (BPM). Spectrogram demonstrating the resulting frequency oscillation is shown in Fig. 4.

The second example is focused on the importance of using some kind of sample interpolation. As already stated, the audio effect plug-in uses linear interpolation to suppress any audible artifacts. Visual comparison of the difference between output signal with and without linear interpolation is shown in Fig 5[1].

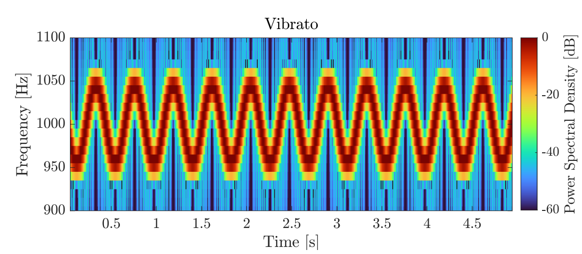



Fig. 4. Spectrogram of the vibrato effect.

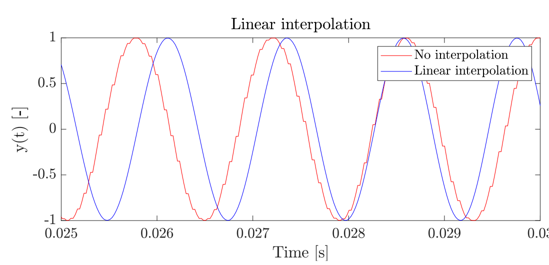



Fig. 5. Comparison of output signal without and with linear interpolation.

[1]Both signals are slightly offset in time, the key feature to look at is the suppression of abrupt changes in the waveform of the interpolated signal.

The third example shows the result of combining the input signal with its delayed copy. In this case, the input signal was white noise and the delay time was set to 1 ms. This combination was done by setting the gain of the feedforward route to 100 %. The result of this type of processing is an effect called the comb filter (see Fig 6), widely known thanks to the chorus and flanger effects commonly used in music.

The last two examples focus on the use of the frequency filters. Fig 7 shows the resulting amplitude spectrum of white noise that was filtered using the bandpass filter. The cutoff frequency of the filter was set to 1 kHz and the filter Q was set to the value of 10. Example in Fig. 8 also uses the bandpass filter to filter white noise, only this time the cutoff frequency (set to 3 kHz) of the filter was modulated using the low-frequency oscillator (set to the triangle wave). The modulation range was set to 1 kHz and the rate of the low-frequency oscillator was set to an eight note.

As already mentioned, a series of examples were prepared to prove the functionality of all of the implemented signal processing types. These include tests of all seven frequency filter types available in the audio effect plug-in, several applications of the delay line and further low-frequency oscillator experiments. All of these additional files (including the source code) can be found in its dedicated GitHub repository [3].

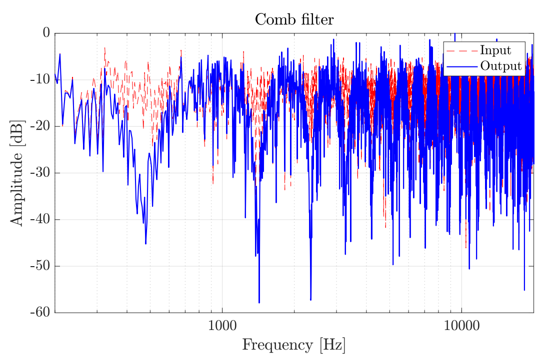



Fig. 6. Amplitude spectrum of the signal before and after the formation of the comb filter.

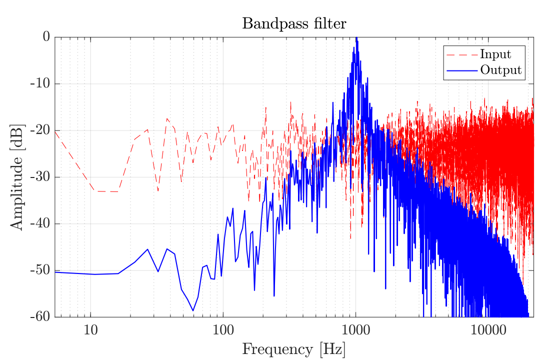



Fig. 7. Amplitude spectrum of the signal before and after being filtered using the bandpass filter.

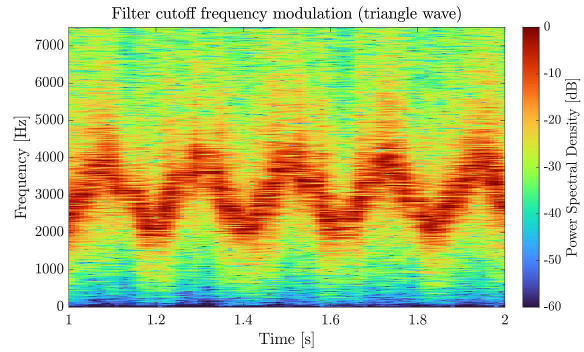



Fig. 8. Filter cutoff frequency modulation.

## V. Conclusion

This paper introduced the audio effect plug-in which was developed to test the basic functionality of the individual types of digital signal processing. The goal of the subsequent work is to create a C++ version of the plug-in with the use of the JUCE framework[2]. This version should feature expanded capabilities of what is possible to create with the effect – mainly due to higher amount of available bands and easier workflow with improved graphical user interface. Another topic that will be explored are advanced sample interpolation techniques.


## Acknowledgment

I would like to thank my master's thesis supervisor doc. Ing. Jiří Schimmel, Ph.D. for guidance during the preparation of this paper and the development of the presented audio effect.


## References


[1] STEINBERG MEDIA TECHNOLOGIES. *Our Technologies*. Online. Available: https://www.steinberg.net/technology/. [cit. 2024-02-29].

[2] NATIVE INSTRUMENTS SOFTWARE SYNTHESIS. *NI-Spektral Delay Operation Manual*. Online. Available: https://ia902802.us.archive.org/35/items/spektraldelaymanualenglish/Spektral%20Delay%20Manual%20english.pdf. [cit. 2024-02-29].

[3] LEITGEB, David. *Experimental audio effect based on dynamic signal filtering*. Online. GitHub. 2024. Available: https://github.com/David-Leitgeb/Experimental-audio-effect-based-on-dynamic-signal-filtering. [cit. 2024-03-09].

[4] ZÖLZER, Udo. *DAFX: digital audio effects*. 2nd ed. Chichester: Wiley, 2011. ISBN 978-0-470-66599-2.

[5] SCHIMMEL, Jiří. *Studiová a hudební elektronika*. Online, PDF. 2. vydání. Vysoké učení technické v Brně, Fakulta elektrotechniky a komunikačních technologií, Ústav telekomunikací, 2015. ISBN 978-80-214-4452-2.

[6] REISS, Joshua D. a MCPHERSON, Andrew P. *Audio effects: theory, implementation and application*. Boca Raton: CRC Press, 2014. ISBN 978-1-4665-6028-4.

[7] RAW MATERIAL SOFTWARE LIMITED. *JUCE*. Online. Available: https://juce.com/. [cit. 2024-02-29].


[2]JUCE is a popular open-source framework created by Julian Storer often used for audio application development [7].

# Bioimpedance Measuring Unit for Smart Agriculture


Alžbeta Kostelanská
*Department of Telecommunications*
*Brno University of Technology*
Brno, Czech Republic
xkoste13@vut.cz

Norbert Herencsár
*Department of Telecommunications*
*Brno University of Technology*
Brno, Czech Republic
herencsn@vut.cz



***Abstract*—The monitoring and precise measurement of plants' physical condition provides a valuable tool for the optimization of processes in agriculture. One of the tools that provides the possibility to accurately determine the state of the different parts of the plant body is impedance spectroscopy. This research deals with the design of a prototype unit that is capable of automated bioimpedance measurement of plant tissues. Utilizing the designed prototype, test measurements were carried out. First measurement was used to evaluate the prototype unit. Subsequently, bioimpedance measurements for a stem of the chosen plant were performed.**



***Index Terms*—bioimpedance, electrical impedance spectroscopy, EIS, AD5933, LoRa, smart agriculture**

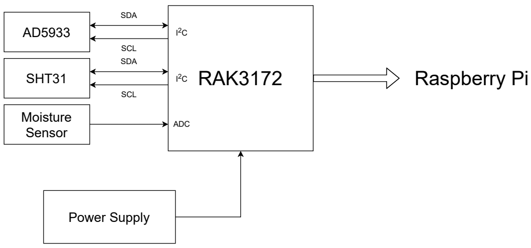

## I. Introduction

Precise and efficient methods for measuring environmental conditions and the cultivated plants can provide a reliable and sustainable means for agricultural resource management. Based on the environmental parameters, in which the plant is located, irrigation schedules, nutrient and fertilizer supply, or sufficient light supply can be determined. The most common methods of measuring factors influencing the plant growth include measurement of the relative humidity of the environment and soil, temperature, the amount of nutrients in the soil and the amount of light provided to the plant. These factors can be monitored using various types of application-specific sensors. The primary use of bioimpedance spectroscopy for determining the condition of plants is measuring of the water content in the plant tissues with a non-destructive method. The possibility of evaluation of the plant health and status using bioimpedance spectroscopy is described in [1]. The study carried out in [2] involves the observations of changes in the plant tissues due to stress caused by different watering intervals. Electrical impedance spectroscopy (EIS) can be utilized to monitor the growth and health condition of the plant's root system. This is being described partially in [3] using the example of plants grown hydroponically. The article [4] presents the use of bioimpedance measurements as a tool for better visualization of the root system as well as the tree trunk in the assessment of tree growth condition. Because the roots are responsible for main nutrient intake of the plant, the size and health of the root system can be used to determine growth of the plant, its health and the environmental suitability. EIS can be useful for assessing the impact of frost on plants and plant products [3]. With the effect of frost in mind, impedance spectroscopy can also be used to observe and record the adaptation of plants to the change of the season and the environmental temperature. Based on the bioimpedance values measured on various types of fruits and vegetable, it is possible to determine not only their ripeness [5], but also their overall quality and nutritional value [3]. From the analysis of the fruit tissues, an estimation of the ripening time can be determined and thus better planning of the fruit harvest can be established. This can lead to more efficient food production. EIS is useful not only for assessing the condition of food products, but also for observing their aging process. The concept presented in [6] describes the possibility of using bioimpedance as a mean for early diagnosis of plant tissue diseases. Bioimpedance spectroscopy is used for assessment of the plant's health status and the possibility of the spreading of diseases that can damage the plant tissues and cells. The study presented in [7] focuses on the effects of nutrient deficiency on the plant growth. From the presented observations, it draws conclusion, based on the bioimpedance values it is possible to develop more accurate system for distribution of the required nutrients to the plants in order to ensure their optimal growth.

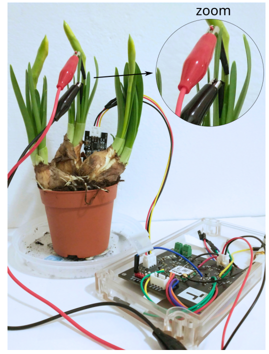
In this paper, the design of a prototype of a bioimpedance analyzer usable in agriculture has been introduced. The proposed unit is capable of measurements in the range of 1 kHz to 100 kHz, and the first validation measurements were performed on the stem of a selected plant.

## II. Developed Measuring Unit

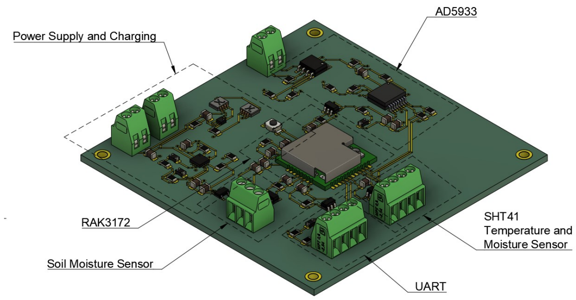
For the purposes of automated bioimpedance measurement, a custom sensor unit was designed. Considering that the unit is to be used primarily for autonomous measurement in an agricultural environment, and also primary for outdoor use, it is necessary to ensure the following features: long range for the data transmission, continuous automatic measurement, long battery life, automatic processing of the measured data and possible notification of the user or of the connected control system. The block diagram of the proposed prototype unit is in shown in Fig. 1(a). Proposed unit is battery powered, one rechargeable Li-Ion battery cell is used. For the case of outdoor use, solar battery charging system is also implemented. As described in the previous chapter environmental conditions can




(a)




(b)




(c)

Fig. 1: (a) Block diagram of the measuring unit, (b) measurement setup with wild daffodil, and (c) the 3D model of the measuring unit prototype.

greatly influence the measured tissue impedance, therefore soil moisture monitoring sensor, air moisture sensor and temperature sensor are also present in the proposed device prototype.

Proposed unit for measuring and evaluating of plant tissues is based on impedance converter AD5933. This interface is capable of measuring impedance in range of 100 Ω to 10 MΩ in frequency spectrum up to 100 kHz, which makes it suitable for measurement of the plant tissue characteristics. The unit is using the LoRa module RAK3172 based on microprocessor STM32WLE5 for processing and transmitting data. LoRa modulation allows to send measured data over large distances, which makes the proposed unit suitable for autonomous measurements in the outdoor environment. The device implements LoRa Point-to-Point communication, the measured complex data is sent directly to the data storing and collection node, which consists of Raspberry Pi with attached LoRa receiver module. It serves as a server for storing and evaluating the measured data. The price for one unit is around 20 €, which makes it a cost-effective alternative to the precise desktop impedance analyzer. Proposed prototype used for the measurements done for this paper and its 3D model are on the Figs. 1(b) and (c), respectively.

## III. Experimental Results

### A. Materials and Methods

The measurements were performed in the indoor environment. For the measurement, wild daffodil (*Narcissus Pseudonarcissus* L.) was chosen. Selected plant species has a relatively short life cycle in indoor conditions (average temperature of 21 °C and an air humidity of 50–60 %). The time period from sprouting to flowering was approximately two weeks. Plant was measured during the flowering stage of it's life cycle. In the case of leaf measurements, Ag/AgCl electrodes were used where the measurement was taken outside the plant tissue. In the case of stems, needles inserted directly into the plant tissue were used.

Firstly, the unit was tested on passive components. 20 kΩ resistor was used for this part of the measurement. Both lay within the possible measurement range of the AD5933 interface. The second measurement focused on the differences in impedance values for the various parts of the observed plant. In this case, the bulbs, stems, older bottom leaves and younger upper leaves of the aforementioned plants were measured. The measured data was compared with the USB oscilloscope and impedance analyzer *Analog Discovery 2*.

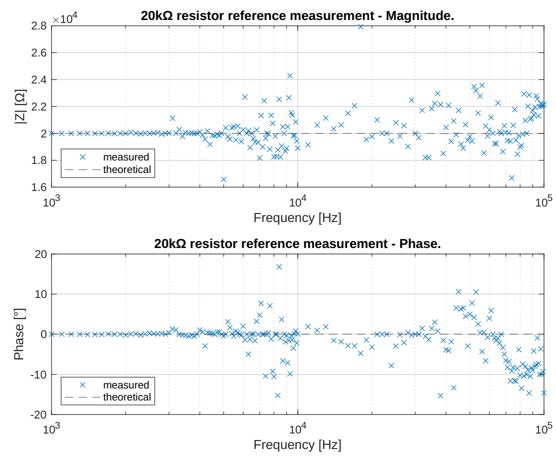



Fig. 2: Measured impedance and phase for 20 kΩ resistor.

### B. Passive Component Verification

An initial test measurement of the prototype bioimpedance unit was carried out by measuring passive component values. In the Fig. 2, the results for the 20 kΩ resistor can be seen. It is evident from the results of the resistor measurements that the proposed prototype shows good match, especially for the lower frequencies of the measured range. For higher frequencies, the measurement also included the interference, whose source may vary. A 20 kΩ resistor was used in the measurement. The measured value was then adjusted by a calibration constant, which represents the impedance and phase shift of the prototype unit itself. The measured value of impedance lies primarily in the value range of 18 kΩ to 22 kΩ. As the frequency increases, the deviation of the values from the 20 kΩ reference becomes more pronounced. The measured phase lies primarily in the

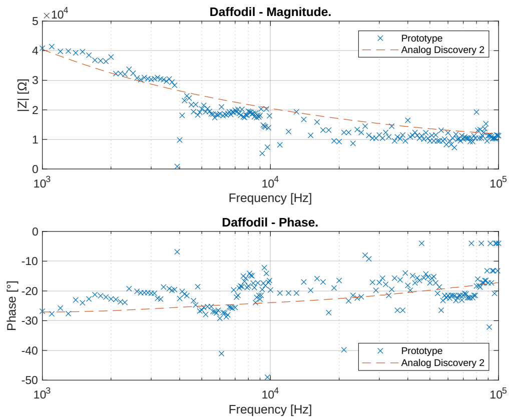



Fig. 3: Measured bioimpedance on the stem of the chosen plant.

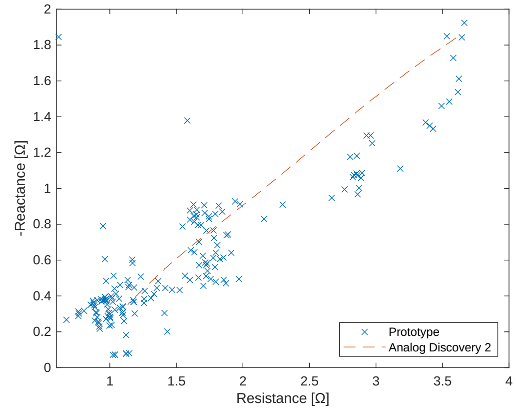



Fig. 4: Comparison of bioimpedance measured with proposed unit and Analog Discovery 2 device - Nyquist diagram.

range 10° to −10°. Here as well, a larger dispersion of values is visible for higher frequencies.

### C. Plant Stem Bioimpedance Measurement

The second part of the measurement involved testing the prototype unit in measuring the bioimpedance of the plant stem. Deviations from the reference measurement obtained with the *Analog Discovery 2* device are particularly evident in the phase values. In this case, the measured values show a large scattering compared to the reference measurement. The results of the daffodil stem measurement are presented in the Fig. 3. The Nyquist diagram of the measured impedance of the daffodil stem is shown in the Fig. 4.

## IV. Discussion and Concluding Remarks

In this paper, the design of a prototype of a bioimpedance analyzer usable in agriculture has been introduced. First validation measurements were subsequently performed for the proposed unit. The proposed unit is capable of measurements in the range of 1 kHz to 100 kHz. As shown in the tests, the prototype unit shows better results for the lower frequencies. The prototype unit was calibrated using a precise 10 kΩ resistor. For the calibration of the AD5933 interface, the best results can be obtained when the value of calibration resistor is $R_{\mathbf{CAL}} = (Z_{\mathbf{MIN}} + Z_{\mathbf{MAX}}) \times \frac{1}{3}$, where $Z_{\mathbf{MIN}}$ and $Z_{\mathbf{MAX}}$ are expected minimal and maximal measured impedance values. Furthermore, it is advised to calibrate the interface for as narrow as possible impedance range, taking into account the measured unknown impedance range. In the case of a 20 kΩ resistor measurement, this value is within the recommended range, but at the same time the measured results indicate the need for a finer calibration of the device to achieve better measurement accuracy.

A future work and follow-up experiments, it would be appropriate to develop a better solution for the device calibration for different unknown impedance ranges. This can be done either as a software or a hardware solution by switching between the calibration resistors or switching between multiple calibration lookup tables. In addition to this, there is also a need to provide better protection of the prototype against random noise generated by the environment and also the prototype system itself, which is evident in the measured results. In order to improve the measurement performance of the proposed prototype, it is necessary in the subsequent experiments and measurements, to ensure a better isolation of the measuring electrodes and the wires that connect them to the prototype, as they can be a source of interference and inaccuracies in the obtained data. Within this measurement, an adjustment of the measured results for this interference has been conducted, but since in the case of the wires this interference is very varied. These variations can be caused by the environment and electrode and wire placement. The software adjustment does not seem as sufficient. A second significant source of errors in the results could be the PCB layout of the prototype. In addition to used correction lookup table, further tests for this problem are needed.

## References

[1] U. Garlando, S. Calvo, M. Barezzi, A. Sanginario, P. Motto Ros, and D. Demarchi, "Ask the plants directly: Understanding plant needs using electrical impedance measurements," *Computers and Electronics in Agriculture*, vol. 193, p. 106707, 2022. doi: 10.1016/j.compag.2022.106707

[2] D. Jamaludin, S. Abd Aziz, D. Ahmad, and H. Z. Jaafar, "Impedance analysis of Labisia pumila plant water status," *Information Processing in Agriculture*, vol. 2, no. 3, pp. 161–168, 2015. doi: 10.1016/j.inpa.2015.07.004

[3] I. Jócsák, G. Végvári, and E. Vozáry, "Electrical impedance measurement on plants: a review with some insights to other fields," *Theoretical and Experimental Plant Physiology*, vol. 31, no. 3, pp. 359–375, Aug. 2019. doi: 10.1007/s40626-019-00152-y

[4] K. Kobata and S. Honda, "Evaluating plant growth by utilizing Electrical Impedance Analysis," in *2014 Proceedings of the SICE Annual Conference (SICE)*, 2014, pp. 113–118. doi: 10.1109/SICE.2014.6935184

[5] P. Ibba, A. Falco, B. D. Abera, G. Cantarella, L. Petti, and P. Lugli, "Bio-impedance and circuit parameters: An analysis for tracking fruit ripening," *Postharvest Biology and Technology*, vol. 159, p. 110978, 2020. doi: 10.1016/j.postharvbio.2019.110978

[6] E. Borges, A. P. Matos, J. M. Cardoso, C. Correia, T. Vasconcelos, and N. Gomes, "Early detection and monitoring of plant diseases by Bioelectric Impedance Spectroscopy," in *2012 IEEE 2nd Portuguese Meeting in Bioengineering (ENBENG)*, 2012, pp. 1–4. doi: 10.1109/ENBENG.2012.6331377

[7] B. Kashyap and R. Kumar, "Sensing Methodologies in Agriculture for Soil Moisture and Nutrient Monitoring," *IEEE Access*, vol. 9, pp. 14 095–14 121, 2021. doi: 10.1109/ACCESS.2021.3052478

# Test benches for microcontroller kits


1st Ivo Žaludek
*Department of Control and Instrumentation, FEEC*
*Brno University of Technology*
Brno, Czech Republic
xzalud08@vutbr.cz

2nd Jakub Arm
*Department of Control and Instrumentation, FEEC*
*Brno University of Technology*
Brno, Czech Republic
arm@vut.cz

3nd Zdeněk Bradáč
*Department of Control and Instrumentation, FEEC*
*Brno University of Technology*
Brno, Czech Republic
bradac@vut.cz



***Abstract*— This paper deals with the design of a test bench for modules used in educational labs. The first section outlines the requirements for the tester, followed by a description of the printed circuit board design process. Implementation details of the software are then examined, including a graphical user interface based on Model-View-Controller (MVC) architecture that allows for easy result printing. The obtained results are presented, and further work is discussed.**



***Keywords—automated testing, hardware development, functional testing, megaAVR, MVC architecture***


## I. Introduction

With the increasing complexity of electronic systems, printed circuit boards (PCB) testing is becoming an essential element to ensure the reliable operation and safety of electronics. Testing not only identifies potential manufacturing defects, but also contributes to the detection of errors in circuit design or device program. There are a variety of PCB tests, ranging from optical inspection and electrical tests to complex functional tests, which typically verify the correct operation of both the hardware and software of the device. [1]

This paper deals with the design of a test bench for various PCB modules. These modules are used in the laboratories for educational purposes, and it is required to perform a functional test of tens of devices a few times a year to ensure that they have not been damaged. There is a kit which contains all modules as listed below. Research has been conducted on a suitable tester. However, due to the high cost of such a device and the wide variety of PCB types, it was decided to design our own device from scratch.

## II. Requirements

The first step was to determine the requirements for the test bench. To meet the user’s needs, the test bench must allow for easy selection of tests and result printing. A PC program was therefore developed. This simplifies test setup, execution and report logging.

Determining test coverage for a functional test can be challenging. There are methodologies available to estimate test coverage for system test, embedded test or stress test. However, due to a lack of documentation for some tested modules, this approach was considered difficult to achieve and such detailed functional test coverage report is out of scope of this project. [2]

Nonetheless, some inspiration from these methodologies was taken into consideration during the test specification creation phase. Three methodologies are presented in [2] and the Declaration and Deduction methods were used. The Declaration model uses schematic and layout viewers for the test coverage input, while the Deduction model is based on the premise that certain components within a functional block are covered by specific test sequences. [2]

An analysis was conducted for each module, and based on this analysis, the test specification was created. The testing procedure is briefly described below:

### A. ATmega328P Xplained Mini

The digital inputs and outputs (DIO) and analogue-to-digital converter (ADC) were tested using the ATmega328P Xplained Mini evaluation kit. Internal ADC references are tested as well. The kit communicates via UART (Universal asynchronous receiver-transmitter); therefore, the functionality of this peripheral is also verified.

### B. Relay module

The relay module consists of two independent electromechanical relays, which have a limited operating lifetime due to mechanical and contact wear [3]. Therefore, a circuit test is conducted separately for each relay output in both the on and off states.

### C. Module with real time clock (RTC) and EEPROM memory

The RTC module undergoes testing by writing an example time and date into the module's registers and then reading the value back. Additionally, the module's square wave output, which has a configurable frequency, is also tested. The module also includes EEPROM (Electrically Erasable Programmable Read-Only Memory) memory, which is tested by writing a test pattern, reading the memory content, and comparing it with the pattern.

### D. Incremental encoder

The incremental encoder has two outputs, A and B, which enable determination of the direction of rotation and the encoder goes through four states in one click (movement). During testing, it is verified that all four states were reached in both clockwise and counterclockwise directions. The button on the encoder axis is also tested.

### E. A matrix keyboard

The matrix keyboard test is simple: the user presses all the buttons, and the test device verifies that all the buttons have been pressed.

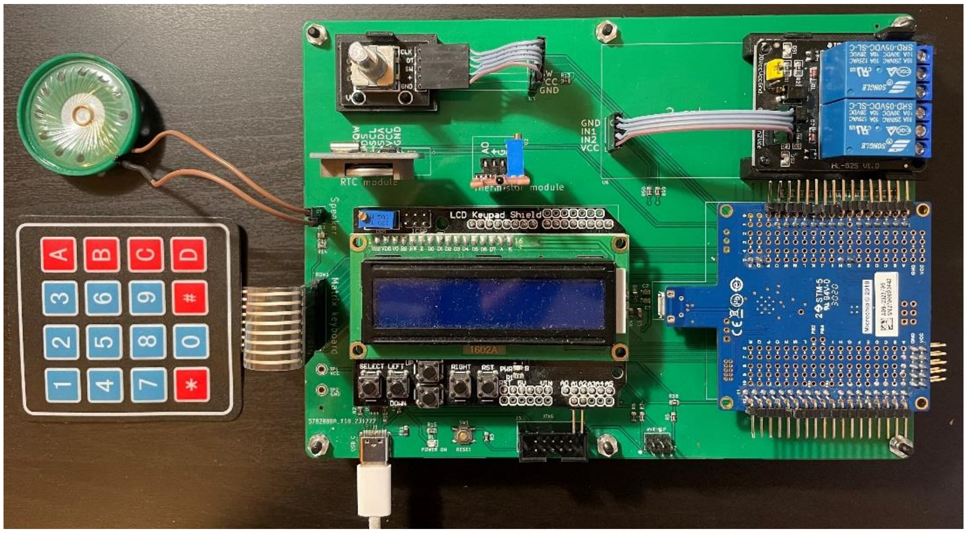


Fig. 1. Test bench with connected modules.

### F. 1602A LCD module with keypad

The monochrome liquid-crystal display (LCD) module is tested in three stages: first, all pixels are activated and then a test pattern is displayed. The first stage tests the functionality of the display itself, while the second stage verifies that the character generator works as desired. In the third stage, the user is prompted to press all buttons on the LCD module.

### G. A speaker

The speaker test consists of exciting the loudspeaker with a 2kHz signal and detecting the sound coming from the speaker.

### H. Module with thermistor

In order to test the module with a thermistor, calibration is necessary. As this process can be lengthy, the test bench only reads the analog value and prints a graph on the graphical user interface (GUI). The user can then determine if the thermistor is functioning properly.

## III. Circuit Solution

Most of the devices under test (DUTs) are evaluation kits, so they provide access to all important pins within their connectors. This eliminates the need for expensive test probes while still allowing for easy connection of all modules. The relay module is the only exception, as it features screw terminals on its output. Spring-loaded (pogo) pins are used in this case to avoid lengthy wire installation into the screw terminals.

### A. Microcontroller

The choice of a suitable microcontroller (MCU) for running and evaluating all tests was influenced by several requirements. All modules typically operate with 5 V logic, so the MCU must also be compatible with this voltage. The MCU must have a UART, I2C, and SPI interface, as well as at least 8-channel ADC. In addition, the MCU has to have a sufficient number of GPIOs. Other factors that were considered include availability, good documentation, and manufacturer support. The ATmega2561 from Atmel's AVR family was selected for the task, given the desired characteristics. Additionally, this choice ensures the preservation of a unified programming environment with tested ATmega328P.

### B. Communication and debugging interface

The ATmega2561 can be programmed using either the JTAG (Joint Test Action Group) or ISP (In-system programming on Serial Peripheral Interface) interface. Both the JTAG and ISP interfaces were used due to a collision between the ISP interface and the UART interface pins. [4] Microchip Studio environment was used for writing and debugging the program for both ATmega2561 and tested ATmega328P kit. Atmel ICE was used as a debugger. USB interface was chosen for communication between the PC and the test bench. CH340 series USB-to-serial converter was connected to ATmega2561 and transient-voltage-suppression diodes were added on the USB input pins as a protection against electrostatic discharge (ESD) and overvoltage.

### C. Power supply and test bench protection

A USB C connector was used in the Upstream Facing Port configuration. The USB 2.0 standard specifies that the host can supply up to 500 mA [5]. Table I. shows that the available power from USB is sufficient to operate the entire test bench even if all components were drawing maximum current simultaneously. Resettable fuse was used as a protection against overcurrent or short circuit.

TABLE I. TEST BENCH TOTAL CURRENT CONSUMPTION TABLE

| ***Component*** | ***Max. current consumption [mA]*** |
|---|---|
| ATmega2561 | 14 |
| ATmega328P Xplained Mini | 14 |
| 1602A LCD module | 30 |
| Relay module | 135 |
| Module with RTC and EEPROM | 8,5 |
| Module with thermistor | 7,2 |
| Speaker | 7 |
| USB-to-serial converter | 20 |
| Total current | 221,7 |

All output pins of ATmega2561 were protected against overcurrent using resistors, ESD protection was not needed in this case because microcontrollers pins have their own protection diodes on chip. The correct functionality of ATmega328P ADC is tested on 3 voltage references: 0V, VCC and bandgap voltage. For this reason, voltage dividers were added on ATmega328P pins with ADC. The speaker is connected to a high pass RC filter with a cut off frequency set to 800 Hz.

## IV. The Software Architecture

In total, three separate software programs were implemented: the GUI for PC, the program for ATmega2561 and the program for tested ATmega328P. The ATmega2561 has two UARTs, one is used for communication with the GUI application and the other communicates with ATmega328P. The communication is based on a set of instructions formed from text strings. Instructions sent from PC application are used to initiate or terminate a test of certain module and the test bench responses with pass or fail answer. In case of fail, the test bench also prints which component of the module failed (e.g. a pin on ATmega328P module). Communication using text strings is human readable, so it is easy to communicate with the test bench even without GUI via a serial terminal like PuTTY.

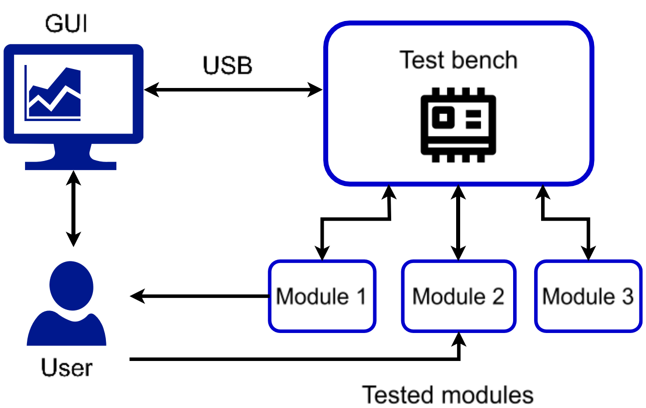



Fig 2.. High-level diagram of key components.

Some of the modules are user interface modules. To test these modules, the user must perform some action such as pressing down a button or checking the LCD module. Although automation of these tests is possible, it was deemed too costly to add a robotic arm to interact with the modules. Therefore, it was decided that the user would interact with the modules during these tests. As a result of this discussion, the tests were classified into three different categories.

a) Tests which are completely automated

b) Tests where the user interacts with the tested module (e.g. rotates the encoder) and the test bench evaluates the response

c) Tests where the test bench initializes the module and user evaluates the correct functionality (e.g. correct pattern on the LCD module)

### A. *Test bench software*

Microchip Studio environment was used for writing and debugging the software for both ATmega2561 and tested ATmega328P kit. The Atmel START tool was used to configure the software framework. However, own drivers were implemented for individual peripherals due to specific requirements.

The test bench periodically reads the input UART buffer and waits until the a line of text is sent. After a command is sent, the ATmega2561 processes the command and executes the task if it corresponds to a certain instruction. These tasks typically involve individual module tests.

For tests in category a), the test bench only runs the test and prints the result. Tests in category b) run in a loop. The user is prompted via GUI to perform an action with the tested module. If the test bench detects all states (e.g. rotation of encoder to both sides), the test passes. If some input is missing, the user is notified and can perform the action again. However, if there is a problem with the module and reaching the missing state is impossible, the user can terminate the test, and the failure of the part is detected. The user determines the appropriate behavior of modules in tests category c). The test bench is unable to verify the proper functioning of the module from category c) due to the lack of a sensor (e.g. a camera in the case of an LCD module) or the need for calibration before running the test (in the case of a thermistor test where only the dynamic behavior is tested). The test bench only verifies proper communication with the module and initiates the test.

The test cases implementation follows the test specification outlined in the requirements with a focus on achieving maximal coverage. During the development of test cases, fail states were simulated to ensure that the tests can identify them.

### B. *ATmega328P software*

The software architecture of the ATmega328P firmware is similar to that of the ATmega2561. However, the command list differs. The ATmega328P has commands that allow for printing of values converted by ADC, including values from internal voltage references, reading of digital values on its pins, and setting of logical values on digital pins when requested via the serial interface.

### C. *Graphical user interface*

The GUI is implemented in the Python programming language. There are many GUIs written in Python, such as Qt, wxPython or Kivy, but Tkinter was chosen because it is part of the standard Python library and is widely used.

The GUI has several functions: it allows the user to select which modules to test, which COM port the test bench is connected to, whether the module test passed or failed, and in case of a failure, it prints out the reason why the test failed.

The application is responsible for serial communication, executing different types of tests and displaying the results. An appropriate design pattern must be chosen to enable reusability, easy editing, and modularity of the application. The Model-View-Controller (MVC) architecture was chosen because it aims to separate the data, user interface and control logic of the application. The Model manages the storage, retrieval, and manipulation of data in the architecture. In our case, the model is mainly responsible for serial communication. The View architecture component displays data to the user and receives user input. The View should be independent of the data and control logic. This is the component where the user interface is implemented. The Controller acts as an intermediary between the Model and the View. The execution logic is implemented in this component. It is responsible for handling user input from the View, propagating data to the Model and updating changes back to the View. [6]

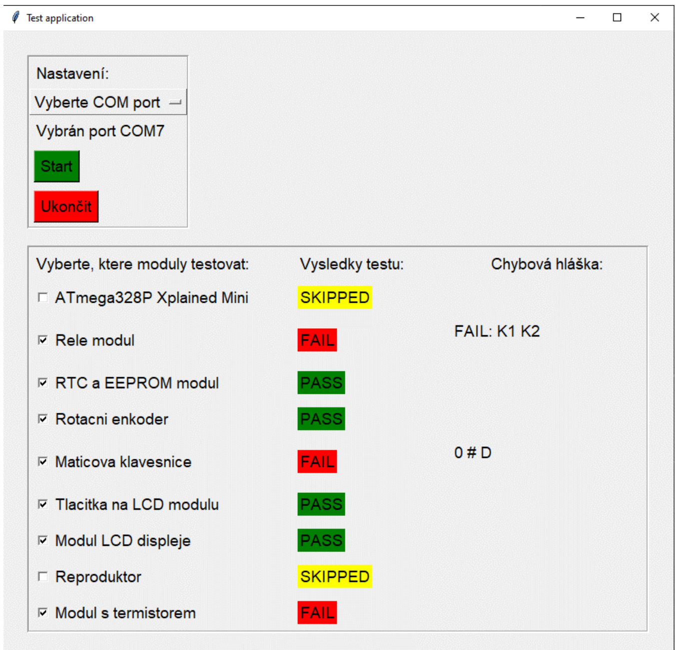



Fig. 3. Graphical user interface (GUI) after running a set of tests.

The use of the MVC model proved to be appropriate for this application. The main challenge in implementing this model was to split the application functionality appropriately into different components of the model. The GUI of the application is in Czech language.

## V. Obtained Results

All test cases have been implemented according to the test requirements and correct functionality was tested on real hardware. Some modules are tested with user interaction, other tests are fully automated. All the modules in a kit can be tested in tens of seconds, depending on how quickly the user interacts with each module. The speed of the test execution is considered to be pleasant for the fact that the tests are only carried out a few times a year with tens of kits. The GUI was found to be useful for repeated testing of a set of modules, as it remembers the previous settings, so the user simply plugs in the modules and presses the start button. Test results are printed, and failures are listed.

## VI. Conclusion

A test bench for various types of PCB modules used in laboratories was designed. The first part of the work was to determine the requirements and design a PCB that would meet them. The next part was to implement software for both the test bench and the tested ATmega328P Xplained Mini kit. A graphical user interface (GUI) was created to run on a PC connected to the test bench via USB. The GUI provides easy test execution and a clear view of results.

In the following work the reliability of the solution will be examined and more failures will be simulated to verify the ability of the test bench to detect all possible failures. Another task is to develop software that enables the test bench to flash firmware directly into the tested ATmega328P, eliminating the need for it to be connected to a PC first.

## Acknowledgment

The completion of this paper was made possible by the grant No. FEKT-S-23-8451 - "Research on advanced methods and technologies in cybernetics, robotics, artificial intelligence, automation and measurement" financially supported by the Internal science fund of Brno University of Technology.

## References

[1] M. Moganti, F. Ercal, C. H. Dagli, and S. Tsunekawa, "Automatic PCB Inspection Algorithms: A Survey," *Computer Vision and Image Understanding*, vol. 63, no. 2, pp. 287–313, Mar. 1996, doi: https://doi.org/10.1006/cviu.1996.0020.

[2] C. Lotz, P. Collins, and D. Wiatrowski, "FUNCTIONAL BOARD TEST -COVERAGE ANALYSIS what does it mean when a functional test passes?", in *2006 European Board Test Workshop*, 2006. Accessed: Nov. 11, 2023. [Online]. Available: https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=538a0db1a206e2a45dbfef0359b4105235c8f358.

[3] P. Roettjer, "Testing Techniques to Improve Relay Reliability," *Relay Testing Services, LLC*, 2023. Accessed Jan. 12, 2024. Available: http://www.relaytestingservices.co/ee_article_pdf.pdf.

[4] Microchip Technology, "AVR® Microcontroller Hardware Design Considerations," *Microchip Technology*, 2013. Accessed Nov. 14, 2023. Available: https://ww1.microchip.com/downloads/aemDocuments/documents/MCU08/ApplicationNotes/ApplicationNotes/AN2519-AVR-Microcontroller-Hardware-Design-Considerations-00002519B.pdf.

[5] Microchip Technology, "Basic USB Type-C™ Upstream Facing Port Implementation," *Microchip*, 2015. Accessed Nov. 09, 2023. Available: https://ww1.microchip.com/downloads/en/AppNotes/00001914B.pdf.

[6] J. Deacon, "Model-View-Controller (MVC) Architecture," *Computer Systems Development*, 2009. Accessed Feb. 02, 2024. Available: https://www.academia.edu/download/50526307/MVC.pdf.

# Accident Tolerant Fuel simulation loaded in advanced nuclear power reactor during severe accident conditions

1st Lukas Hamrik
*Department of Electrical Power Engineering*
*Brno University of Technology*
Brno, Czech Republic
Lukas.Hamrik@vut.cz

2nd Peter Mician
*Department of Electrical Power Engineering*
*Brno University of Technology*
Brno, Czech Republic
Peter.Mician@vut.cz

***Abstract*—This paper introduces accident tolerant fuels with respect to their current state of development, while the leading causes are briefly mentioned. The main reasons for the development of other cladding materials are presented. Selected types of accident tolerant fuel cladding materials are described in short and then are used for a simulation in a model of advanced power reactor undergoing a severe accident scenario in the MELCOR code.**

***Index Terms*—nuclear fuel, accident tolerant fuel, safety analysis, severe accident, MELCOR**

## I. Introduction

Nowadays, advanced reactor designs have a much higher level of safety with improved safety margins. However, the possibility of an accident is a concern, although the risk is low. There are still ways to improve safety margins; an exceptional way is by introducing accident tolerant fuels (ATF). These materials exhibit enhanced protection against severe accident conditions, reducing the potential for core damage and the release of radioactive materials. ATF considers the use of advanced materials that show superior properties under accident conditions, such as high thermal stability and enhanced oxidation kinetics. There are three main categories under the ATF concept: high thermal conductivity fuels, advanced cladding materials, and zirconium alloys coated with a protective layer. The main events that helped promote and accelerate development are those that occurred at the Three Mile Island and Fukushima-Daichi plants [1].

## II. Accident tolerant fuels

Fuel is protected by cladding tubes from contact with water and steam. Oxidation of the cladding by air and steam occurs under conditions that lead to severe accidents, during which hydrogen is produced. The accumulation of hydrogen produced in the reactor coolant system (RCS) and containment could lead to detonation when ignited at certain concentrations with a containment atmosphere. Combustion could damage the RCS and containment and radioactive release could occur.

Nowadays, in commercial pressurised water reactors (PWRs), zirconium-based cladding is used. Although these alloys have advantages such as high thermal stability and low thermal neutron cross section, the main issue of Zr is still present. Zr alloys are subjected to steam oxidation, which produces a substantial amount of hydrogen and heat, and the process is greatly accelerated by increasing the temperature [3] [4].

The ATFs are based on the principle of Do No Harm, which means that under all operating conditions the ATF fuel system must perform at least as well as or better than the current fuel system. In addition, the new fuel concepts must be compatible with previously used fuel systems in all areas such as storage, proliferation prevention, etc.

## III. Cladding materials

The two most viable options that could be used as cladding materials are advanced steels, FeCrAl alloys, and refractory ceramic, silicon carbide. An extensive amount of data in non-irradiated state for both materials is available from laboratory testing or other industries of use and are compatible with light-water reactor (LWR) chemistry.

### *A. FeCrAl alloys*

These alloys based on iron, chromium and aluminium are representatives of advanced steels that have greatly improved oxidation resistance. Nuclear-grade alloys have optimised chemistry and microstructure. They have a lower content of chromium (10-13 wt%) and aluminium (4-6 wt %), with minor additions of niobium and molybdenum to improve overall strength. Strength allows for a reduction in the thickness of the cladding tube, compensating for the higher neutronic penalties. In a steam environment, the protective layer of $Al_2O_3$ is formed on the surface of the material, reducing the further oxidation of the base material by steam, also the $Cr_2O_3$ layer is formed, improving the overall stability of the protective layers.

### *B. Silicon Carbide*

Has an established use in environments of highly corrosive nature and where parts must withstand high temperatures. Compared with Zr alloys, SiC is chemically inert, has a lower neutron absorption cross section, and is able to cope with higher temperatures. As it is a ceramic material, it is brittle; therefore, the proposed designs are in the form of duplex and triplex structures to improve mechanical strength. The structure may be based only on SiC or may be assisted by metal. The SiC matrix is subject to dissolution phenomena under LWR conditions, the environmental protection barrier is required; if the barrier is not present, the environment in the long term would affect the structure as it would be degraded [5] [6].

## IV. Methodology

The most severe scenario has been chosen to be unmitigated Large Break Loss Of Coolant Accident (LB LOCA), it is a DBA which the plant is designed to cope with; however, if is not mitigated by engineered safety features, it progresses to severe accident.

### *A. MELCOR code*

The MELCOR code allows for advanced modelling of the progression of such a scenario in LWR technology plants. Materials are defined using available data on FeCrAl - C35M from ORNL [2] and on SiC from [7]. Because the materials produce oxides as they are subjected to the steam environment, the oxides and the conditions under which they form have to be addressed in the code by the use of generic oxidation modelling (GOM). The thermophysical properties of the forming oxides were defined using available data from the NIST JANAF database.

The main area of interest in the simulation is the produced hydrogen and heat produced by the oxidation of the cladding. It is a must to note that only the material of the cladding was changed and its properties; no changes to the fuel rod geometry and neutron kinetics models were introduced.

### *B. Advanced power reactor model*

For investigation of ATF complex impact, an APR1400 model is used. The RCS configuration of the used model has two reactor coolant loops. Each loop consists of a hot leg, a steam generator, two cold legs, and reactor coolant pumps at each of the cold legs. The nuclear steam supply system is designed to operate at a rated thermal output of 4000 MW. The RCS model includes all the most important systems – detailed primary circuit (reactor pressure vessel with internals, two loops, pressuriser), the most important components of the secondary circuit (steam generators, steam lines, main steam safety and isolating valves) and safety systems (Safety Injection Tanks, Safety Injection Pumps, Safety Depressurisation System).

In Fig. 1 a schematic view of the control volumes of RCS and the secondary system and its layout is shown. The RCS consists of 32 control volumes; the secondary system consists of 11 control volumes. The boundary volumes are the turbine connected to the main steam hub and the main feedwater system.

The reactor core is modelled in detail using one control volume and the COR package. The COR package covers modelling schemes not only for the core, but also for the lower plenum. In total, 15 axial levels and 4 radial rings are used. The nodalization of the COR package is illustrated in Fig. 2, where the green colour represents the active fuel and the blue represents the lower plenum. From that, active fuel is present in 10 axial levels and 3 radial rings.

### *C. Accident scenario*

The initiating event for the scenario analysed is a cold leg guillotine break. Break size is of 200 % cold leg cross-section area and is located between the cold leg volume CV480 of the loop without pressuriser and the RPV downcomer

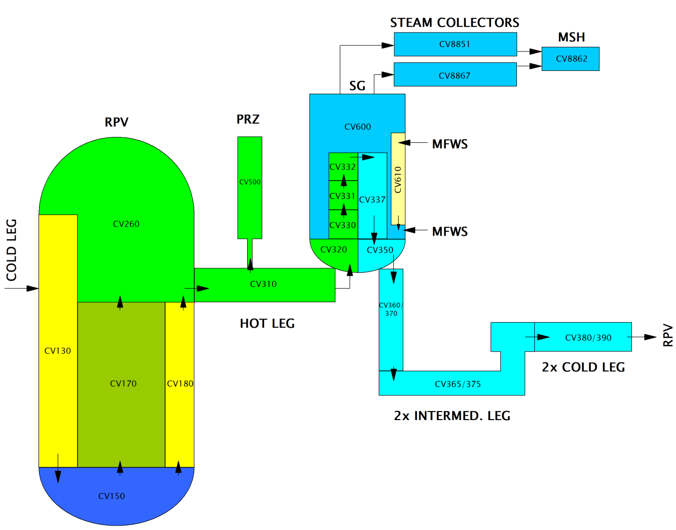



Fig. 1. Control volumes and flow path nodalization of the model (for simplification only one loop with pressuriser and single cold leg are presented) [8].

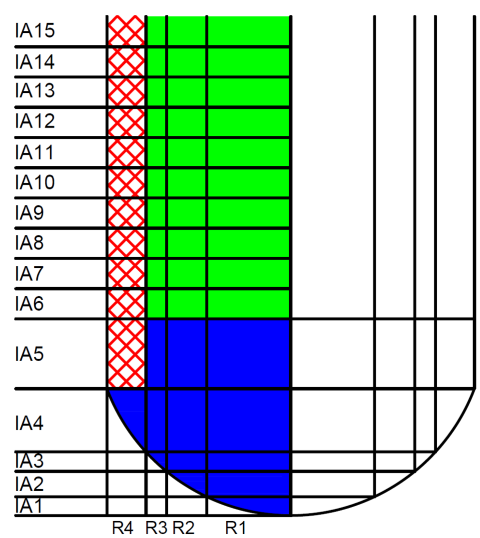



Fig. 2. Core nodalization of core package [9].

volume CV130. In terms of engineered safety features and accident mitigation, all safety injection pumps are assumed to be unavailable, and only passive Safety Injection Tanks are available.

### D. Oxidation modelling

The generic oxidation model in MELCOR uses the Arrhenius correlation for the calculation of oxidation. It is assumed that the material would undergo oxidation by parabolic rate characteristics. The modelling of the SiC oxidation reaction is as follows, with consideration of oxidation of both the Si and C components under high-temperature steam.

$$SiC + 3H_2O \rightarrow SiO_2 + 3H_2 + CO \tag{1}$$

For the C35M, two approaches are determined, while the oxidation parabolic kinetics rate remains unchanged. The first approach considers oxidation of Fe to $Fe_3O_4$ oxidation, while being accompanied by Cr to $Cr_2O_3$ and Al to $Al_2O_3$ oxidations, the oxidation rate of each component is weighted by the percentage of weight of the components in the alloy. The reactions are as follows:

$$Fe + H_2O \rightarrow FeO + H_2 \tag{2}$$

$$2Cr + 3H_2O \rightarrow Cr_2O_3 + 3H_2 \tag{3}$$

$$2Al + 3H_2O \rightarrow Al_2O_3 + 3H_2 \tag{4}$$

The second approach considers only a formation of $Al_2O_3$ protective layer based on the research [10], which shows that the formed layer of oxides is primarily of $Al_2O_3$. MELCOR is capable of calculating only three simultaneous oxidation processes, only the elements with the highest weight fraction are considered for oxidation for the C35M alloy. Formed oxides are not treated separately in the code, and one generic oxide is formed. Thus, the thermal properties of each oxide were weighted by their respective molar weights to the total molar weight of the generic oxide formed and by the percentage of element weight in the alloy to determine an estimate on the thermal properties of the generic oxide.

## V. Results and Discussion

As the simulated materials are not in the MELCOR library, for use in scenario, further optimisation and debugging was required, with some estimates imposed. The results of the simulation are shown in the graphs below.

In Fig. 3, the graph of heat produced by cladding oxidation is shown. The lower the heat production by the oxidation, the better, because the heat produced in this way could significantly accelerate the fuel melting in combination with the decay heat from the fuel. As mentioned above, zirconium alloy oxidation produces an excessive amount of heat. In the graph legend, FeCrAl is marked with the FCA shortcut.

Both proposed materials, the SiC and FeCrAl alloy, have shown significantly lower heat production because the reaction heat of the base material is lower. Lower heat production should extend the coping time before the lower head of the RPV fails due to the stresses posed by the melted core; also,

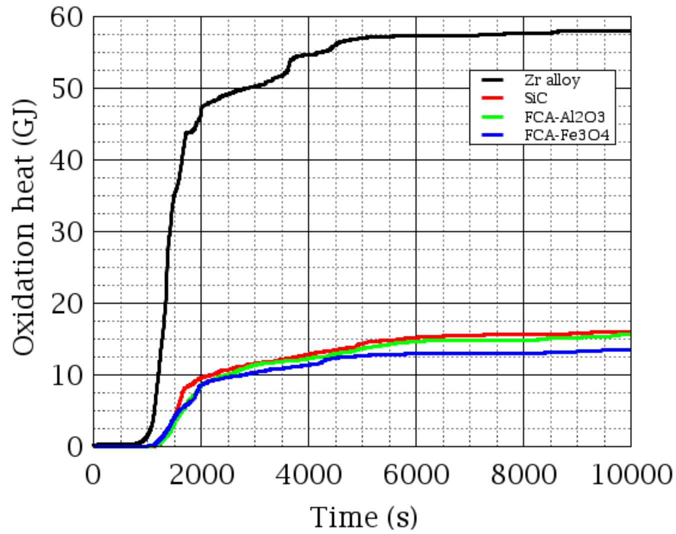



Fig. 3. Total oxidation heat produced due to oxidation of in-vessel components.

the cladding mass plays its role. In case of FeCrAl, differences could be observed between the approaches taken; the weighted approach shows a lower heat production, as the generic oxide would be primarily of $Fe_3O_4$, that has by orders of magnitude lower values of heat of reaction than $Al_2O_3$.

Fig. 4 shows the amount of hydrogen produced due to in-vessel oxidation of components. Similarly as in the case of heat production; the hydrogen produced from the oxidation of SiC and FeCrAl is lower, about half that of the Zr alloy. Hydrogen production by SiC oxidation is comparable to that of FeCrAl using the only $Al_2O_3$ approach. There are certain differences between FeCrAl approaches; the lowest total amount of hydrogen is in case of weighted $Fe_3O_4$ approach, where by an amount of the reacting base metal a lower amount of hydrogen is produced.

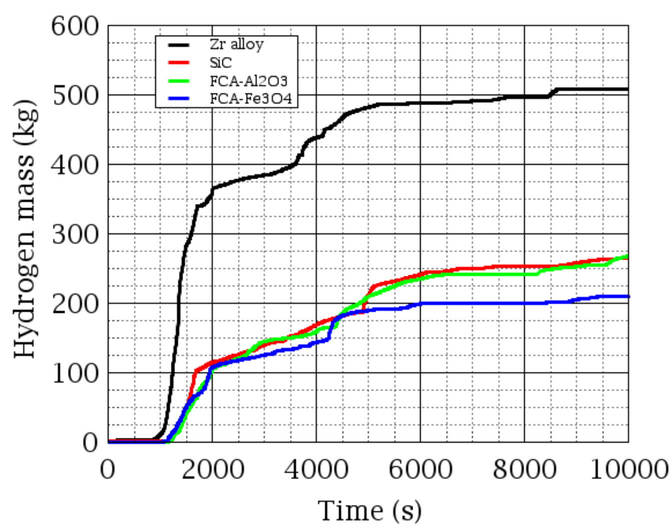



Fig. 4. Total hydrogen mass produced due to in-vessel oxidation of components.

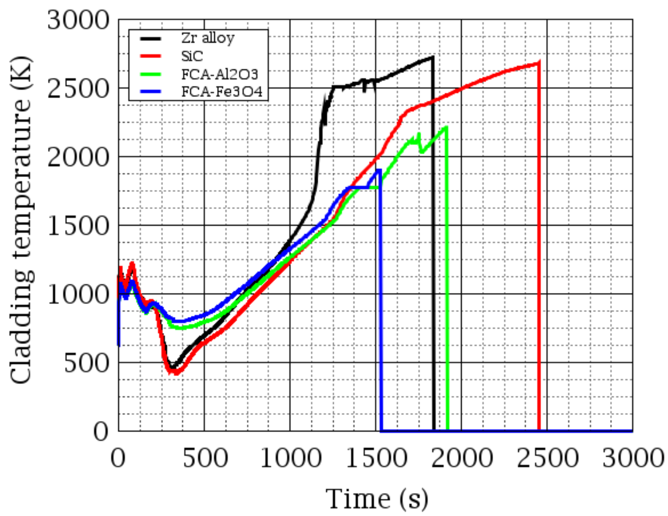



Fig. 5. Cladding temperature in core axial node 12, radial ring 2 - the first node that melts and collapses.

Fig. 5 shows the surface temperature of the cladding tube in the radial ring 2 and the axial node 12. This node is the first to collapse due to melting of the cladding in the case of the Zr alloy and FeCrAl. Because of the high melting point of SiC, the fuel inside the rod is melted first, resulting in the collapse of the cladding tube by the damage function.

The oxidation heat of the Zr alloy has effects on temperature, and, as can be seen, the rate of increase in temperature is higher than that of other materials, also the $ZrO_2$ thermal conductivity is lower, reducing the amount of heat transferred. The FeCrAl alloy melts at lower temperatures as a result of its lower melting point. SiC shows the latest collapse time because it has a higher thermal conductivity and melting point compared to other materials.

In Table I, times of RPV lower head and the first cladding failures are listed. In all cases of the use of ATF cladding, lower head failures occurred later compared to Zr alloy cladding. As mentioned above, both FeCrAl and SiC materials exhibit lower cumulative hydrogen production, reducing the risk of structural damage posed by possible hydrogen ignition.

The report [11] examines the influence of the ATF cladding on the TMI-2 accident scenario and shows that the amount of oxidation heat and hydrogen oxidation produced has been significantly reduced. Compared to the production rates, considering that the oxidation in this study progresses at higher temperatures, where the oxidation is faster, it is possible to state that the results are correlating in trends.

## VI. Conclusion

In this study, the ATF cladding was simulated using the MELCOR code in a severe accident scenario. Two different materials, advanced nuclear grade stainless steel C35M and ceramic SiC, were defined and simulated in the advanced power reactor model. The results have shown a significant decrease in the production of oxidation heat and total hydrogen

TABLE I
Times of RPV and first cladding failures and cumulative hydrogen production

| | RPV failure (s) | First clad failure (s) | $H_2$ mass (kg) |
|---|---|---|---|
| Zr alloy | 8435 | 1835 | 509,1 |
| SiC | 10215 | 2450 | 265,3 |
| FCA-$Al_2O_3$ | 9555 | 1915 | 267,9 |
| FCA-$Fe_3O_4$ | 9960 | 1525 | 210,8 |

mass, however, in the case of FeCrAl, with a trade-off of a lower melting point, which may result in a sooner breakdown of the cladding. Research is needed on how oxides are formed by oxidation of FeCrAl alloys as there are still many estimates, primarily regarding their thermal properties and composition. Furthermore, the generic oxidation model would require further optimisation and verification, especially when modelling the heat released from the material by oxidation and its kinetics.

## Acknowledgment

This research work has been carried out at the Centre for Research and Utilization of Renewable Energy (CVVOZE). Authors gratefully acknowledge financial support from the Technology Agency of the Czech Republic (project No. TN 02000012/001).

## References

[1] K. G. Field, Y. Yamamoto, B. A. Pint, M. N. Gussev, and K. A. Terrani, Accident tolerant FeCrAl fuel cladding: Current status towards commercialization. In: The Minerals, Metals & Materials Series, Springer International Publishing, 2017, pp. 1381–1389.

[2] K. G. Field, Handbook on the material properties of FeCrAl alloys for nuclear power production applications. ORNL, 2018.

[3] E. S. Sarkisov, N. T. Chebotarev, A. A. Nevzorova, and A. I. Zverkov, Oxidation of zirconium at high temperature and structure of the primary oxide film. The Soviet Journal of Atomic Energy. 1958, vol. 5, no. 5, pp. 1465–1470.

[4] M. Steinbrück, and M. Böttcher, Air oxidation of Zircaloy-4, M5® and ZIRLO™ cladding alloys at high temperatures. Journal of Nuclear Materials, 2011, vol. 414, no. 2, pp. 276–285.

[5] T. Koyanagi, Y. Katoh, G. Singh, and M. Snead, SiC/SiC cladding materials properties handbook. ORNL, 2017.

[6] K. A. Terrani et al., Silicon carbide oxidation in steam up to 2 MPa. Journal of the American Ceramic Society, 2014, vol. 97, no. 8, pp. 2331–2352

[7] B. Qiu et al., A review on thermohydraulic and mechanical-physical properties of SiC, FeCrAl and Ti3SiC2 for ATF cladding. Nuclear Engineering and Technology, 2020, vol. 52, no. 1, pp. 1–13.

[8] P. Mician, T. Cerny, and K. Katovsky, Simplified Modeling of Natural Circulation during Station Blackout with MELCOR. *2023 23rd International Scientific Conference On Electric Power Engineering, EPE 2023*, 2023.

[9] T. Cerny, P. Mician, K. Katovsky and M. Ptacek, Impact of External Injection from FLEX Equipment during LOCA in APR1400. *2023 23rd International Scientific Conference On Electric Power Engineering, EPE 2023*, 2023.

[10] C. Badini, and F. Laurella, Oxidation of FeCrAl alloy: influence of temperature and atmosphere on scale growth rate and mechanism. Surface and Coatings Technology, 2021, vol. 135, no. 2–3, pp. 291-298.

[11] B. J. Merrill, S. M. Bragg-Sitton, P. W. Humrickhouse, Modification of MELCOR for severe accident analysis of candidate accident tolerant cladding materials. Nuclear Engineering and Design, vol. 315, 2017, pp. 170-178.

# SUPER HOT CHIPS


ASICentrum
A COMPANY OF THE SWATCH GROUP

Naše srdce planou pro navrhování čipů!
Hledáme další nadšence, co by rozšířili naše řady!

## JOIN US!

ASICENTRUM PRAHA / BRNO

WE LOVE TO **DESIGN** THEM.

## ČIPY KTERÉ JINDE NENAJDETE

- Navrhujeme ULTRA LOW POWER integrované obvody
- Naše čipy mají vždy něco jedinečného, co konkurence neumí
- Bluetooth, RFID NFC, RFID UHF, Sensor hubs, Optical Sensors
- Na tom všem se můžete  podílet

## ČESKÁ FIRMA SE ŠVÝCARSKÝM ZÁZEMÍM

- ASICentrum je středně velká firma s rodinnou atmosférou
- Jsou zde špičkoví odborníci, od kterých se můžete hodně naučit
- Švýcarské zázemí díky mateřské firmě EM Microelectronic
- V centrále firmy EM Microelectronic ve Švýcarsku máme vlastní továrnu na výrobu integrovaných obvodů

## PŘIJĎ NAVRHOVAT ČIPY

- Nabízíme brigády, diplomky a zaměstnání
- Potřebujeme návrháře, verifikátory i softwarové inženýry
- Studenty naučíme moderní techniky návrhu integrovaných obvodů
- Zaměstnanci získají obrovské zkušenosti s celým procesem vývoje čipů
- Skvělé platové ohodnocení a benefity
- Co očekáváme – zájem, chuť do práce a nezbytné základní znalosti ze školy

## HLEDÁ SE KOLEGA HOMO TECHNICUS

- Digital IC Designer
- Digital IC Verification Engineer
- Embedded Software Engineer
- Digital Backend Engineer
- Analog IC Designer

hr@asicentrum.cz
**www.hotchips.cz**

em microelectronic
A COMPANY OF THE SWATCH GROUP

# Corrosion of Sintered Materials Based on Iron

Petra Slotová
*Brno University of Technology*
*Department of Electrical and Electronic Technology*
Brno, Czech Republic
xsloto00@vutbr.cz

Marie Sedlaříková
*Brno University of Technology*
*Department of Electrical and Electronic Technology*
Brno, Czech Republic
sedlara@vut.cz

***Abstract*—This paper addresses the topic of biodegradable bone implants. Presently, titanium alloys, known for their ability to reinforce bone structure, are utilized for fracture fixation. Nevertheless, post-healing, their extraction from the body becomes necessary. Biodegradable implants serve the same purpose but undergo gradual degradation upon exposure to bodily fluids. Consequently, the need for additional surgical procedures for their removal is obviated.**

***Index Terms*—biodegradable, corrosion, pH, conductivity, iron, sample**

## I. Introduction

This paper deals with Fe-based biodegradable sintered materials and their corrosion. Nowadays, mainly materials based on titanium compounds are used, which have excellent mechanical strength and properties, but after partial healing of the fracture, they must be surgically removed [1].

Biodegradable materials based on inorganic substances, such as iron, magnesium or zinc, could help prevent further surgical intervention in the body. The fracture healing process itself could be much better when using materials that are able to degrade in the body. The bone support will be gradually absorbed by corrosion processes and safely removed from the patient's body. Different compositions of such materials are investigated in different corrosion environments in order to evaluate different properties [2].

Corrosion is a spontaneous, gradual transformation of metals or non-metallic organic and inorganic materials. It arises as a result of a chemical or electrochemical reaction of the basic material with the external environment. It consists of an anodic and a cathodic reaction, which are interconnected and one cannot occur without the other unless an external current is passed through the corroding metal. The anodic reaction represents the oxidation of the metal. The cathodic reaction represents the reduction of oxygen of oxygen dissolved in the electrolyte [3].

## II. Preparation of Samples

The manufacturing of samples of biodegradable bone implants took place in the following steps. First of all, it was necessary to prepare the material for sintering. The mixture was created by mixing metal powder (Fe) with magnesium (Mg) and with different amounts of polystyrene (PS) for individual samples. The exact composition of the individual samples is written in Table 1. The next step after preparing the correct ratio of the individual samples was the sintering of the material. By sintering, a solid material was created, prepared for insertion into a corrosive environment. Due to the different weight of the individual samples, their dimensions and also the surface area are different. These are irregular blocks that have a width of 1 cm, a height of 0.5 cm and a length of 5 - 6 cm, depending on their weight. Subsequently, a saline solution (0.9% sodium chloride solution, NaCl) was mixed into which the individual samples were inserted. Each sample was immersed in 30 milliliters of solution and it was a static bath. The samples soaked in the solutions were then placed in an environment with a temperature of 37 °C, which had the task of simulating the temperature of the human body [4].

TABLE I
Composition of 4 different samples

| Sample composition | |
|---|---|
| Sample 1 | 9 g Fe, 1 g Mg, 1.5 g PS |
| Sample 2 | 9 g Fe, 1 g Mg, 1 g PS |
| Sample 3 | 9 g Fe, 1 g Mg, 0.5 g PS |
| Sample 4 | 9 g Fe, 1 g Mg, 2 g PS |

## III. Measurements

The aim of the measurements was to monitor the development of changes in pH and conductivity of individual solutions caused by the corrosion processes of the samples. Thanks to the monitoring of these changes, it is then possible to determine the speed of individual corrosion processes, or to determine the amount of metal released into the body. It is also subsequently possible to estimate whether the amount of metal released is in line with expectations or could have some adverse effect on the human body.

Before the very beginning of the measurements of individual samples, the pH and conductivity values of the saline solution were measured. A pure NaCl solution should have a pH value of around 7. The measurement found a value of around 7.5. As for the conductivity, around 51 mS/cm was measured.

The measurement of individual samples started 10 days after they were placed in the solutions. In the next subsections, the individual measurements are shown in more detail, and it can be seen that the pH and conductivity values of the solution changed very quickly after insertion.

The completion of this paper was made possible by the grant FEKT-S-23-8286 - "Materials and technologies for electrical engineering V".

## A. pH measurement

One of the measurements was the measurement of the pH of the solutions. Based on it, it is possible to find out how quickly the corrosion processes take place in individual samples and how the corrosion speed changes itself. From the measurements, it is possible to find out how much metal is gradually degraded and how much remains in the body. Subsequently, it is possible to roughly estimate whether the values are acceptable for the human body or could already be dangerous.

All four samples were placed in a saline solution, which had the task of simulating the environment of the human body and was kept at a temperature of 37 degrees Celsius for the entire time. The first measurement took place after 10 days of being placed in the solution and was subsequently repeated approximately every week for 3 months.

Changes in pH values are illustrated in Figure 1. It can be observed that at the beginning, when the surface of the material is clean, faster oxidation occurs. As a result, water begins to be reduced and hydrogen and OH- are formed, resulting in alkalization. However, the released OH- will begin to gradually bind to iron, which will cause the pH values to decrease again. For a closer overview, Table 2 also summarizes the actual numerical values from measurements at certain time intervals.

TABLE II
MEASUREMENT OF PH SOLUTIONS

| pH of solutions | | | | | | |
|---|---|---|---|---|---|---|
| **Number of days** | ***10*** | ***19*** | ***33*** | ***54*** | ***68*** | ***91*** |
| Sample 1 | 11.34 | 11.18 | 10.71 | 10.77 | 10.67 | 10.35 |
| Sample 2 | 10.76 | 10.66 | 10.54 | 10.44 | 10.42 | 10.02 |
| Sample 3 | 10.74 | 10.71 | 10.47 | 10.38 | 10.42 | 9.97 |
| Sample 4 | 11.30 | 11.28 | 1071 | 10.75 | 10.71 | 10.26 |

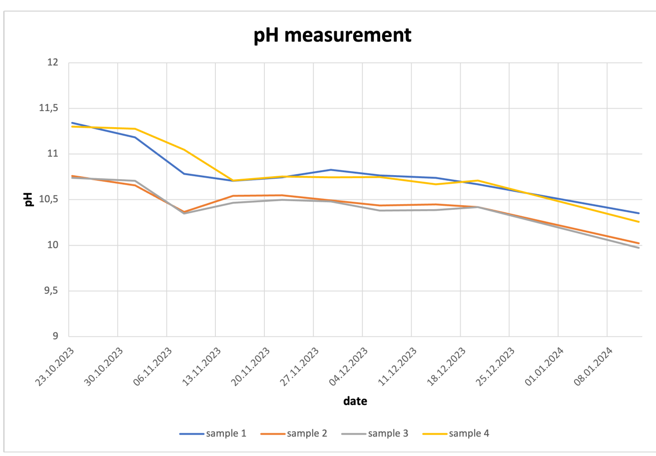



Fig. 1. Difference in corrosion solution pH depending on time.

## B. Conductivity measurement

The second of the measurements was the measurement of the conductivity of the solutions. The measurement was similar to that of pH. The first measurement also took place after 10 days from being placed in the solution and then repeated approximately every week for 3 months.

Changes in conductivity values are shown in Figure 2. The curves have a decreasing character of conductivity, which is caused by the process of precipitation of ions in the solution. It can be seen that while in sample 4 there was a rather sharp decrease in sample 2, the decrease in wonder was significantly slower. Table 3 is prepared for a closer overview of the values.

TABLE III
MEASUREMENT OF CONDUCTIVITY SOLUTIONS

| Conductivity of solutions [mS/cm] | | | | | | |
|---|---|---|---|---|---|---|
| **Number of days** | ***10*** | ***19*** | ***33*** | ***54*** | ***68*** | ***91*** |
| Sample 1 | 11.56 | 11.79 | 11.56 | 11.24 | 11.05 | 10.93 |
| Sample 2 | 11.37 | 11.35 | 11.17 | 10.80 | 10.61 | 10.53 |
| Sample 3 | 12.06 | 12.02 | 11.75 | 11.46 | 11.11 | 10.97 |
| Sample 4 | 11.13 | 11.30 | 10.96 | 10.58 | 10.05 | 10.03 |

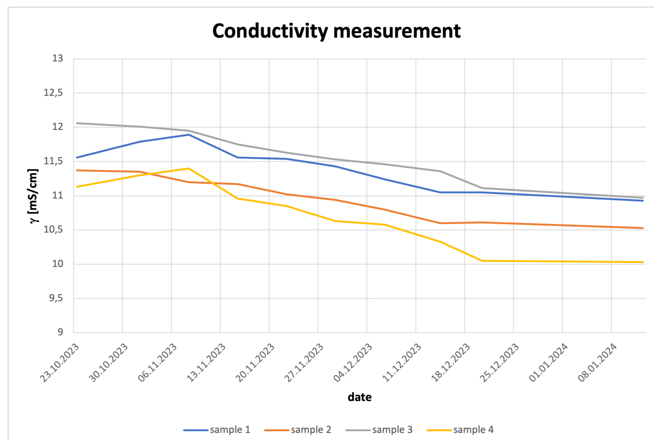



Fig. 2. Difference in corrosion solution conductivity depending on time.

# IV. ENERGY DISPERSIVE X-RAY ANALYSIS

Energy dispersive X-ray analysis, shortly EDAX, is an analysis that is used to determine the occurrence of elements in a sample and their percentage representation. The principle of this method is based on the generation of X-ray radiation, which is created by the impact of the primary beam of electrons on the examined sample. When an electron beam hits the sample, elastic scattering can occur, in which the beam of primary electrons is deflected by the force of the nucleus, its energy is reduced and the energy difference is emitted in the form of continuous radiation. In an inelastic collision, the electron is excited to a higher level and the atom goes into an unstable state. To ensure stability, the transition of an electron from a higher level to the place of a freed electron occurs, and the energy difference of these levels is emitted in the form of characteristic X-ray radiation. In energy dispersive X-ray analysis, there is an effort to detect characteristic radiation, because this radiation contains a narrow spectrum of energies and an element can be determined based on them [5].

EDAX analysis was performed before placing the samples in the solution (before corrosion) and 96 days after placing the samples in the saline solution (after corrosion). The tables below compare the values obtained from the analyzes for individual samples. An important role in determining the elemental composition could be played by the area from which the elemental analysis was made.

### *A. EDAX analyses of Sample 1*

The analysis of Sample 1 before soaking in NaCl showed that the most abundant element in the sample was iron. The next element that was found was carbon. It probably stayed there as the rest of the foam, which was made of polyurethane material. The amount of carbon was almost the same as the amount of magnesium. After wetting, the elemental composition changed. The amount of iron has decreased, which means that iron leaks out of the sample during corrosion. The amount of oxygen increases after wetting, which could be expected. Chlorine is also present in the sample, which is caused by the NaCl solution. The exact quantities are given in Table 4 and the composition of the sample elements is shown in Fig. 3.

TABLE IV
PERCENTAL AMOUNTS OF ELEMENTS PRESENT IN SAMPLE 1

| Sample 1 | | |
|---|---|---|
| **Elements** | **Sample 1 [%]** | **Sample 1 NaCl [%]** |
| Iron | 64.73 | 55.00 |
| Carbon | 14.75 | 7.00 |
| Magnesium | 14.71 | 15.63 |
| Oxygen | 5.81 | 21.81 |
| Chlorine | - | 0.57 |

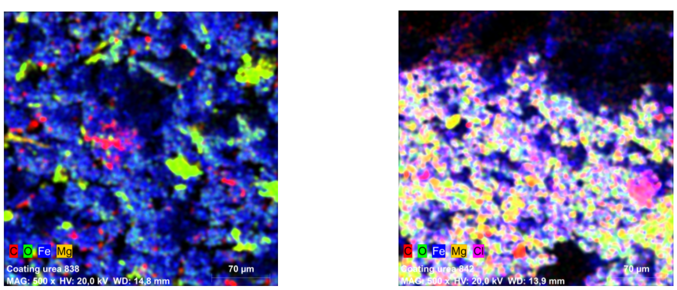



Fig. 3. Distribution of elements in Sample 1 before and after wetting in NaCl

### *B. EDAX analyses of Sample 2*

For Sample 2, the analysis before wetting showed similar results as for Sample 1. The most represented element is iron, followed by carbon. After soaking in the solution, the amount of iron and carbon decreases again. The amount of magnesium and oxygen increases, and a trace amount of chlorine also appears. The exact amount of individual elements is again shown in Table 5, and the composition of the elements of the sample before soaking in the solution and after soaking is shown in Figure 4. In the individual sample images it can be seen that the individual elements are evenly distributed in each pattern.

TABLE V
PERCENTAL AMOUNTS OF ELEMENTS PRESENT IN SAMPLE 2

| Sample 2 | | |
|---|---|---|
| **Elements** | **Sample 2 [%]** | **Sample 2 NaCl [%]** |
| Iron | 70.48 | 48.35 |
| Carbon | 13.67 | 6.54 |
| Magnesium | 10.86 | 19.18 |
| Oxygen | 4.98 | 24.97 |
| Chlorine | - | 0.96 |

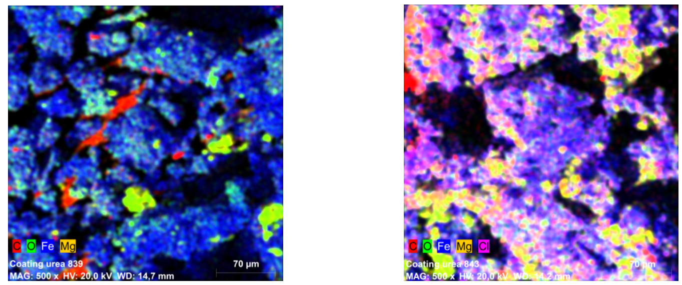



Fig. 4. Distribution of elements in Sample 2 before and after wetting in NaCl

### *C. EDAX analyses of Sample 3 and Sample 4*

EDAX analysis of Samples 3 and 4 turned out very similar to the previous two. In any case, iron had the largest presence, the amount of which decreased after soaking. The only difference is that in Sample 3, chlorine elements were not found after soaking as in the previous two. On the other hand in Sample 4, after soaking, not only chlorine, but also sodium was recorded, both appeared there due to NaCl solution. The exact amounts of individual elements are again clearly shown in Tables 6 and 7.

TABLE VI
PERCENTAL AMOUNTS OF ELEMENTS PRESENT IN SAMPLE 3

| Sample 3 | | |
|---|---|---|
| **Elements** | **Sample 3 [%]** | **Sample 3 NaCl [%]** |
| Iron | 60.27 | 56.92 |
| Carbon | 4.98 | 6.43 |
| Magnesium | 25.33 | 19.61 |
| Oxygen | 9.42 | 17.04 |

TABLE VII
PERCENTAL AMOUNTS OF ELEMENTS PRESENT IN SAMPLE 4

| Sample 4 | | |
|---|---|---|
| **Elements** | **Sample 4 [%]** | **Sample 4 NaCl [%]** |
| Iron | 78.69 | 55.43 |
| Carbon | 13.97 | 719 |
| Magnesium | 5.25 | 16.14 |
| Oxygen | 2.08 | 19.45 |
| Chlorine | - | 0.69 |
| Sodium | - | 1.1 |

ACKNOWLEDGEMENT

The completion of this paper was made possible by the grant FEKT-S-23-8286 - ”Materials and technologies for electrical engineering V" financially supported by the Internal science fund of Brno University of Technology.

## CONCLUSION

This work is primarily focused on the long-term monitoring of the degradation of samples in physiological solutions. Conductivity measurement, pH measurement and EDAX analysis were used for qualitative and quantitative assessment of degradation. Four different samples that differed in the amount of polystyrene were examined and compared. It can be clearly seen from the measured curves that both pH and conductivities have a decreasing character. From the EDAX analysis, it is possible to say that the iron really dissolves and decreases with time. In the individual images of the samples, it can be seen that the individual elements are evenly distributed in each pattern. In conclusion of this work, it can be said that iron-based materials containing magnesium and polystyrene really show a relatively satisfactory rate of degradation. However, it is necessary to observe these measurements over a longer time horizon and find out how they will behave after a longer period than three months.

## REFERENCES

[1] Sun, Jie, Annika Hämmerle, Günter Fafilek, et al. Electrochemical investigation for understanding the bactericidal effect of Cu2Se and Ag2Se for biomedical applications. Journal of Applied Electrochemistry. 2021, (52), 15, . DOI : https://doi.org/10.1007/s10800-021-01617-2

[2] Biodegradable materials for bone defect repair. BMC, research in progress [online]. 2020 [cit. 2024-03-07]. Dostupné z: https://mmrjournal.biomedcentral.com/articles/10.1186/s40779-020-00280-6

[3] M. Sedlaříková, M. Zatloukal1, J. Kuchařík, P. Čudek, G. Fafilek, E Doleželová, "Corrosion processes of sintered materials based on Fe," Journal of Physics: Conference Series, vol. 2382, Nov. 2022. doi:10.1088/1742-6596/2382/1/012019

[4] Hrubovcakova M., Kupkova M., Dzupon M. Fe and Fe-P Foam for Biodegradable Bone Replacement Material: Morphology, Corrosion Behaviour, and Mechanical properties. September 29, 2016. Institute of Materials Research of SAS, Kosice, Slovakia.

[5] Scimeca, M., Bischetti, S., Lamsira, H. K., Bonfiglio R., Bonanno, E. Energy Dispersive X-ray (EDX) microanalysis: A poweful tool in biomedical research and diagnosis. European Journal of Histochemistry. 2018, 62, . DOI: https://doi.org/10.4081/ejh.2018.2841

# *Parking Areas at BUT for energy use*

Enas Al Halabi, Martin Paar,
*Department of Electrical Power Engineering*
*Brno University of Technology*
Brno, Czech Republic
203498@vut.cz, paar@vutbr.cz

***Abstract*— The article focuses on the concept of solar carports in parking lots for energetic utilization and charging electric vehicles (EVs), describes their structure, components, configuration, also paper presents an overview of the analysis of selected parking areas at Brno University of Technology (BUT) for energy utilization, finally a designed project of solar carports in these areas are highlighted.**

***Keywords*— *Solar energy, PV system, Electric vehicle, Car parking lots, EVs charging, Solar carport***

## I. Introduction

The concept of solar parking lots aims to coupling the development of clean solar electricity and electric mobility [1].

Solar carports are multifunctional structures that turn everyday parking spaces into sustainable power sources. They do not only protect vehicles from weather conditions but also harness solar energy, leading to monetary benefits. They offer a variety of designs that cater to different needs, whether it is for a home or a vast commercial space. Besides their economic and environmental benefits, they offer smart charging which has many optimizations on grid and the power storage in the case of smart solar carport.

This paper is divided into two basic parts. Chapter 1 outlines the definition, structure, configuration and overview of the solar carport. An overview of designed carports in selected parking areas at BUT are presented in chapter 2.

## II. An Overview of the solar carport

The solar carport is defined as a parking area provided with an overhead canopy that offers shelter for vehicles besides generating electricity from the solar PV system. Since is similar to the ground-mounted panels, that are installed on the ground, it differs from the ground-mounted panels not just in structure but also in space efficiency, whereas the solar carports are taller and more space-efficient [2].

### *A. Design and structure*

Solar carport could be classified depending on the next points:

**Size:** A solar carport could cover one, two or more rows [2], and can be for one car (single carport), two cars (solar carport double) or many cars (solar carport multi) fig.1 [3], these types match residential, commercial and industrial settings.

**Frames:** Frames may be wooden or metallic from Aluminum or steel [2], designs L, T, Y, W, N. [4]

**Shape:** Carport roof can be tilted in one direction upwards, tilted and curved in one direction, extremely slightly tilted, tilted in two directions like east-west, double row [5].

Fig. 1. Solar carport multi, pitch, single column, T structure [6]

### *B. The configuration and operation of solar parking lots*

The solar parking lots could be off-grid or on-grid, providing PV charging alone or PV-grid charging, the second one is more frequent, Conceptually, the layout of an Electric Vehicles Solar Parking Lots (EVSPL) is as mentioned in Fig.2. Vehicles are connected to a DC link after a DC-DC converter. Or an AC link after a DC-AC inverter a Control Center (CC) is responsible for managing the system, system can be with battery system or without it [1].

The energy production of the PV system is variable. On the other hand, the charging requirements of the EVs vary according to the behavior of the users, this makes the coupling of the two a problem. To avoid this, many variables must be considered with a minimal time-step:

Solar generated electric power, distribution in time of cars arriving, the parking duration of each car, state-of-charge of the batteries at the time of vehicles arrival, energy demand of each vehicle [1].

For maximizing solar EV charging, PV production must be matched as much as possible to the EV load profiles. This is considering in the controlled charging.

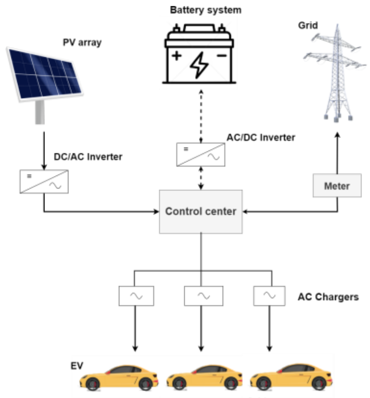



Fig. 2. Configuration of a PV EVSPL (AC charging)

## III. Analysis of Selected Parking Areas at BUT for Energy Utilization

### A. The selected parking areas for the solar carport at BUT

The selection was between four parking areas near faculty FEEC, for the solar carport the parking areas Technická 12 (T12), Technická 14 (T14), and Centre of Sports Activities of BUT (CESA) are suitable. The largest one is T12; however, the shading analysis indicates that not the entire area is suitable for PV installation.

Fig. 3. Location of the selected parking areas for solar carport

### B. Design of solar carports for three car parks on the university campus

The designing of photovoltaic systems was done in the software PV SOL premium 2024, as 3D analysis. For taking photos of car parks and dimensions calculation application Google Earth Pro was used, dimensions of the parking building and surrounding buildings also were verified from photos taken personally in the car park locations and a tour of the site, the values are approximate.

As the first step, parking areas were selected, followed by the design of the PV system, as a roof-mounted PV system. This process was carried out in the following steps: Painting a 3D structure of the parking houses and surrounding objects, define solar array and placement of panels, shading analysis, calculation of solar output capacity. Sketches of the designed solar carports in car parks are shown in Fig.4, Fig.5, Fig.6.

Fig. 4. Sketch of the designed solar carports in car park T14

Fig. 5. Sketch of the designed solar carport in car parks T12

Fig. 6. Sketch of the designed solar carport in car parks CESA

### C. Parameters of PV panels in carports

As photovoltaic panel in all systems module FU 500 Silk premium is chosen, which has the following parameters:

PV power output: 500 Wp, dimensions: 2185 x 1098 x 35 mm, type: monocrystalline, half-cut cells, efficiency: 20.84%.

### D. Parameters of carports structures:

#### 1) The high and shape:

In general, the height of carports is around 2.5 m and 4.5 m [7], the height of structure here 2.7 meters (high of bottom edge) in T14 and T10, the highest floor in parking areas T12. However, in T12 the structures in middle floor have a high of 3.7 m and in the lowest floor has high of 4.7 m. The structure in all car parks mounted with a slope of 15 degrees, structures can be from aluminum with shape of single column, T structure, titled 15 degrees in one direction like in Fig.1.

*2) Solar carport width:*

The width of solar carport roof is in general about 3 to 7 m, according to size and number of car rows to cover the cars [3].

In T14: carport roof has 3 vertical modules on row width, then depth of array row is 6.34 m.

In T12 and CESA: carport roof has 5 horizontal modules on row width then the depth of array row is 5.3 m.

### *E. Solar PV capacity and sufficient parameters of systems*

In the following table the other sufficient parameters are determined for designed solar carports:

*Table 1: sufficient parameters determined for designed solar carports.*

| **Car park** | **PV output (kWp)** | **Modules number** | **Array direction** | **Number of areas** |
|---|---|---|---|---|
| T14 | 177 | 354 | South-west | 1 |
| T12 | 350 | 700 | South-west<br>South-east | 3 |
| CESA | 182.5 | 365 | South | 3 |

### *F. Shading analysis:*

The system was designed with consideration the shading of surrounding objects and the shading of solar modules, shading analysis was performed, the allowed limit was chosen as 5%, that means the shading on the solar panel did not exceed more than 5%, as seen in the following figures.

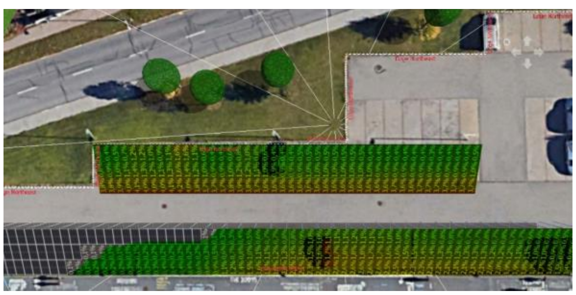


Fig. 7. View of shading analysis of designed solar carport in T14 car park, maximum 4.2%

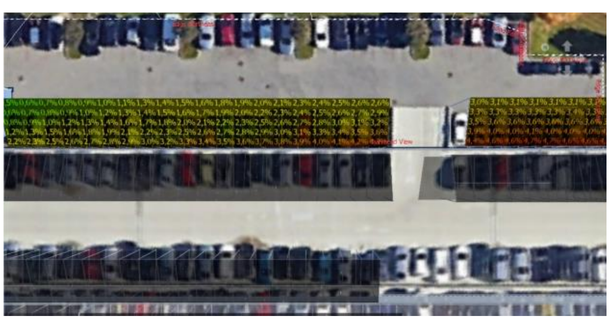


Fig. 8. View of shading analysis of designed solar carport in T12- first area car park, maximum shading 4.7%

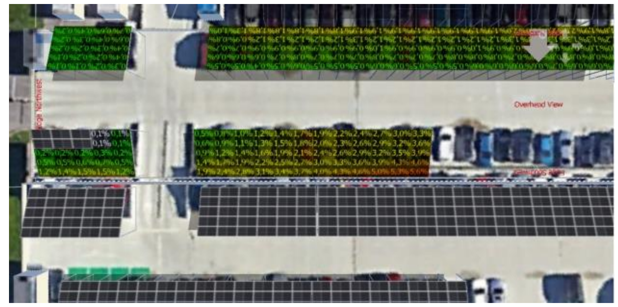


Fig. 9. View of shading analysis of designed solar carport in T12-middle area car park, maximum shading 5.6%

Fig. 10. View of shading analysis of designed solar carport in T12- middle area car park other side, maximum shading 2.4%

Fig. 11. View of shading analysis of designed solar carport -upper area car park, maximum shading 5.1%

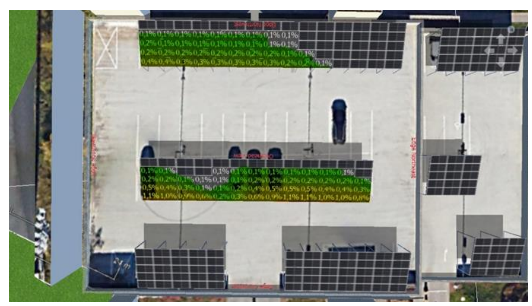


Fig. 12. View of shading analysis of designed solar carport in CESA parking house - first area (upper), maximum shading 1.1%

Fig. 13. View of shading analysis of designed solar carport in CESA parking house - second area, maximum shading 6% (one panel), 2.9%

Fig. 14. View of shading analysis of designed solar carport in CESA parking house -third area, maximum shading 3.4%

## IV. REQUIREMENTS FOR CONNECTING CATEGORY B1 PRODUCTION FACILITIES TO THE LOCAL DISTRIBUTION SYSTEM

The following rules summarize the requirements that need to be considered when connecting category B1 electricity production facilities to the MV network of the Local Distribution System Operator (PLDS):

**Category of Production Module:** B**, limit:** 1 MW**, Subcategory :** B1, **limits PLDS :** ≥ 100 kW and < 1 MW **Requirements:** According to article 14 for production modules B, article 20 for asynchronous production modules category B [7].

### *A. Relevant Legislation and Regulations*

For the establishment of a production facility and electrical storage device, it is necessary to comply with current regulations and standards in PPLDS Attachment 4: Rules for Parallel Operation of Production Facilities and Storage Devices with the Distribution System Operator's Network [7] to ensure suitability for parallel operation with the DSO network and to eliminate disruptive back feeding effects on the network or devices of other consumers.

## V. CONCLUSION

In conclusion, using solar carport carried many benefits, from point of energy production for EVs charging. Electric vehicles consume an average of 0.2 kWh/km, and photovoltaic (PV) power can provide up to 60% of the energy needed for charging electric vehicles [8]. Also, it offers economic benefit, so smart EVSPLs offer additional services that reduce charging costs for both EV owners and grid operators [1].

The article provided an essential overview of solar carports, evaluate the use of selected parking areas owned by Brno University of Technology in terms of installation of PV systems. The solar capacity of parking parks, in terms of generated power output, which the designed photovoltaic system could generate as mentioned in Tab.1: In parking area **T14: 177 kWp**, in parking area **T12: 350 kWp**, in parking area **CESA: 182.5 kWp.**

Later, the aim of the study will be to evaluate the use of selected parking areas at BUT in terms of parked electric vehicles within the vehicle-to-grid concept. In addition, will explore the possibility of EV charging, and the utilization of EVs as battery storage. The study considerations will include time-of-use analysis of the parking lots and driver behavior patterns.

## REFERENCES


[1] NUNES, Pedro, Raquel FIGUEIREDO and Miguel C. BRITO. 2016. The use of parking lots to solar-charge electric vehicles. Renewable & sustainable energy reviews OXFORD: Elsevier, 66, 679-693[online]. J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.

[2] R. Istók, P. Kádár and L. Reizer. 2021. Solar Carport with Energy Storage Capabilities. IEEE 4th International Conference and Workshop Óbuda on Electrical and Power Engineering (CANDO-EPE), Budapest, Hungary, 2021, pp. 89-94, [online].

[3] LETS GROW S.R.O. SOLAR CARPORT [online]. Available at: https://www.solarcarport.cz/blog/

[4] What are the Different Solar Carport Designs? Energytheory [online]. Available at: https://energytheory.com/types-of-solar-carpor/

[5] Kensington Dual-Pitch Solar Canopy. ABLE CANPOIES [online]. Available at: https://www.ablecanopies.co.uk/products/canopies/solar-carports-canopies/kensington-dual-pitch-solar-canopy/

[6] Multi-Housing Carport Designs. CARPORT STRUCTURES CORPORATION. [online] Available at: https://www.carportstructures.com

[7] ENERGETICKÝ REGULAČNÍ ÚŘAD. *PPLDS Příloha 4:*PRAVIDLA PRO PARALELNÍ PROVOZ VÝROBEN A AKUMULAČNÍCH ZAŘÍZENÍ SE SÍTÍ PROVOZOVATELE DISTRIBUČNÍ SOUSTAVY. 2014.

[8] SHEPERO, Mohamoud, Joakim MUNKHAMMAR, Joakim WIDÉN, Justin D.K. BISHOP and Tobias BOSTRÖM. 2018. Modeling of photovoltaic power generation and electric vehicles charging on city-scale: A review, Renewable and Sustainable Energy Reviews, Elsevier [online]. Available at:

https://www.sciencedirect.com/science/article/pii/S1364032118300625

# Data acquisition, visualisation and data processing of measuring of electrical energy consumption in BUT buildings

1st Radim Dvořák
*Department of Electrical Power Engineering*
*Brno University of Technology*
Brno, Czech Republic
220760@vut.cz

2nd Daniel Janík
*Department of Electrical Power Engineering*
*Brno University of Technology*
Brno, Czech Republic
janikd@vut.cz

***Abstract*—The implementation of energy management is beginning to play a crucial role in terms of reducing environmental impacts in the long term. Brno University of Technology (BUT) has decided to start implementing this system to measure up to 95 % of energy outlets in individual buildings and campuses of BUT. Specifically, it addresses the deployment of detailed submetering for electrical energy consumption and the creation of a user-friendly environment for working with the collected data. Included is the establishment of a database for data storage and enabling communication between measurement elements at BUT and the storage system.**

***Index Terms*—Energy management, KMB, Grafana, Node-Red, InfluxDB, NAS, data collection, data processing, data visualisation.**

## I. Introduction

Energy management is a key tool for efficient and sustainable use of energy in buildings, industry and organisations. It aims to optimise energy consumption, minimise costs and reduce negative environmental impacts. According to the ČSN EN ISO 50001 standard, the energy management system (EnMS) is based on the Plan-Do-Check-Act (PDCA) cycle which can be seen in fig. 1. This approach provides a framework for continuous improvement and integrates energy management into existing organizational procedures.

A key element in energy management processes is the systematic and regular measurement and collection of energy consumption data. These data provide the necessary basis for analysis and decision-making processes. The analytical approach enables the identification of potential energy savings and the optimisation of energy systems based on detailed assessments.

This is followed by planning and implementation of measures, where specific measures are proposed to improve energy efficiency based on the analysis and evaluation. Continuous monitoring and control of the results of the implemented measures are necessary to maintain the achieved savings.

The importance of data visualisation lies in the user-friendly display of measured data and the rapid identification of deviations. With the help of visual outputs, changes in consumption or other variables can be easily and quickly identified and reacted to. In parallel, the metering system is a key element to obtain real-time information on energy consumption.

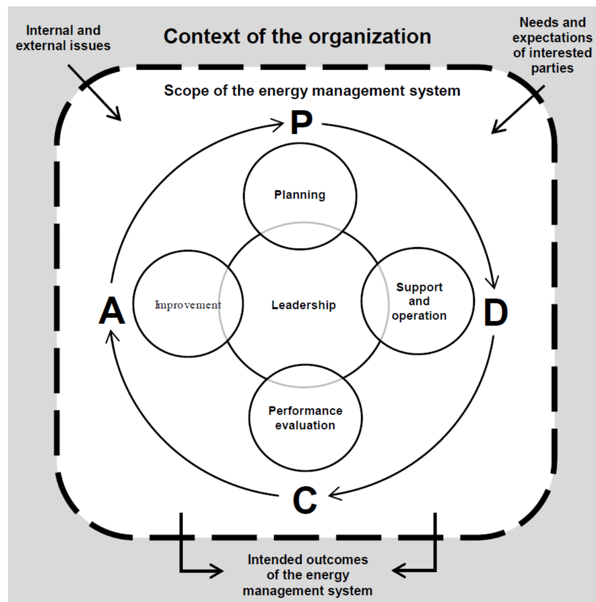



Fig. 1. Plan-Do-Check-Act Cycle [1]

The database is important for data storage and management, which is a basic requirement for data archiving and subsequent evaluation. It is also necessary to ensure reliable communication between the metering elements, which is essential for the collection of up-to-date data and the proper functioning of the whole energy system.

Once the measurement, data collection, visualization and database are in place, a thorough evaluation of the results is crucial. This includes a detailed analysis of the savings achieved, identification of areas for further optimization, and designing concrete steps to maximize the overall energy efficiency of the organization. Data evaluation is an essential step in the continuous improvement cycle, enabling organisations, buildings or industry to respond flexibly to dynamic changes in the energy environment. [1]

In practice, DIN rail meters are used for secondary metering of electricity consumption. The disadvantage of these meters is that they take up a lot of space in the distributor. This was not possible for our purposes, so we turned to multi-channel power consumption monitors. One of the manufacturers of this technology is Blue Panther instruments with their ELNet MC products, which comes in 3 versions, for measuring 2, 8 or 12 current outlets. The reason why these meters were not used was because of the impossibility of potential expansion. Therefore, the KMB instruments were chosen. Compared to other solutions, this solution is advantageous mainly because of its good price per measured outlet while maintaining good accuracy and also because it measures not only consumption, which is crucial for energy management, but also a number of other parameters that provide additional valuable information about the state of the network.

## II. Measurement system

For the implementation of energy management at BUT, instruments from KMB will be installed, which deal with energy management, power quality and software for configuration and access to the instruments. The instruments BCPM 233.012, EMI 12, EMI 12 FLEX will be used for remote monitoring of power consumption and power quality and are designed for measurements in 3x230/400 V, 50 Hz networks. These instruments are suitable for a wide range of applications in the power industry. They are most commonly used in smart grids, for automation of production processes and buildings, and for remote monitoring of infrastructure.

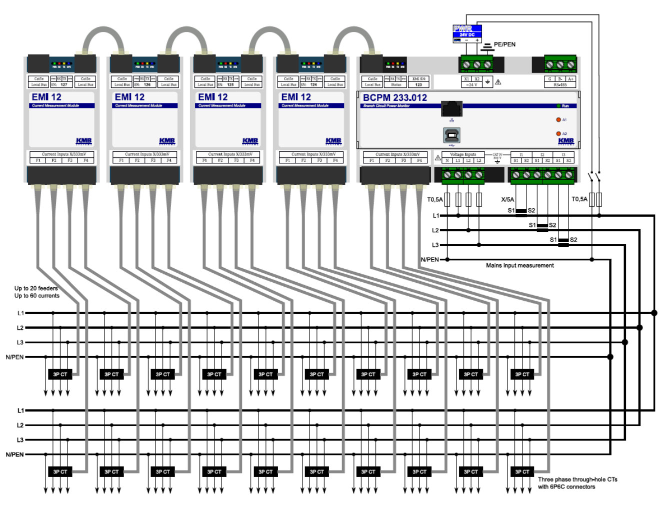



Fig. 2. Example of a common measurement circuit. [2]

The PQ analyzer BCPM 233.012 measures three voltage and current values, which it uses as input values. Without additional modules, it can measure the basic electrical parameters of up to five three-phase circuits (15 currents) with a sampling frequency of 28.8 kHz. Main current inputs use current measuring transformer with the conversion of X/5 A and the rest use X/333 mV. Using the LocalBus, up to four expansion modules can be connected to it, which can be either EMI 12, EMI 12 FLEX or EMI 12 HALL. Up to twenty-one circuits (63 currents) can be measured with these expansion modules. From the measured values of voltages and currents, it can calculate or evaluate voltage and current characteristics, which include rms values and phasors (of voltages and currents), total harmonic distortion, individual harmonics (up to 128 components) and interharmonic components, active, reactive, apparent powers, distortion and fundamental active and reactive powers, rapid voltage changes, voltage events, voltage deviations and asymmetries, flickr, frequency, power factors, HDO signals. In addition, the instrument allows the measurement of DC quantities and systems with nominal frequencies of $16\frac{2}{3}$ Hz, 80 Hz, 400 Hz or frequency converters up to a nominal frequency of 500 Hz.

Recording of measured data is done using a built-in real-time circuit with battery backup and memory, for recording measured data and events with a capacity of 512 MB, and an aggregation interval of 200 ms to 24 hours.

RS-485, Ethernet and USB communication interfaces are used for data transfer, instrument setup and firmware updates. An Ethernet interface with RJ-45 connector will be used for remote communication with the instrument, to access the instrument and obtain measured data using the Modbus TCP/IP protocol.

In fig. 2 is an example of a circuit where five modules are connected by one LocalBus. The BCPM 233.012 automatically detects the connected modules and configures them. The measured values are immediately available after connection and the user can configure the newly detected modules. [2].

## III. Software solution

As a software solution for the database and the visualisations, Node-Red in connection with InfluxDB and Grafana was chosen. This solution was used after an extensive search for open-source solutions. Synology Inc.'s DiskStation DS1522+ was selected as the platform to run Node-RED, Grafana and InfluxDB software due to its cybersecurity measures. The storage is currently composed of two 8 TB drives in RAID, out of a total of five which leaves room for expansion. This Network Attached Storage (NAS) device has built-in security features such as firewall, SSL/TLS support for communication encryption, and the ability to deploy more security measures. The interconnection of the individual meters, NAS and software is shown in Fig. 3.

The communication between the meters and the NAS is via a local Ethernet network. This network is directly connected to the DiskStation DS1522+ and is isolated from the public Internet and is solely used for communication between the meters and the storage. This was done to reduce the risk of potential cyber threats and external attacks, contributing to the overall security of the system.

Node-RED, Grafana and InfluxDB are all open source applications, which means that their code is open for auditing and review. Because of this identifying potential security threats and vulnerabilities can be done quicker which enables

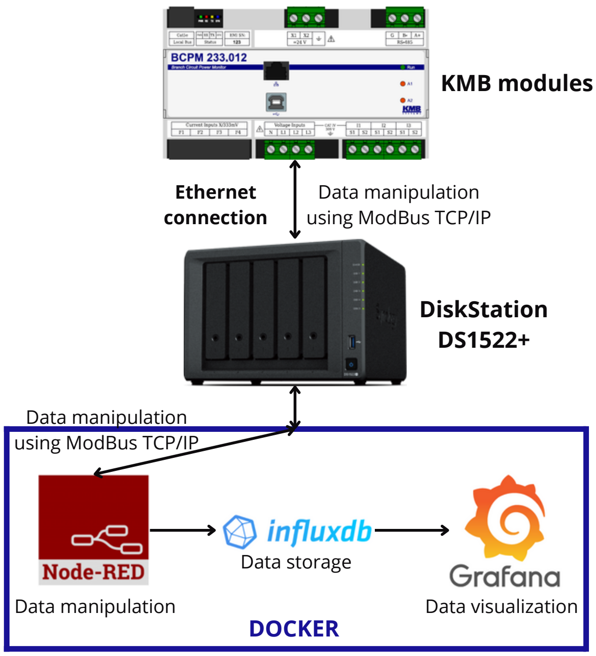



Fig. 3. Measurement, data collection, visualization and database solution.

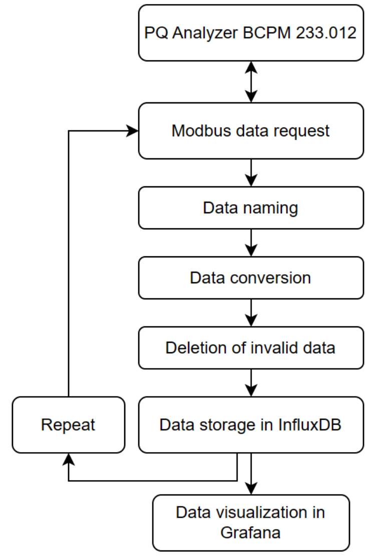



Fig. 4. Flow chart of data processing and visualization.

quick responses to potential problems. Synology DiskStation DS1522+ was chosen for its ability to run Docker containers, which allows for application isolation and minimizes the risk of potential threats transferring data between applications. Due to this, a higher level of security and integrity of data stored and processed by Node-RED, Grafana and InfluxDB applications can be achieved, which is important from a cybersecurity perspective and can be used in an effort to minimize the potential risks associated with running these applications on the BUT network.

## A. Node-RED

Node-RED is a flow-based, visual programming development tool. It provides a web browser-based flow editor, which can be used to create data manipulation and data processing. It is widely used in the areas of Internet of Things (IoT), home automation and data management. It is easy to use and has a lot of libraries that make it easier to use. [4]

In this case, Node-RED is used to periodically retrieve actual measurement data from all the nodes of the BCPM 233.012 instruments at minute intervals using Modbus TCP/IP protocol using the node-red-contrib-modbus library. The readings from the instruments are retrieved using a local Ethernet network that was created for the purpose of measuring power consumption within the BUT buildings. To get all of the measured data from one device, more than 40 ModBus requests need to be made getting 70 or more registers at one time. The amount of data that will be sent by one query will be up to 1120 bits (140 bytes). There will be 6 BCPM 233.012 meters installed on buildings T8, T10 and T12, therefore up to 6720 bits (840 bytes) can be sent to the NAS at one time. The Modbus protocol allows reading up to a maximum of 125 registers in a single request, therefore the requests need to be divided in time and the response can take up to 200 ms. The received data is than converted from two integer values 16 b each to a float decimal, since one value is composed of 32 b, which is two registers and names are added to the data. The processed data is then sent to InfluxDB for storage and further processing using the node-red-contrib-influxdb library.

## B. InfluxDB

InfluxDB is an open-source time series database that is optimized for storing and querying time-series data. From the beginning, this database has been developed for time-series data, making it the ideal solution for storing measured and processed data from BCPM 233.012 instruments. A key feature of InfluxDB is its ability to process millions of records per second, performing various types of analyses and aggregations over the data. [5] Processed data from Node-RED, is received and stored with a time stamp which makes it easy to track historical trends, analyze fluctuations in energy consumption. The estimated amount of data that will be stored from all meters is each day is 64 MB, with the main voltage measurement, phase and cumulative, its harmonic distortion, phase shift, direct, inverse and zero component being stored. The currents, harmonic distortion, phase shift, direct, inverse and zero component, power factor, active, reactive, apparent and stray power of each phase and also three phase will be read for each of the 115 measured outputs. With these readings it would take up 23 GB of storage per year. It is also intended to read voltage and current harmonics (amplitudes and phases) up to the 50th power. However, these are not yet shown in the estimation, as the extent of their reading is not decided yet. The data stored is then ready for subsequent integration with

Grafana, which will perform visualizations of the measured electricity consumption data.

### C. Grafana

Grafana is a powerful open-source software used for data visualization. It is not tied to its own database, but is rather used to display data. [3] Therefore, it will be used in combination with InfluxDB. Due to this integration, data from InfluxDB will be presented in the form of graphs and dashboards. This will make it easy to monitor trends in electricity consumption, analyse the evolution of the measured data and if necessary take steps to reduce electricity consumption. The measured data will be presented through different types of visualisations that can be further customised if required. A major advantage of Grafana is its easy-to-use environment, which will be a great benefit in terms of subsequent use by the users. In fig. 5 we can see a example visualization in Grafana.

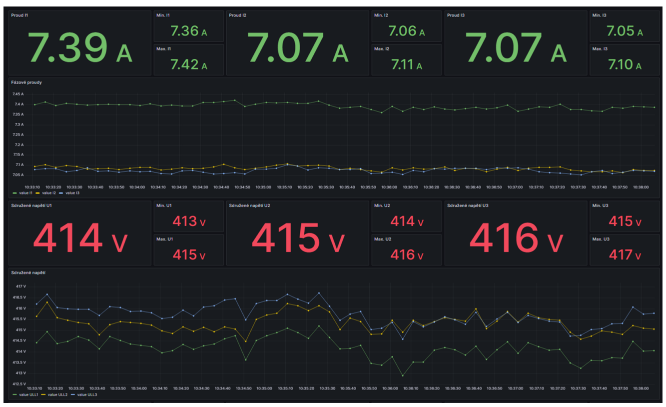



Fig. 5. Data visualization in Grafana.

## IV. Conclusion

In the context of the emerging initiative of active energy management at the BUT and based on the current state of metering of electricity consumption and other parameters and on the basis of the energy audit that was implemented in 2022, the BUT decided to deploy sub-metering of electricity in selected nodes in order to better monitor the energy flows within the electrical network in the buildings.

In particular, there is a need to monitor the consumption of technologies necessary for the proper functioning of the building and for maintaining the quality and comfort of the indoor environment, IT technologies that operate in continuous operation and the operational consumption of individual floors/departments. This system will also be integrated via Modbus RTU into the university-wide BMS system that controls other building technologies. Data from both systems will be used in parallel for operational purposes. This measurement can lead to motivation to implement energy saving measures to reduce electricity consumption.

In practice, visualization is achieved using software such as Home Assistant, which is a free platform designed primarily for home automation and serves as an integration platform independent of the IoT ecosystem. A better solution than Home Assistant for our purposes might have been Thingsboard, which is an open-source platform focused on data management, visualization and collection. It provides an extensive set of tools for monitoring and controlling devices, gives users the ability to create advanced dashboards, and makes it easy to work with data from different devices. The reason why this system was not selected was its primary focus on IoT. Because of Grafana's greater flexibility for different types of projects and data sources, and especially because of its simple and user-friendly environment, Grafana was chosen.

Currently, meters have been installed at 60 out of 115 outlets in BUT buildings T8, T10, and T12. In fig. 6 already installed meters can be seen. The plan is to meter up to 95 % of the outlets. The database and the visualizations are currently working but changes will be made.The installation of the meters is currently underway. All will be operational during the conference EEICT.

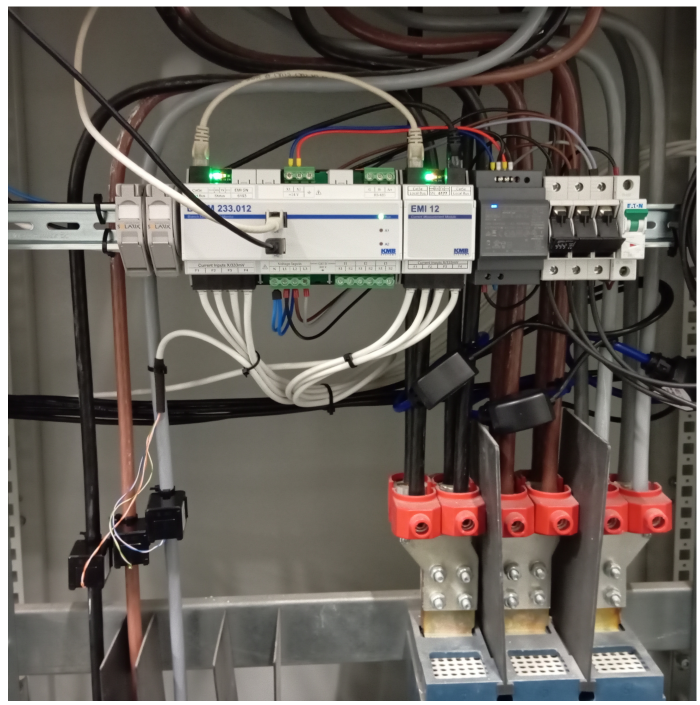



Fig. 6. Installed meters in building T12.

## References

[1] Office for Technical Standardization, metrology and state testing. ČSN EN ISO 50001, *Energy management systems – Requirements with guidance for use.* 1st eddition. Prague, 2019.

[2] KMB SYSTEMS, S.R.O. *Analyzátor sítě a monitor podružných vývodů BCPM 233.012.* [online]. Published 12.4.2021. [cit. 2024-03-01]. Available from: https://kmb.cz/cs/ke-stazeni/dokumentace/24-manualy-pro-merici-pristroje?download=159:bcpm233-012-manual-v3-0-cz

[3] Grafana Labs *Grafana open source documentation.* [online]. 2024. [cit. 2024-03-2]. Available from: https://grafana.com/docs/grafana/

[4] Node-Red *Node-Red documentation.* [online]. 2024. [cit. 2024-03-2]. Available from: https://nodered.org/docs/

[5] InfluxDB *InfluxDB documentation.* [online]. 2024. [cit. 2024-03-2]. Available from: https://docs.influxdata.com/

# Training set generation system for reconstruction of electrical impedance tomography images

Samia Zemiti
*Institut supérieur d'électronique de Paris*
Paris, France
samia.zemiti@eleve.isep.fr

Clark Aloph
*Institut supérieur d'électronique de Paris*
Paris, France
clark.aloph@eleve.isep.fr

Jan Mikulka
*Brno University of Technology*
*Technicka 12, 616 00*
Brno, Czech Republic
mikulka @vut.cz

***Abstract*— This paper introduces an innovative approach to simulating Electrical Impedance Tomography (EIT) through MATLAB, aimed at advancing the accuracy and reliability of internal imaging using electrodes. We address the critical challenge of reconstructing interior images with high fidelity by simulating inhomogeneous mediums. Our methodology involves the generation of synthetic datasets, encompassing various inhomogeneity scenarios, followed by applying forward solutions to ascertain voltage measurements indicative of interior conductivity variations. The research emphasizes the creation of an adaptive framework capable of simulating real-world scenarios within a controlled digital environment, thereby enhancing the predictive capabilities of EIT systems in diverse applications ranging from medical imaging to industrial inspection.**

***Keywords*— *EIT, image reconstruction, dataset generation, forward problem, conductivity distribution, inhomogeneity mapping, MATLAB.***

## I. Introduction

Electrical Impedance Tomography (EIT) represents a revolutionary step in non-invasive imaging technologies, offering unique advantages over conventional methods such as X-ray, MRI, and CT scans. Unlike these traditional techniques, which often come with high operational costs, require stationary equipment, and necessitate specialized personnel for operation, EIT provides a cost-effective, portable, and user-friendly alternative. This innovative approach leverages the measurement of electrical conductivity and permittivity within a given domain, facilitated by an array of electrodes placed around the subject of interest. By applying a small current through these electrodes and measuring the resulting potentials, EIT reconstructs the conductivity distribution inside the object, offering insights into its internal structure.

The term "tomography" comes from the Greek word "tomos," meaning "slice" or "section," and refers to imaging by sections or sectioning. In the context of Electrical Impedance Tomography (EIT), it involves creating a detailed image of the cross-section of an object by interpreting the electrical signals measured on the object's surface. These signals are influenced by the internal structure of the object, and by analyzing them, we can infer what's inside without physically cutting into the object. This technique is useful for non-invasive examinations, which is particularly valuable in medical diagnostics and material testing [2].

The principle of EIT is grounded in the relationship between the electrical properties of materials and their internal composition. This relationship is quantified through the calculation of voltage differences and the subsequent generation of conductivity matrices, visualized as two-dimensional or three-dimensional images. "The voltages are measured using voltage measurement electrodes. An image reconstruction method is then used to calculate the internal conductivity distribution from the boundary data." [3]. The technique's versatility allows for applications ranging from medical diagnostics, where it can offer non-invasive views of internal organ function, to industrial process monitoring, where it can detect changes within containers or pipelines without direct access [4]. Despite its potential, the adoption of EIT faces challenges, primarily due to the complexity of accurately reconstructing images from the measured boundary voltages.

This paper addresses these challenges by introducing a digital simulation framework designed to enhance the accuracy and reliability of EIT. Through a comprehensive approach that includes the generation of synthetic datasets, the application of advanced image reconstruction algorithms, and the integration of machine learning techniques, we aim to push the boundaries of what EIT can achieve. By providing a detailed exploration of our simulation environment, complete with a variety of inhomogeneous models, this study contributes to the theoretical understanding of EIT and lays the groundwork for its broader application in both clinical and industrial settings.

## II. Background and Related Work

The advent of Electrical Impedance Tomography (EIT) marks a pivotal advancement in non-invasive diagnostic imaging. This technique, originally pioneered for geophysical exploration, has been adapted to healthcare, enabling the monitoring of pulmonary functions, breast cancer screening, and even brain activity assessment [3].

EIT stands out by employing an indirect approach to visualizing internal conductivity variations through surface electrical measurements. Central to EIT's methodology is the challenge of the inverse problem which is reconstructing an object's internal conductivity map from these external measurements [8]. The problem is notoriously ill-posed; small errors in data can lead to significant reconstruction inaccuracies, necessitating sophisticated computational techniques to resolve meaningful images from noisy data.

A MATLAB-based simulation framework was crafted, generating synthetic datasets under both uniform and variable conductivity scenarios within a model environment. It can be used to code an artificial neural network that would reconstruct the images instantly from the electrodes' measures.

For further context on the significance and historical development of EIT, Brown's comprehensive review provides an extensive overview, while Bera et al.'s work on a MATLAB-based virtual phantom demonstrates the practical applications of such simulations in a controlled environment [4][7]. Soleimani and Lionheart, as well as Mosquera et al., discuss the nuances of solving the forward problem in EIT, shedding light on the computational challenges faced in the field [1][6].

As EIT continues to evolve, the collective knowledge from these and other scholarly works will be indispensable in guiding future innovations. The data compiled in this project could serve as a foundation for machine learning applications, potentially streamlining the reconstruction process to meet the demands of real-time imaging.

## III. Methodology

The development of this simulation framework for Electrical Impedance Tomography (EIT) encompasses several key steps, following a structured approach to accurately simulate and reconstruct EIT images. This framework is designed to operate within a MATLAB environment, facilitating the generation of synthetic EIT data sets for inhomogeneous scenarios and employing established reconstruction techniques to visualize the internal conductivity distributions. The process is delineated into distinct phases as you can see in Fig. 1, each contributing to the framework's overall functionality and performance.

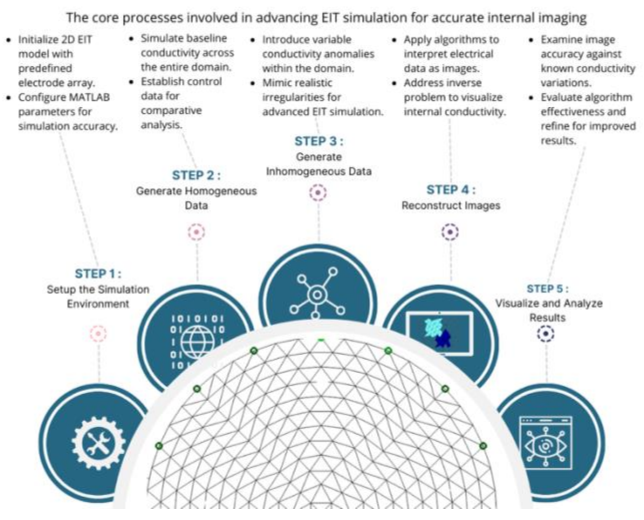



Fig. 1. Five-Step Workflow for EIT Data Simulation and Reconstruction

### A. Environment Setup and Model Initialization

The foundation of the simulation framework is established through the creation of a 2D circular domain, reflecting common geometric configurations in EIT applications. This initial step involves setting up a simulated environment within MATLAB that mimics the physical setup of EIT, including the placement of electrodes around the perimeter of the circular domain. The model's design is pivotal in ensuring that subsequent simulations accurately reflect the potential distributions and interactions expected in real-world EIT scenarios.

At the core of our methodology is the simulation of electrical impedance data. This phase is bifurcated into two primary processes: generating data for a homogeneous medium, which serves as a control or baseline, and generating data for inhomogeneous media, which introduces various anomalies or irregularities into the model. The latter is particularly crucial for evaluating the efficacy of EIT in detecting and characterizing differences in internal conductivity, akin to distinguishing between healthy and pathological tissues or identifying defects in non-biological materials.

### B. Forward Problem and Inhomogeneity Parameterization in EIT Simulations

In this phase of our study, we have addressed the forward problem of EIT, which focuses on predicting voltage measurements given specific conductivity distributions. Our MATLAB-based environment has been utilized to generate these simulations, particularly emphasizing the response to inhomogeneous conditions. Randomly shaped anomalies with varying conductivities were introduced into the model, and the fwd_solve function computed the resulting electrical potential distribution. Our framework provides flexibility by allowing users to customize the characteristics of inhomogeneities. This includes setting the number of anomalies, selecting shapes, and defining a range for their conductivities. Specifically, users can determine the impedance of these anomalies and manipulate their sizes, which can vary between one-half and one-tenth of the mesh size. These parameters are essential for tailoring the simulations to mimic various real-life scenarios.

TABLE 1. Customizable Simulation Parameters and Applied Values

| Parameter | Description | Values-chosen |
|---|---|---|
| Number of anomalies | Total of inhomogeneities within the model | Random integer from 0 to 3 |
| Shapes | Geometrical shape of the anomalies | Circle, Rectangle, or Triangle (randomly selected) |
| Conductivity Range | The range of possible conductivities for anomalies | Random value between 0 (exclusive) and 1 (inclusive) |
| Size of Anomalies | Relative size of each anomaly | Between 10% (0.1) and 50% (0.5) of the mesh size |

The parameters detailed in Table 1 are critical to the versatility of our simulation framework. The customizable aspects of the zone of inhomogeneities' number, shape, conductivity, and size afford us the capability to simulate nuanced environments that closely resemble the complexities of real-world scenarios. These simulations serve as a tool for rigorous testing.

### C. Technical Specifications of the Simulation Framework

Our simulation framework is built within the MATLAB computational environment, designed to model the Electrical Impedance Tomography (EIT) process. The central component is a 2D circular model mimicking the cross-section of the target medium, fitted with 16 virtual electrodes around its boundary as you can see in Fig. 2. These electrodes are simulated to have an ideal contact impedance, and their placement follows the equipotential lines for a uniform electrical field. The forward model, underpinning the simulations, utilizes a finite element method (FEM) approach to solve the associated boundary value problem, ensuring the accurate prediction of potential distributions given any conductivity scenario within the model.

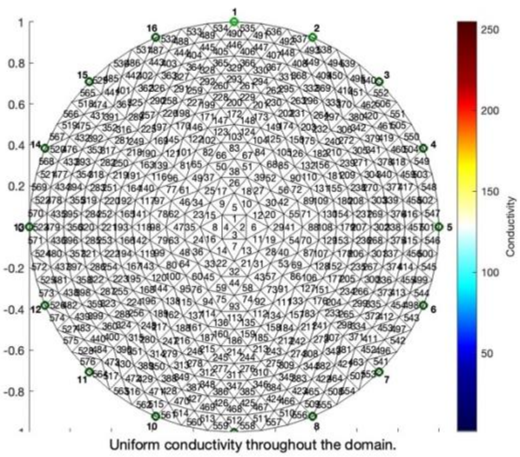



Fig. 2. EIT Mesh Configuration: Node Indexing on Homogeneous Ground

## D. Generation of Synthetic Datasets

Synthetic datasets are generated to reflect two distinct scenarios: a homogeneous medium with uniform conductivity and an inhomogeneous medium with varied conductivities and randomly placed anomalies. The homogeneous dataset serves as a baseline to calibrate the system and validate the reconstruction algorithm's ability to recover a known conductivity distribution. In contrast, the inhomogeneous datasets are created to simulate more complex and realistic scenarios that are likely to be encountered in practical applications. Random shapes such as circles, rectangles, and triangles, with random conductivities different from the background, are introduced to test the algorithms' sensitivity and specificity in detecting and imaging internal structures (Fig. 3.).

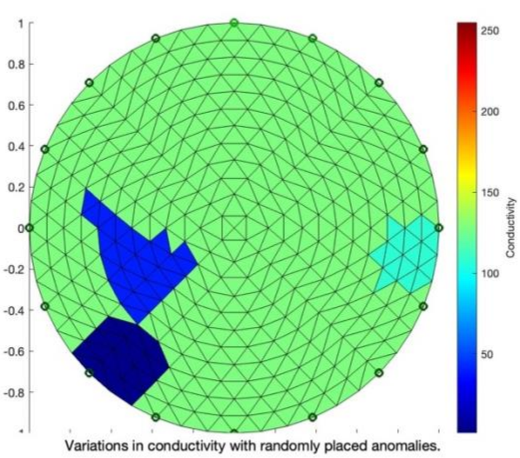



Fig. 3. Inhomogeneous Conductivity with Anomaly Mapping

## E. Image Reconstruction Algorithms Implemented

The framework implements a range of image reconstruction algorithms, from linear back-projection methods to iterative techniques such as the Levenberg-Marquardt algorithm [9]. These algorithms are chosen for their relevance in current EIT research and their proven track record in medical imaging applications. The selection allows for a comparative analysis of different approaches, providing insights into their effectiveness in various simulation conditions.

# IV. Results

In this section, we present the results derived from the simulations of Electrical Impedance Tomography (EIT) using synthetic datasets. These datasets were designed to assess the effectiveness of EIT reconstruction algorithms under controlled conditions, simulating both uniform and varied conductivity environments.

## A. Benchmarking with Homogeneous Models

The homogeneous simulations established a benchmark for the accuracy of our reconstruction algorithms. With a consistent conductivity set across the entire domain, the algorithms were tested for their ability to render an accurate representation of a well-defined and predictable environment. The results showed a high degree of precision, with reconstructed images matching the known parameters closely, thereby confirming the system's calibration accuracy and the reliability of the base algorithms.

## B. Analyzing Complex Conductivity Profiles

In our inhomogeneous simulation phase, we wanted the domain with random conductivity anomalies, to take the form of geometric shapes placed at random locations. These shapes were introduced to assess the robustness of our reconstruction algorithms under conditions mimicking real-world complexities. Figures 4, 5, and 6 illustrate these random placements with triangle and rectangle combinations, a single circle, and a composite of different shapes, respectively. The algorithms demonstrated proficiency in segregating these anomalies from the homogeneous backdrop, as can be observed. These results underscore the promise of the employed EIT techniques in navigating complex diagnostic scenarios.

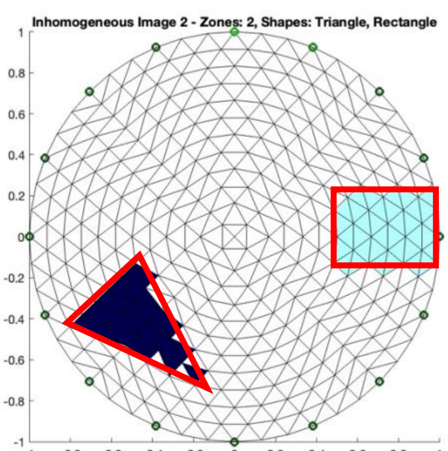



Fig. 4. EIT Reconstruction of triangular and rectangular objects

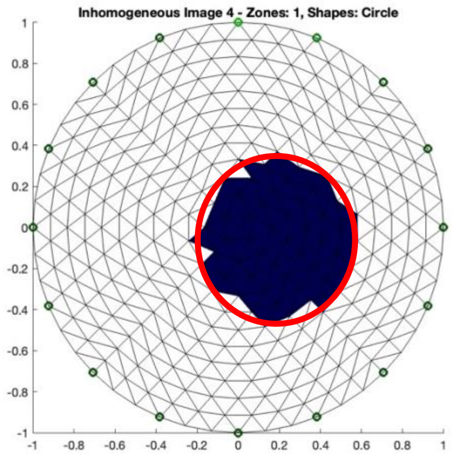



Fig. 5. EIT Reconstruction circular object

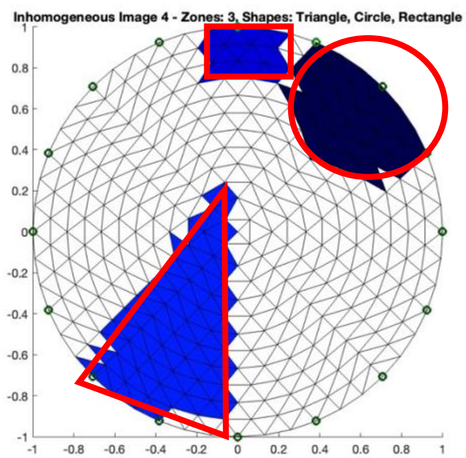



Fig. 6. EIT Reconstruction of triangular and rectangular and circular objects

### C. *Current advances in EIT: Enhanced Imaging techniques*

While our project concentrated on addressing the forward problem in EIT by generating synthetic images from voltage measurements, the groundwork has been laid for future comparative analysis. Modern techniques showcased a marked improvement, particularly in refining image resolution and expediting processing time, a pivotal advancement for applications necessitating swift and precise imaging. Notably, recent research by Smith J. (2023) corroborates these findings, highlighting the efficiency of modern algorithms in dynamic imaging scenarios [10]. Furthermore, Bawa's (2024) comprehensive review in his book "Advances in Medical Imaging, Detection, and Diagnosis" [11] provides an extensive discussion on the evolution of EIT techniques, emphasizing the substantial enhancements in diagnostic capabilities brought about by recent innovations. These works substantiate the significant strides made in the field, illustrating the progress from traditional methods to cutting-edge solutions.

## V. Conclusion

In this research, we've developed an advanced simulation framework for Electrical Impedance Tomography, adept at discerning internal structures through the manipulation of Volt measurements. The resulting datasets represent a foundational step toward enhancing EIT with machine learning efficiencies.

One of the study's key contributions is the preparation of these datasets for neural network integration, which holds the promise of transforming voltage measurements into precise imagery with remarkable speed. This effort is a stride toward real-time, non-invasive imaging techniques that could have significant impacts on medical diagnostics and material analysis. Looking ahead, the potential to train neural networks to quickly interpret EIT data could revolutionize imaging by reducing reliance on time-intensive reconstruction methods.

## Acknowledgment

This work was supported by the Department of Theoretical and Experimental Electrical Engineering, Brno University of Technology.

## References


[1] G Soleimani, M., & Lionheart, W. R. (2005). Nonlinear image reconstruction for electrical capacitance tomography using experimental data. Measurement Science and Technology, 16(10), 1987.

[2] Wahab, Y. A., Rahim, R. A., Rahiman, M. H. F., Rahim, H. A., Aw, S. R., Jamaludin, J., & Fadzil, N. S. M. (2014). A review of process tomography application in inspection system. Jurnal Teknologi, 3(3), 35-39.

[3] Harikumar, R., Prabu, R., & Raghavan, S. (2013). Electrical impedance tomography (EIT) and its medical applications: a review. Int. J. Soft Comput. Eng, 3(4), 193-198.

[4] Brown, B. H. (2003). Electrical impedance tomography (EIT): a review. Journal of medical engineering & technology, 27(3), 97-108.

[5] Bera, T. K. (2018, March). Applications of electrical impedance tomography (EIT): a short review. In IOP Conference Series: Materials Science and Engineering (Vol. 331, No. 1, p. 012004). IOP Publishing.

[6] Mosquera, V., Gonzalez, C., & Ortega, E. (2019). Eidors-Matlab interface for forward problem solving of electrical impedance tomography.

[7] Bera, T. K., Maity, P., Haldar, S., & Nagaraju, J. (2014). A MatLAB based virtual phantom for 2D electrical impedance tomography (MatVP2DEIT): studying the medical electrical impedance tomography reconstruction in computer. J Med Imaging Health Inform, 4(2), 147-167.

[8] Dimas, C., & Sotiriadis, P. P. (2018, May). Electrical impedance tomography image reconstruction for adjacent and opposite strategy using FEMM and EIDORS simulation models. In 2018 7th International Conference on Modern Circuits and Systems Technologies (MOCAST) (pp. 1-4). IEEE.

[9] Klima, O., Kleparnik, P., Spanel, M., & Zemcik, P. (2016, March). Intensity-based femoral atlas 2D/3D registration using Levenberg-Marquardt optimisation. In Medical Imaging 2016: Biomedical Applications in Molecular, Structural, and Functional Imaging (Vol. 9788, pp. 113-124). SPIE.

[10] Smith, J. (2023). Novel Hybrid-Learning Algorithms for Improved Millimeter-Wave Imaging Systems. arXiv preprint arXiv:2306.15341.

[11] Raj Bawa, Gerald F. Audette, S. R. Bawa, Bela Patel, Bruce D. Johnson, Rajeev Khanna (2024). Advances in Medical Imaging, Detection, and Diagnosis.

# 3D-bioprinted Gelatin/Alginate loaded with Carbon Nanotubes for tissue engineering application


1st Minoo Partovi Nasr
*Department of Biomedical Engineering*
*Faculty of Electrical Engineering and Communication*
*Brno University of Technology*
Brno, Czech Republic
253556@vutbr.cz

2nd Inna Zumberg
*Department of Biomedical Engineering*
*Faculty of Electrical Engineering and Communication*
*Brno University of Technology*
Brno, Czech Republic
zumberg@vutbr.cz

3rd Larisa Chmelikova
*Department of Biomedical Engineering*
*Faculty of Electrical Engineering and Communication*
*Brno University of Technology*
Brno, Czech Republic
chmelikoval@vutbr.cz

4th Zdenka Fohlerova
*Department of Biomedical Engineering*
*Faculty of Electrical Engineering and Communication*
*Brno University of Technology*
Brno, Czech Republic
Zdenka.Fohlerova@ceitec.vutbr.cz

5th Valentine Provaznik
*Department of Biomedical Engineering*
*Faculty of Electrical Engineering and Communication*
*Brno University of Technology*
Brno, Czech Republic
provaznik@vut.cz



***Abstract*—The objective of utilizing 3D-bioprinted Gelatin/Alginate loaded with Carbon Nanotubes (CNTs) in tissue engineering applications is to create scaffolds that closely mimic the natural extracellular matrix (ECM), thereby enhancing cell growth, proliferation, and differentiation. Gelatin and Alginate, both biocompatible materials, have been widely researched for their potential in bioprinting due to their similarity to the ECM, offering a conducive environment for cell encapsulation and tissue regeneration. The addition of CNTs to these hydrogels significantly improves the mechanical properties and stability of the scaffolds, making them more suitable for supporting tissue development. CNTs, known for their unique properties such as high tensile strength and electrical conductivity, contribute to the development of scaffolds that not only support mechanical stability but also can influence cellular behavior and tissue formation. This integration aims at enhancing the functionality of 3D-bioprinted scaffolds, enabling them to better support the formation and maturation of engineered tissues. Furthermore, the electrical conductivity of CNTs-loaded scaffolds can be exploited to stimulate electrical activity in tissues, such as cardiac and neural tissues, promoting organized tissue development and functionality. The strategic combination of Gelatin/Alginate with CNTs in 3D bioprinting offers a promising approach to tissue engineering, aiming to address the critical challenge of replicating the complex structure and function of natural tissues. This innovative methodology not only enhances the mechanical and structural properties of the scaffolds but also introduces new possibilities in tissue engineering through the electrical stimulation of tissues, paving the way for the creation of more complex and functional tissue constructs.**



***Keywords—3D-bioprinting, Gelatin, Alginate, Carbon Nanotubes, Tissue engineering.***


## I. Introduction

3D bioprinting in tissue engineering represents a transformative advancement in regenerative medicine, offering unprecedented capabilities in the creation of complex tissue constructs that closely mimic natural tissues. Leveraging the precision of 3D printing technologies, this approach enables the fabrication of tissue models with specific geometric and structural designs, utilizing a combination of biomaterials and living cells. This technology has found applications across a wide range of medical fields, including the development of bone, orthopedic tissue, and even entire organs, providing new avenues for research and treatment [1]. One of the key benefits of 3D bioprinting is its ability to replicate the desired biological, structural, and mechanical properties of the target tissue through careful selection of biomaterials and printing techniques and it allows for the spatial distribution of cells within the constructs, enhancing their functionality and regenerative potential [2]. Despite challenges such as the need for improved printing resolution and a wider array of compatible materials, the ongoing advancements in 3D bioprinting continue to push the boundaries of what is possible in tissue engineering and hold great promise for the future of regenerative medicine [3].

In the field of 3D bioprinting for tissue engineering, Gelatin and Alginate stand out as two pivotal biomaterials that have garnered significant attention for their unique properties and applications. Gelatin's significance in tissue engineering is underscored by its ability to closely mimic the properties of native tissues, thanks to its derivation from collagen. Its versatility and functionalization capabilities make it a crucial component in bioink formulations, facilitating the creation of gelatin-based hydrogels that enhance the regenerative potential of engineered tissues [4]. Alginate, renowned for its biocompatibility, biodegradability, and rapid gelling properties, is another biomaterial frequently employed in 3D bioprinting. Its utility in creating scaffolds for tissue regeneration is well-documented, with alginate-based hydrogels being used extensively due to their non-inflammatory reactions, ease of gel formation, and compatibility with various cell types [5]. The synergy between Gelatin and Alginate in 3D bioprinting has led to the development of composite bioinks tailored for specific tissue engineering applications. By manipulating the concentrations of Alginate and Gelatin, researchers have been able to fabricate scaffolds with optimized mechanical properties and structural fidelity, suitable for a range of tissue regeneration purposes. This combination has proven effective in enhancing the printability and structural integrity of bioinks, paving the way for the engineering of complex tissue constructs with improved regenerative outcomes [6].

The integration of CNTs with Gelatin/Alginate (GA) hydrogel represents a significant advancement in the field of 3D bioprinting for tissue engineering. This combination harnesses the unique properties of each component to create composite bioinks that offer enhanced mechanical strength, improved printability, and novel physical characteristics conducive to tissue regeneration [7]. Beyond mechanical enhancements, CNTs can endow the hydrogel with unique physical features such as electrical conductivity, shape memory and tuning rheological properties. The synergistic combination of Gelatin, Alginate, and CNTs opens up a wide range of biomedical applications [7]. Despite these advantages, challenges such as ensuring biocompatibility to avoid toxicity, optimizing printability and shape fidelity, and maintaining structural stability of the hydrogel-CNTs composite need to be carefully addressed. These challenges highlight the importance of meticulous formulation and testing of the bioink to ensure successful outcomes in tissue engineering applications [8].

This study investigated the bioink preparation of 3D-biopritned GA loaded with CNTs for tissue engineering applications. Initially, the bioink composition were fabricated and printed using Cellink INKREDIBLE+ bioprinter with an examination of the printability of the bioink, ensuring that the scaffolds could be produced with high fidelity to the intended designs. Furthermore, the cell viability within these scaffolds was rigorously tested, verifying that the cellular components remained functional and could potentially contribute to tissue growth and healing.

## II. MATERIALS AND METHODS

### A. Bioink preparation

The GA hydrogel ink solutions were prepared using a combination of methods. The amount of Gelatin and Alginate tested in different percentages to have proper printability. Amongst all the percentages of Gelatin and Alginate, 7% Gelatin and 5% of Alginate had proper printability properties (see Table 1). Initially, Gelatin was dissolved in water for 1.5 h at 60 ℃ through constantly stirring at high speed. Afterwards, Alginate was added to the Gelatin solution while mixing at low speed for another 2 h to obtain a pure GA hydrogel ink solution.

TABLE I. SAMPLE COMPOSITION FOR PRINTIBILITY

| Sample name | Gelatin | Alginate | Printability |
|---|---|---|---|
| Gel10-Alg5 | 10% | 5% | No |
| Gel10-Alg10 | 10% | 10% | No |
| Gel7-Alg2.5 | 7% | 2.50% | No |
| Gel7-Alg5 | 7% | 5% | Yes |

Multi-walled CNTs were purchased from Sigma-Aldrich (694185). Later, different concentrations of CNTs (0.25, 0.5, 0.75% w/v) were added to the GA hydrogel ink solution which was stirred for 1 h. Then, the solution was dispersed with probe ultrasonication for 15 min to obtain a GA-CNTs hydrogel ink solution.

### B. 3D-bioprinting and printability

In the current investigation, the micro extrusion-based dual-syringe bioprinter Cellink INKREDIBLE+ (Cellink, Sweden) was used to print the prepared bioink. Stainless steel 22G nozzles were obtained from Cellink with a 13 mm length, inner diameter of 0.41 mm and outer diameter of 0.72 mm. The ink solutions were loaded in a 3 ml plastic cartridge and the nozzle was fixed to the plastic cartridge for printing. The circular scaffold model with a diameter of 15 mm and a height of 0.41 mm was designed using Autodesk Fusion 360 and saved in STL format. Then the model was converted into a G-code file using Cellink HeartWare software with a 0.41 mm layer height (one layer) and 0.27 mm pore size. The printing was performed at room temperature and the pressure ranged between 110 to 150 kPa was selected to extrude ink solutions during the printing of the scaffolds. The nozzle speed was set at 4 mm/s and infill density was 75%. The distance from the nozzle to the print surface was adjusted so that the leading edge of the flow was in line with the nozzle. All the scaffolds were maintained at neutral pH and crosslinked with 100 mM $CaCl_2$ for 5 min immediately after printing followed by washing with PBS for twice an hour.

### C. Cell culture and cell seeding

Mouse fibroblast-like cell line, NIH/3T3 is cultured in High-Glucose Dulbecco's Modified Eagle Medium (HG-DMEM) with 10% fetal bovine serum and 2% penicillin-streptomycin in 37 °C, 5% $CO_2$ incubator with 95% humidity. The cells were regularly subcultured when they reach a certain level of confluency, using a solution of 0.25% trypsin-EDTA solution at a ratio of 1:5.

For cell seeding, the printed scaffolds were exposed to UV irradiation (250 nm) for 20 min within a biological safety cabinet at room temperature. The scaffolds were put into 9.2 $cm^2$ culture dishes (93040, TPP, Switzerland) and suspension of $1{\cdot}10^6$ cells was injected into the scaffolds and incubated at 37 °C, 5% $CO_2$, and 95% humidity. After 1 h of incubation, 1 mL HG- DMEM was added for further culture. Cell seeding was used for subsequent testing of cell attachment and cell morphology analysis.

## III. RESULTS

### A. 3D-bioprinting of GA-CNTs hydrogel

GA hydrogel used as the main component for reaching the best concentration for 3D-biopritning. In the preliminary study, various hydrogels with different concentrations of Gelatin (10, 7%) and Alginate (2.5, 5, 10%) were prepared and printed at room temperature using the INKREDIBLE+ bioprinter. Ink at highest concentration of Gelatin (10%) with Alginate (5%, 10%) had high level of viscosity and it was not printable due to the improper structure of the scaffold. Adding less percentages of Gelatin (7%) with 5% of Alginate showed better viscosity and well-defined structure for printing with high structure fidelity. Following having the best percentage of GA (7%:5%) for printing different concentration of CNTs (0.25, 0.5, 0.75% w/v) were added for printing. The extrusion pressures for printing sample named Gel7-Alg5 with percentages of GA (7%:5%) at room temperature were approximately 120 kPa with nozzle temperature of 27 ℃ (Fig. 1).

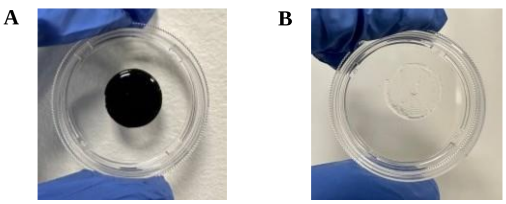



Fig. 1. (A) 3D-bioprinted Gel7-Alg5 consist of 0.5% CNTs; (B) 3D-bioprinted Gel7-Alg5

### B. Cell viability

Gel7-Alg5 printed sample and GA samples consist of different concentration of CNTs (0.25, 0.5, 0.75% w/v) were tested for cell viability. Gel7-Alg5 printed sample demonstrated suitable cell viability after bioprinting within 24 and 48 hours (Fig. 2). The images demonstrate that the scaffold can support growth of NIH/3T3 fibroblast cells. Following the experiments, incorporation of CNTs to Gel7-Alg5 samples were assayed. Amongst three different concentrations of CNTs (0.25, 0.5, 0.75% w/v), sample consist of 0.5% of CNTs had the highest cell survival of NIH/3T3 fibroblast cells within 48 hours. Cell imaging of Gel7-Alg5 samples with CNTs was provided using Leica TCS SP8 X confocal microscope and CellTracker CMFDA dye (Fig. 3).

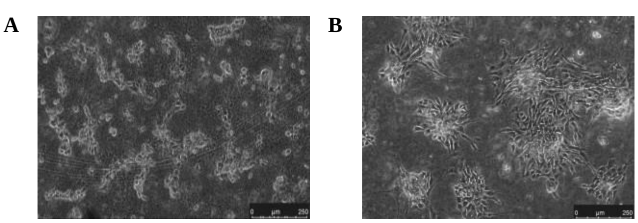



Fig. 2. Testing Biocompatibility of GA hydrogel on NIH/3T3 fibroblast cell line using Leica DMi8 microscope: (A) after 24 hours; (B) after 48 hours

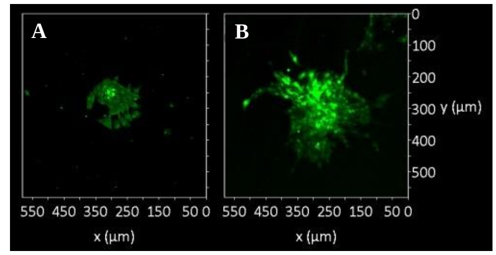



Fig. 3. Testing Biocompatibility of GA-CNTs hydrogel on NIH/3T3 fibroblast cell line with confocal microscope using CellTracker CMFDA dye: (A) after 24 hours; (B) after 48 hours

## IV. CONCLUSION

This study 3D-bioprintability of GA loaded with CNTs were considered for tissue engineering applications. The construct was successfully created by 3D-bioprinting, and it was confirmed that the 3D-bioprinted construct composed of GA loaded with CNTs could maintain viability and promote adhesion of fibroblast cells. These preliminary results show that the 3D-bioprinted GA model loaded with CNTs provides a suitable microenvironment for fibroblast to maintain its activity. This bioink could be a promising candidate for further tissue engineering application in the future.


## ACKNOWLEDGMENT

The completion of this work was made possible by the grant No. FEKT-S-23-8196, financially supported by the Internal science fund of Brno University of Technology.


## REFERENCES


[1] Z. Xie, M. Gao, A. O. Lobo, and T. J. Webster, "3D Bioprinting in Tissue Engineering for Medical Applications: The Classic and the Hybrid," *Polymers (Basel)*, vol. 12, no. 8, p. 1717, Jul. 2020, doi: 10.3390/polym12081717.

[2] H. Abu Owida, "Developments and Clinical Applications of Biomimetic Tissue Regeneration using 3D Bioprinting Technique," *Appl Bionics Biomech*, vol. 2022, pp. 1–12, Dec. 2022, doi: 10.1155/2022/2260216.

[3] P. Jain, H. Kathuria, and N. Dubey, "Advances in 3D bioprinting of tissues/organs for regenerative medicine and in-vitro models," *Biomaterials*, vol. 287, p. 121639, Aug. 2022, doi: 10.1016/j.biomaterials.2022.121639.

[4] I. Lukin *et al.*, "Progress in Gelatin as Biomaterial for Tissue Engineering," *Pharmaceutics*, vol. 14, no. 6, p. 1177, May 2022, doi: 10.3390/pharmaceutics14061177.

[5] M. Young, The Technical Writer's Handbook. Mill Valley, CA: University Science, 1989.

[6] S. Y. Sonaye, E. G. Ertugral, C. R. Kothapalli, and P. Sikder, "Extrusion 3D (Bio)Printing of Alginate-Gelatin-Based Composite Scaffolds for Skeletal Muscle Tissue Engineering," *Materials*, vol. 15, no. 22, p. 7945, Nov. 2022, doi: 10.3390/ma15227945.

[7] D. Moura, R. F. Pereira, and I. C. Gonçalves, "Recent advances on bioprinting of hydrogels containing carbon materials," *Mater Today Chem*, vol. 23, p. 100617, Mar. 2022, doi: 10.1016/j.mtchem.2021.100617.

[8] A. B. Kakarla, I. Kong, I. Turek, C. Kong, and H. Irving, "Printable gelatin, alginate and boron nitride nanotubes hydrogel-based ink for 3D bioprinting and tissue engineering applications," *Mater Des*, vol. 213, p. 110362, Jan. 2022, doi: 10.1016/j.matdes.2021.110362.

# Deep Learning for Agar Plate Analysis: Predicting Microbial Cluster Counts

1st Michal Cicatka
*Department of Telecommunications*
*Brno University of Technology*
Brno, Czech Republic
0000-0002-8119-6679

2nd Radim Burget
*Department of Telecommunications*
*Brno University of Technology*
Brno, Czech Republic
0000-0003-1849-5390

***Abstract*—Manual analysis of agar plates remains a bottleneck in microbiology, hindering automation efforts. This study investigates the feasibility of using machine learning for automated microbial cluster count detection from agar plate images. We employed various methods, including elbow detection (baseline) and supervised learning models (Support Vector Regression, Simple CNN, XGBoost, Random Forest, pre-trained VGG, and pre-trained Inceptionv3). The results demonstrate that machine learning models significantly outperform the baseline, achieving lower prediction errors and higher accuracy in identifying the correct number of clusters. Notably, both pre-trained VGG and InceptionV3 achieved strong performance, highlighting the effectiveness of transfer learning for this task. InceptionV3 exhibited the lowest error rates overall. This study establishes a foundation for developing robust automated systems for quantifying microbial growth, potentially streamlining workflows and improving efficiency in microbiological research and clinical settings.**

***Index Terms*—image processing, machine learning, agar plates**

## I. Introduction

The landscape of biomedical laboratories is shifting towards total laboratory automation (TLA), driven by the desire to cut costs, streamline repetitive tasks, and enhance training efficiency [1]. However, despite advancements, a key limitation remains in analyzing agar plates, a vital tool teeming with microbial life. While automation has taken hold, it hasn't fully embraced the full potential of agar plates, especially in smaller laboratories. Tasks like colony picking, counting, and antibiotic susceptibility testing still rely heavily on manual intervention, hindering efficiency and introducing potential human error.

At the heart of these tasks lies the crucial ability to localize microbial colonies in the agar plate image. Existing approaches to localization primarily rely on segmentation, the process of separating the colonies from the background agar. Traditional image processing techniques [2]–[4] alongside machine learning-based methods [5]–[9] have been successfully employed for this purpose.

However, the realm of advanced microbial analysis remains largely untapped. While existing classification approaches have been utilized for colony counting [10] and differentiating virulent and avirulent colonies [11], and even classifying bacterial species [12], they often lack the generalizability needed for broader application. These solutions tend to be tailored to specific experimental settings and may not perform well in different contexts, limiting their overall impact.

Instead of focusing on differentiating specific microbial types or detecting individual phenomena on agar plates, a critical gap remains in the ability to group colonies of the same type together. This functionality would revolutionize workflows across various applications, including automating the detection of contaminated agar plates. Additionally, preselection algorithms in colony picking robots and advanced colony counting analyses would benefit immensely from this ability.

Existing clustering algorithms typically fall into two categories: those requiring a pre-defined number of clusters (e.g., K-means [13]) and those that predict the number of clusters themselves (e.g., OPTICS [14], DBSCAN [15]). However, the latter category often presents challenges in parameter fine-tuning and post-processing, requiring merging or splitting clusters.

This article presents a novel solution aimed at addressing this gap. We delve into a method for predicting the count of microbes directly from agar plate images, paving the way for a more comprehensive and automated approach to analyzing these ubiquitous laboratory tools.

## II. Methods and Materials

This section outlines our exploration of predicting microbial cluster counts in agar plate images using regression machine learning models. We conducted experiments utilizing a unique dataset specifically designed for this task, provided by the company Bruker Daltonics GmbH & Co. KG.

The provided dataset consisted of 92 agar plates, each captured under six different lighting conditions, resulting in a total of 552 images. Each plate contained 1 to 4 distinct microbial types, cultivated in designated areas to prevent colony merging and facilitate unambiguous annotation with segmentation masks. These masks assigned unique colors to each pixel, corresponding to a specific cluster of the same microbial type (see Fig. 1). While the controlled environment minimized colony merging and facilitated annotation, it's important to note that not all microbial types were successfully cultivated. This occasionally resulted in missing data points in the segmentation masks, as some intended colonies were

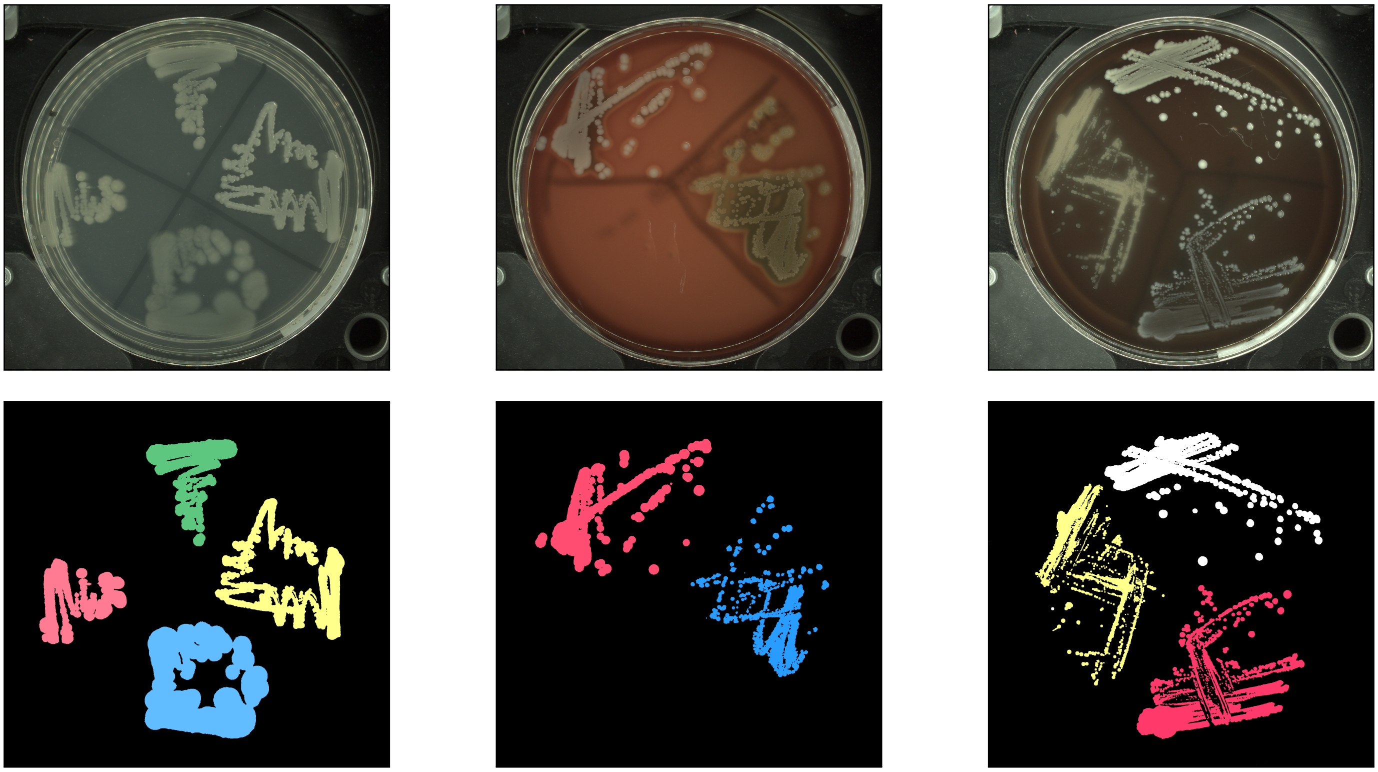


Fig. 1. Examples of dataset images (top) and their corresponding annotation masks (bottom)

absent. This controlled environment and detailed annotation proved instrumental in training and evaluating our proposed machine learning models for predicting microbial cluster counts.

Since each agar plate was captured under six different lighting conditions, we employed a selection strategy to choose the most suitable image for analysis. We opted to select the image with the highest entropy. This choice was made because entropy provides a measure of information content, and images with higher entropy often contain more details and potentially offer richer information for accurate colony identification and count prediction.

As a baseline, we employed the elbow/knee detection algorithm [16], a well-established approach for determining the optimal number of clusters. This method analyzes the "elbow" or "knee" of a curve representing the explained variance within the data as a function of the number of clusters. The optimal count is identified at the "elbow" where the additional gain in explained variance diminishes significantly. While conceptually simple, this baseline provides a benchmark for comparing the performance of more complex models.

Unlike the machine learning models described later, this approach does not require training. Instead, it relies solely on the inherent characteristics of the data. To provide the elbow/knee detection algorithm with informative features, we converted the entire image data into the HSV color space before applying the algorithm. The HSV color space often separates color information from intensity information, potentially aiding the algorithm in identifying distinct clusters based on color variations. This choice aligns with the intuitive notion that different microbial types might be visually distinguishable by their color characteristics. While conceptually simple, this baseline using elbow/knee detection with HSV-converted images provides a benchmark for comparing the performance of more complex models and allows for an initial assessment of the feasibility of the task.

To improve cluster counting accuracy, we explored regression-focused machine learning models (Support Vector Regression, Random Forest, XGBoost) trained on manually extracted image features. These features capture key characteristics of the colonies, allowing the models to learn the relationship between features and actual cluster counts.

We began by converting the segmentation mask to a boolean format to identify foreground pixels corresponding to the colonies. Additionally, the image was converted to the HSV color space, which separates color information from intensity information, potentially aiding feature extraction.

For each color channel (both RGB and HSV), various statistical properties were calculated for the foreground pixels. These included mean, standard deviation, median, minimum, and maximum values, capturing essential information about the overall intensity, dispersion, and central tendency of the pixel values within each colony.

To capture the distribution of pixel intensities within each color channel, histograms were computed for the foreground pixels. These histograms were normalized, ensuring they represent the relative frequency of intensity values within a specific range, and providing insights into the distribution of colors within each colony.

Finally, texture features were extracted from the grayscale version of the image using the Haralick descriptors. These descriptors capture the spatial arrangement of pixel intensities, providing information about the co-occurrence of different intensity values within the colony region.

By combining these diverse feature types (color properties, intensity distribution, and texture), we aimed to create a comprehensive representation of each colony. These features were then used by the machine learning models to learn the relationship between the extracted information and the corresponding cluster count.

To optimize the performance of each model (SVR, Random Forest, and XGBoost), we employed Bayesian optimization. This approach iteratively evaluates the model's performance with different hyperparameter configurations, using the results to guide the selection of the next configuration to be evaluated. This process aims to efficiently identify the hyperparameter combination that yields the best performance for the task, ultimately leading to more accurate predictions.

Conversely, Convolutional Neural Networks (CNNs) automatically learn features directly from raw image data. We employed a simple CNN, pre-trained VGG model, and pre-trained Inceptionv3 model, all trained to predict cluster counts directly from agar plate images. All models used the Adam optimizer with a learning rate of $1e^{-3}$ and mean squared error (MSE) loss function.

The simple CNN architecture had two convolutional layers with ReLU activations and 2x2 max pooling layers, followed by a fully connected layer with 128 neurons and a dropout layer with a 0.5 drop rate. The final output layer had a single neuron with a linear activation for regression.

For the pre-trained models (VGG16 and Inceptionv3), we excluded the top classification layer and froze the weights of the remaining layers to prevent overfitting. We then added a custom head consisting of a flattening layer, a fully connected layer with 256 neurons and a dropout layer with a 0.5 drop rate, followed by a final output layer with a single neuron and a linear activation for regression.

To artificially expand the training dataset and improve generalization, we employed data augmentation techniques such as random rotations, shifts, shears, zooms, and horizontal flips during training. This process helped the models learn robust features that are less sensitive to variations in the image data.

The training data was split into training and validation sets using an 80/20 split. The model was trained on the training set with early stopping and model checkpointing implemented to prevent overfitting and to save the best performing model based on validation loss. Additionally, a learning rate scheduler was used to gradually reduce the learning rate over time, potentially facilitating convergence and fine-tuning.

Overall, the training process aimed to optimize the performance of both models for accurately predicting microbial cluster counts from agar plate images, while incorporating techniques to mitigate overfitting and improve model generalizability.

Finally, all models were evaluated using a combination of metrics to assess their performance: Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and Mean Absolute Error (MAE) provided insights into the magnitude of prediction errors. Additionally, absolute accuracy, calculated as the number of correctly predicted cluster counts (rounded predictions) in the 28-image evaluation set, was determined. We also included the Pearson correlation coefficient to evaluate the linear relationship between the predicted and actual cluster counts.

## III. Results

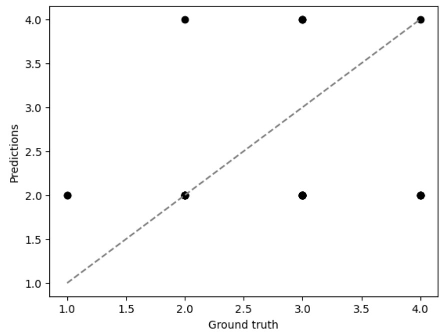
TABLE I: Performance comparison of different methods for microbial cluster counting

| **Architecture** | **MSE** | **RMSE** | **MAE** | **Correlation coefficient** | **Hits (M/N)** |
|---|---|---|---|---|---|
| Elbow detection (inertia) | 1.357 | 1.165 | 0.929 | 0.116 | 8/28 |
| Elbow detection (distortion) | 2.214 | 1.488 | 1.214 | -0.144 | 6/28 |
| Simple CNN | 0.709 | 0.842 | 0.716 | 0.293 | 10/28 |
| Support Vector Regression | 0.702 | 0.838 | 0.711 | 0.347 | 14/28 |
| Random Forest | 0.437 | 0.661 | 0.518 | 0.658 | 17/28 |
| XGBoost | 0.422 | 0.650 | 0.478 | 0.676 | 17/28 |
| Pre-trained VGG | 0.452 | 0.672 | 0.495 | 0.680 | 18/28 |
| **Pre-trained Inceptionv3** | **0.361** | **0.601** | **0.423** | **0.721** | **20/28** |

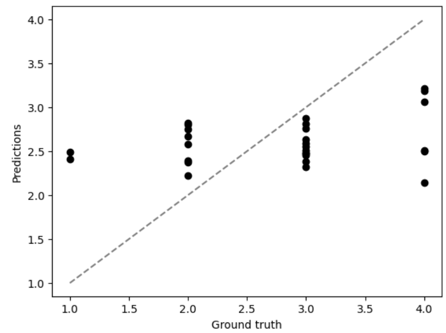
This section presents the evaluation results of the employed methods for predicting microbial cluster counts from agar plate images. The results are presented in Table I. For a more comprehensive understanding of the model performance, we included six figures (Figures 2a, 2b, 2c, 2d, 2e, and 2f) visualizing the predicted versus actual cluster counts for selected models and the elbow detection (inertia) as a baseline. These scatter plots allow for visual inspection of the distribution of errors and provide insights into the models' ability to capture the relationship between image features and cluster counts.

Both elbow detection approaches using inertia and distortion metrics exhibited limitations. While elbow detection with distortion showed the lowest error metrics (MSE = 2.214, RMSE = 1.488), it only correctly predicted cluster counts in 6 out of 28 samples. Elbow detection with inertia achieved slightly lower MSE (1.357) and RMSE (1.165), but its performance was still unsatisfactory. Additionally, the very low correlation coefficients (potentially negative for distortion) reinforce the notion that these baseline approaches struggle to capture the linear relationship between features and actual cluster counts.

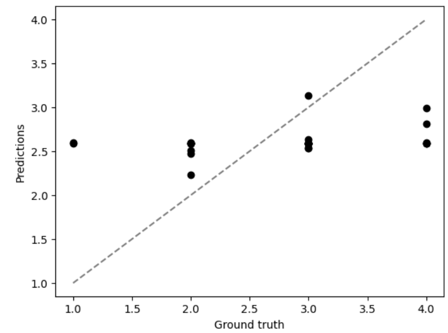
Machine learning models significantly outperformed the baseline elbow detection approach, achieving lower error metrics (MSE, RMSE, MAE) and higher absolute accuracy (Hits)



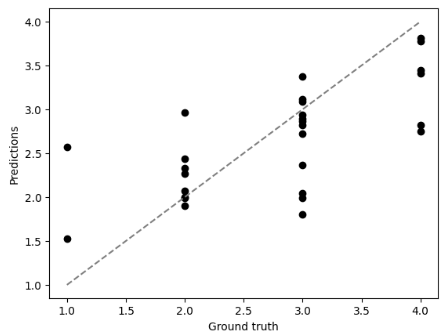

(a) Elbow detection (inertia)



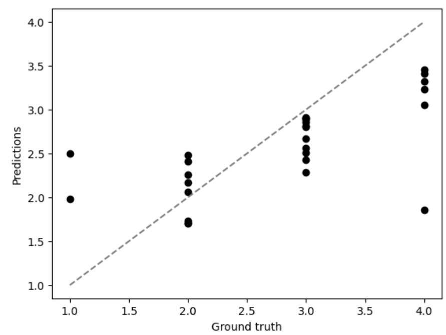

(b) Simple CNN



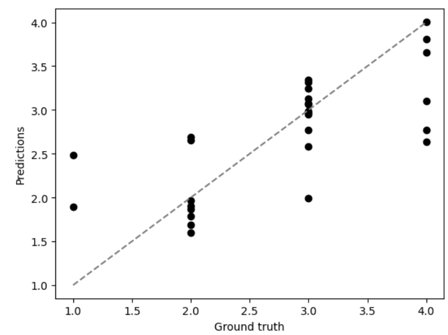

(c) Support Vector Regression (SVR)




(d) XGBoost




(e) Pre-trained VGG




(f) Inceptionv3

Fig. 2. Scatter plots visualizing predicted vs. actual cluster counts for each model and the elbow detection (distortion) as a baseline.

across the board. This suggests improved capability to capture complex relationships within the data.

However, examining corresponding plots (Figures 2b, 2c) reveals limitations in Simple CNN and SVR. These models often returned regressed values aligned with the training data distribution, suggesting a struggle to capture specific relationships within individual samples.

Among the non-pre-trained models, Random Forest achieved the strongest performance, with the lowest MSE (0.437) and a moderately high correlation coefficient of 0.658. This indicates that Random Forest effectively learned the underlying relationships between image features and cluster counts in this specific dataset. Similarly, XGBoost maintained consistent performance with even slightly lower error metrics (MSE = 0.422, RMSE = 0.650) and higher correlation coefficient (0.676).

The highest performance was observed with the pre-trained Inceptionv3 model. Inceptionv3 achieved the lowest error rates overall (MSE = 0.361, RMSE = 0.601, MAE = 0.423) and the highest absolute accuracy (Hits = 20/28), surpassing all other models. The high correlation coefficient (0.721) further strengthens this finding, indicating a strong linear relationship between the predicted and actual cluster counts for Inceptionv3. Notably, the pre-trained VGG model also performed well, achieving a competitive MSE (0.452), RMSE (0.672), and absolute accuracy (Hits = 18/28). While Inceptionv3 showed a slight edge, both models demonstrate the value of leveraging pre-trained architectures for this image analysis task.

It is important to note that even in some cases where the models struggled to achieve perfect accuracy, differentiating between certain microbial colonies can be challenging even for trained professionals, as some visually appear very similar (see Figure 3). This highlights the inherent complexity of the task and emphasizes the relative success of the machine learning models in achieving significant improvements compared to the baseline, despite these limitations.

Overall, the results suggest that machine learning models significantly outperform the baseline elbow detection approaches. This experiment highlights the potential benefits of both exploring different model architectures (like Random Forest and Inceptionv3) and potentially using hyperparameter optimization techniques to further refine the performance of all models for achieving even higher accuracy and generalizability across various datasets and microbial types.

## IV. Conclusion

This study investigated the feasibility of using machine learning to predict microbial cluster counts from agar plate images, aiming to facilitate automated analysis and potentially improve efficiency in microbiological workflows.

We employed various methods, including elbow detection as a baseline and several supervised learning models (Support

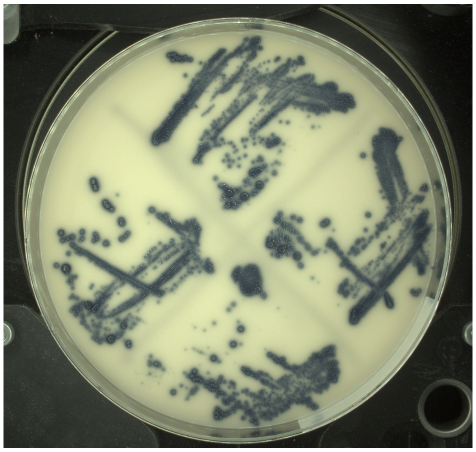


Fig. 3. Example image from the evaluation set showcasing visually similar microbial colonies, where even human experts might find differentiation difficult

Vector Regression, Simple CNN, XGBoost, Random Forest, pre-trained VGG, and pre-trained Inceptionv3). The results convincingly demonstrate that machine learning models significantly outperform the baseline approach. They achieve lower prediction errors and exhibit a higher capability to identify the correct number of clusters in unseen images.

Among the machine learning models, both the pre-trained VGG and InceptionV3 networks achieved strong performance. InceptionV3 exhibited the lowest error rates overall, while VGG remained competitive. This finding h ighlights t he effectiveness of leveraging pre-trained architectures, particularly those specifically designed for image recognition, for this task.

However, further research is warranted to explore additional model architectures, hyperparameter optimization techniques, and the use of larger or more diverse datasets. These efforts aim to further improve accuracy and generalizability of the model.

This study provides a valuable foundation for developing robust and accurate automated systems for quantifying microbial growth from agar plate images. Such systems have the potential to contribute to streamlined workflows a nd efficient analysis in various research and clinical settings. Furthermore, the ability to accurately predict cluster counts paves the way for applying deep clustering algorithms, such as k-means, to classify colonies based on their type. By providing an initial estimate of the number of clusters, this work serves as a crucial first s tep t owards a utomated m icrobial identification and characterization from agar plates.

## ACKNOWLEDGMENT

This article has been made in close cooperation with Bruker Daltonics GmbH & Co. KG. The company provided annotated dataset of agar plate images with microbes and other useful resources and know-how.

## REFERENCES

[1] G. Lippi, G. D. Rin, Advantages and limitations of total laboratory automation: a personal overview, Clinical Chemistry and Laboratory Medicine (CCLM) 57 (6) (2019) 802–811 [cited 2023-09-07]. doi:doi:10.1515/cclm-2018-1323.
URL https://doi.org/10.1515/cclm-2018-1323

[2] Q. Geissmann, Opencfu, a new free and open-source software to count cell colonies and other circular objects, PloS one 8 (2013) e54072. doi:10.1371/journal.pone.0054072.

[3] A. u. M. Khan, A. Torelli, I. Wolf, N. Gretz, Autocellseg: robust automatic colony forming unit (cfu)/cell analysis using adaptive image segmentation and easy-to-use post-editing techniques, Scientific Reports 8 (1) (2018) 7302. doi:10.1038/s41598-018-24916-9.
URL https://doi.org/10.1038/s41598-018-24916-9

[4] D. R. Stirling, M. J. Swain-Bowden, A. M. Lucas, A. E. Carpenter, B. A. Cimini, A. Goodman, Cellprofiler 4: improvements in speed, utility and usability, BMC Bioinformatics 22 (1) (2021) 433. doi:10.1186/s12859-021-04344-9.
URL https://doi.org/10.1186/s12859-021-04344-9

[5] D. Nie, E. A. Shank, V. Jojic, A deep framework for bacterial image segmentation and classification, in: Proceedings of the 6th ACM Conference on Bioinformatics, Computational Biology and Health Informatics, BCB '15, Association for Computing Machinery, New York, NY, USA, 2015, p. 306–314. doi:10.1145/2808719.2808751.
URL https://doi.org/10.1145/2808719.2808751

[6] P. Andreini, S. Bonechi, M. Bianchini, A. Mecocci, F. Scarselli, A deep learning approach to bacterial colony segmentation, in: V. Kůrková, Y. Manolopoulos, B. Hammer, L. Iliadis, I. Maglogiannis (Eds.), Artificial Neural Networks and Machine Learning – ICANN 2018, Springer International Publishing, Cham, 2018, pp. 522–533.

[7] G. Zhu, B. Yan, M. Xing, C. Tian, Automated counting of bacterial colonies on agar plates based on images captured at near-infrared light, Journal of Microbiological Methods 153 (2018) 66–73. doi:https://doi.org/10.1016/j.mimet.2018.09.004.
URL https : / / www . sciencedirect . com / science / article / pii / S0167701218306146

[8] P. Andreini, S. Bonechi, M. Bianchini, A. Mecocci, F. Scarselli, Image generation by gan and style transfer for agar plate image segmentation, Computer Methods and Programs in Biomedicine 184 (2020) 105268. doi:https://doi.org/10.1016/j.cmpb.2019.105268.
URL https : / / www . sciencedirect . com / science / article / pii / S0169260719311216

[9] M. Cicatka, R. Burget, J. Karasek, J. Lancos, Increasing segmentation performance with synthetic agar plate images, Heliyon 10 (3) (2024) e25714. doi:10.1016/j.heliyon.2024.e25714.
URL http://dx.doi.org/10.1016/j.heliyon.2024.e25714

[10] A. Ferrari, S. Lombardi, A. Signoroni, Bacterial colony counting with convolutional neural networks in digital microbiology imaging, Pattern Recognition 61 (2017) 629–640. doi:https://doi.org/10.1016/j.patcog.2016.07.016.
URL https : / / www . sciencedirect . com / science / article / pii / S0031320316301650

[11] T. Naets, M. Huijsmans, P. Smyth, L. Sorber, G. de Lannoy, A mask r-cnn approach to counting bacterial colony forming units in pharmaceutical development (2021). doi:10.48550/ARXIV.2103.05337.
URL https://arxiv.org/abs/2103.05337

[12] S. A. Nagro, M. Kutbi, W. M. Eid, E. J. Alyamani, M. H. Abutarboush, M. A. Altammami, B. K. Sendy, Automatic identification of single bacterial colonies using deep and transfer learning, IEEE Access 10 (2022) 120181–120190.

[13] S. Lloyd, Least squares quantization in pcm, IEEE Transactions on Information Theory 28 (2) (1982) 129–137. doi:10.1109/TIT.1982.1056489.

[14] M. Ankerst, M. M. Breunig, H.-P. Kriegel, J. Sander, Optics: Ordering points to identify the clustering structure, in: Proceedings of the 1999 ACM SIGMOD International Conference on Management of Data, SIGMOD '99, Association for Computing Machinery, New York, NY, USA, 1999, p. 49–60. doi:10.1145/304182.304187.
URL https://doi.org/10.1145/304182.304187

[15] M. Ester, H.-P. Kriegel, J. Sander, X. Xu, et al., A density-based algorithm for discovering clusters in large spatial databases with noise., in: kdd, Vol. 96, 1996, pp. 226–231.

[16] V. Satopaa, J. Albrecht, D. Irwin, B. Raghavan, Finding a "kneedle" in a haystack: Detecting knee points in system behavior, in: 2011 31st International Conference on Distributed Computing Systems Workshops, 2011, pp. 166–171. doi:10.1109/ICDCSW.2011.20.

# Bioinformatics study of the third generation sequencing platforms applied on a thermophile


Mohammad Umair
*Department of Biomedical Engineering*
*Faculty of Electrical Engineering and Communication, Brno University of Technology*
Brno,Czech Republic
Mohd.Umair@vut.cz

Veronika Řeháková
*Department of Food Chemistry and Biotechnology*
*Faculty of Chemistry, Brno University of Technology*
Brno, Czech Republic
Veronika.Rehakova@vut.cz

Iva Buchtikova
*Department of Food Chemistry and Biotechnology*
*Faculty of Chemistry, Brno University of Technology*
Brno, Czech Republic
iva.buchtikova@fch.vut.cz

Matej Bezdicek
*Department of Internal Medicine Hematology and Oncology, University Hospital*
Brno, Czech Republic
Bezdicek.Matej@fnbrno.cz

Stanislav Obruca
*Department of Food Chemistry and Biotechnology*
*Faculty of Chemistry, Brno University of Technology*
Brno, Czech Republic
obruca@fch.vut.cz

Karel Sedlář
*Department of Biomedical Engineering*
*FEEC,Brno University of Technology*
Brno,Czech Republic
sedlar@vutbr.cz



***Abstract*—This study compares the efficiency of Pacific Biosciences technology (PacBio) and Oxford Nanopore Technology (ONT) in sequencing, assembling, and annotating the *Aneurinibacillus* sp AFn2 bacterium. We aim to evaluate the performance based on contiguity, depth, and functional annotation of the resulting genome. Using ONT we generated 152,047 long reads assembling into 2 contigs with total base count of 0.4 billion which provided us efficient assembly while PacBio produced 139,701 reads, assembling into 21 contigs with a total base count of 1.4 billion. Functional annotation revealed differences in the number of coding sequences, with PacBio detecting more comprehensive gene sets than ONT. The comparative analysis done in this research shows the strengths and limitations of both the platforms, with ONT providing higher assembly contiguity and PacBio offering greater detail in genetic content. We aim to offer insights of both the sequencing technologies, guiding researchers in selecting the appropriate technology.**



***Keywords*— Aneurinibacillus, Nanopore, PacBio, Sequencing, Genome Assembly**


## I. Introduction

DNA sequencing is a method used to figure out the order of nucleotides in the entire genome. Thanks to the development of computing tools, it has achieved enormous growth in recent years[1]. Knowledge of the complete genome provides groundbreaking information about the studied organism.

Oxford Nanopore Technologies (ONT), developed by Oxford Nanopore Technologies, has revolutionized the field of genomics by providing a method for long-read sequencing that allows for the direct, real-time analysis of DNA or RNA molecules. This innovative approach has significantly improved the ability to sequence complex genomic regions, facilitating a better understanding of structural variations and epigenetic modifications. The technology's adaptability and portability, exemplified by the MinION device[2], has expanded its applications from clinical diagnostics to environmental monitoring, presenting new opportunities for in-field testing and rapid data acquisition. It is a fascinating combination of biology and technology[3]. It consists of a tiny hole known as nanopore which is embedded in synthetic membrane with a continuous flow of electric current through it[4]. When the DNA/RNA base is passed through the nanopore, it causes varying disruptions depending on the type if nucleotide base (A, T, G, C), with each nucleotide affecting the flow uniquely. These disruptions generate distinct electrical signals, referred to as squiggles, which are recorded and stored by sequencer, and are later interpreted into sequences by sophisticated algorithms. Despite its advantages, nanopore sequencing faces challenges such as higher error rates compared to traditional sequencing technologies. However, ongoing research and development are focused on enhancing the accuracy and reliability of nanopore sequencing reads. By integrating advanced data analysis tools and improving sequencing chemistry, scientists aim to mitigate these limitations, thereby broadening the technology's applicability and making it a more robust tool for genomic studies. The continued evolution of nanopore sequencing promises to unlock further insights into genomic complexities and drive innovations in personalized medicine, public health, and biodiversity conservation. The technology's application extends from healthcare to environmental monitoring, showcasing its versatility and potential in various fields. However, challenges such as higher error rates and data interpretation persist, requiring continuous refinement and combined approaches to reach sufficient accuracy[2].

Pacific Biosciences (PacBio) sequencing technology is renowned for its high accuracy and long-read capabilities, offering a comprehensive view of genomes and transcriptomes.

This technology, known as Single Molecule, Real-Time (SMRT) sequencing, provides not only the sequence but also the modification status of DNA bases, which is crucial for understanding the epigenetic context and complex genomic regions. It has been instrumental in resolving highly repetitive sequences and facilitating the de novo assembly of genomes, which is pivotal in areas like genetic diversity, species identification, and evolutionary studies. The development and application of PacBio sequencing has significantly impacted the understanding of microbial communities, plant genetics, and human healt[5]. Its ability to generate long reads with uniform coverage and without bias is particularly advantageous for fully characterizing genomes, leading to more accurate and complete genomic information. Researchers have utilized this technology in various fields, including the comprehensive mapping of human genetic variation, understanding complex disease mechanisms, and exploring the biodiversity of uncharted biological samples. However, despite its advantages, challenges such as higher costs and lower throughput compared to other sequencing technologies remain. Ongoing developments aim to reduce these limitations, making PacBio sequencing more accessible and applicable across broader research areas. The future of PacBio technology holds promise for further enhancing genomic research, offering deeper insights into genetic structures and functions.

Thermophilic members of the *Aneurinibacillus* genus are gram-positive spore-forming bacteria species recently identified as astonishing producers of polyhydroxyalkanoates (PHAs)[6, 7]. PHAs, synthesized intracellularly in granular form by various microorganisms, have garnered significant attention due to their remarkable resemblance to traditional petrochemical plastics, coupled with their biodegradability, biocompatibility, and non-toxic nature. Consequently, research interest from numerous global scientific communities has been drawn to these materials, given their vast potential across diverse industries such as food, medicine, and agriculture[8]. The most prevalent PHA synthesized is the homopolymer poly(3-hydroxybutyrate) (P(3HB)), known for its high crystallinity, resulting in stiffness and brittleness. However, introducing other PHA monomers into the polymer chain disrupts crystallinity to varying degrees, rendering the PHA more amorphous and significantly improving its material properties, including melting temperature and elongation at break. This enhancement enables PHAs to better compete with synthetic plastics[9]. Thermophilic aneurinibacilli have been identified as unique PHA producers due to their abundance of non-specific PHA synthases, allowing for the incorporation of various substrates into the PHA structure. These bacteria have been reported to produce a range of PHA monomers such as 4-hydroxybutyrate, 4-hydroxyhexanoate, 3-hydroxyvalerate, 4-hydroxyvalerate, and 5-hydroxyvalerate, utilizing specific precursors such as lactones and dioles[10]. This broad spectrum of PHA monomers sets thermophilic aneurinibacilli apart from other PHA producers, underscoring the importance of thoroughly exploring their full potential. One approach to achieving this is through genome analysis, particularly examining functional annotation, which can provide valuable insights into the metabolism of these unique microbes, better understand the physiological capabilities of bacteria, including their metabolic pathways, energy production, and nutrient utilization. This knowledge is essential for understanding how bacteria interact with their environment, how they cause disease, and how they can be exploited for biotechnological applications such as bioremediation or biofuel production. *Aneurinibacillus* sp. AFn2 underwent PacBio and ONT sequencing to perform assembly and check structural and functional annotation of the AFn2 genome. Additionally, the obtained datasets allowed the comparison of the effectiveness of both sequencing methods.

## II. Materials and Methods

### A. Microorganisms, cultivation, and DNA isolation

*Aneurinibacillus* sp. AFn2 was originally isolated from activated sludge from a wastewater treatment plant (Bystřice pod Hostýnem, Czech Republic, 2019) using an original isolation procedure, osmoselection[10]. The bacterium was maintained as the frozen stock culture at −80 °C in the presence of 10 % (v/v) glycerol and later used for the inoculum development. To enhance growth for the DNA isolation, strain AFn2 was inoculated in complex Nutrient Broth media composed of 10 g/L beef extract, 10 g/L bacteriological peptone, and 5 g/L NaCl in 100 ml Erlenmeyer flasks. The cultivation occurred under bacterium optimal growth conditions; 50°C at a shaking frequency of 180 rpm for 24 hours. Following the cultivation, a required amount of bacterial cells was harvested (centrifugation at 6000 rpm, 5 min), washed with a buffer, and centrifugated again under identical conditions. DNA isolation was performed using PowerSoil Pro (Qiagen, NL) isolation kits.

### B. Sequencing

Sequencing of *Aneurinibacillus* sp. AFN2 was perfomed by both the technology PacBio and ONT. For ONT the sequencing library was prepared using Rapid Barcoding Kit V14 (Oxford Nanopore Technologies, UK). The sequencing was performed using the R10.4.1 flowcell and the PromethION platform. PacBio sequencing was performed using PacBio Sequel IIe platform (EMBL, Genomics Core Facility, Heidelberg, Germany).

### C. Genome assembly

Genome assembly was done on the sequenced data of *Aneurinibacillus* sp. AFn2 using both PacBio and ONT. The HiFi reads from PacBio and sequence from ONT were assembled using unicycler v0.5.0[11]. Which is an assembly tool designed for bacterial genomes, integrating both short-read and long-read sequencing data. By applying hybrid approach, combining both de Bruijn graph and overlap-layout-consensus methods, optimizing the assembly process for accuracy and completeness. The assemblies were stored in fasta file. Post that the fasta files of assemblies were used for the quality assessment of the genome assembly using QUAST tool v5.0.2[12]. Subsequently PhaBOX server was run to look for phage DNA with the goal to have a better understanding of the assembly result[13].

### D. Annotation

Structural and functional annotation of the assemblies from both sequencing technologies were performed. For genome annotation we used prokka tool v1.12 which uses a combination of ab initio prediction, similarity searching against databases, and heuristic methods to annotate genomic features[14]. It integrates tools like Prodigal for gene prediction and HMMER for identifying protein domains, alongside custom databases for matching against known sequences. Standardized output in the form for fasta format and. tsv was given out. fasta file of amino acid from the structural annotation was then used for functional annotation. Functional Annotation was performed using online version of BlastKOALA[15]. The methods flowchart of this study is available in Fig. 1.

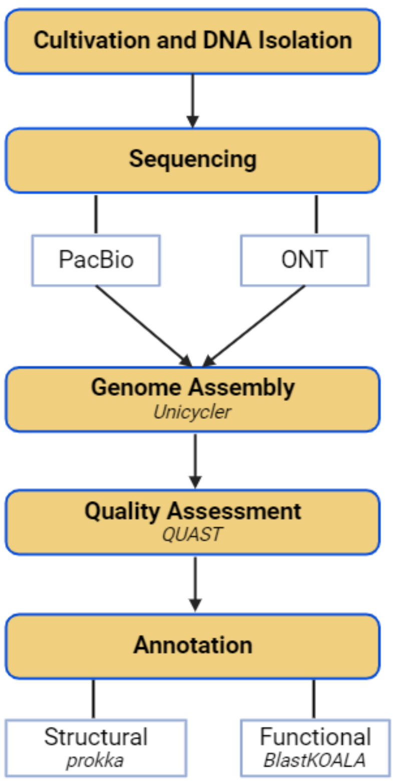



Fig. 1. Methodology Overview

## III. Results and Discussion

### A. Genome Assembly

The genome of the *Aneurinibacillus* species AFn2 from both ONT and PacBio was reconstructed. Sequences from PacBio generated 139,701 reads containing a total of 1.4 billion base counts and ONT generated 152,047 long reads containing a total of 0.4 billion bases creating. Assembly of PacBio sequences resulted in generation of 21 contigs with 1 circular and 20 linear unitigs of total size 4,326,806 bp. While ONT gave us only 2 contigs, 1 linear and 1 circular unitig with the total size 3,699,339 bp. This distribution of contigs with their size can be seen in QUAST report in Fig. 2. Since most of the linear contigs were small fragments to elucidate the biological significance and potential functionality of these contigs, we ran the PhaBOX server to characterize the contigs and check for any inconsistencies in our sample. Upon further analysis with PhaBOX , three of the linear contigs from the PacBio assembly were identified as originating from phage DNA belonging to the Autographiviridae family, exhibiting virulent characteristics. Additionally, the linear unitig from ONT was also found to be phage DNA with virulent properties. Following this to understand more about the short contig of PacBio we compared them with the discarded reads of the ONT to assess their conservation across different long read platforms, no such similarity or any additional information was found.

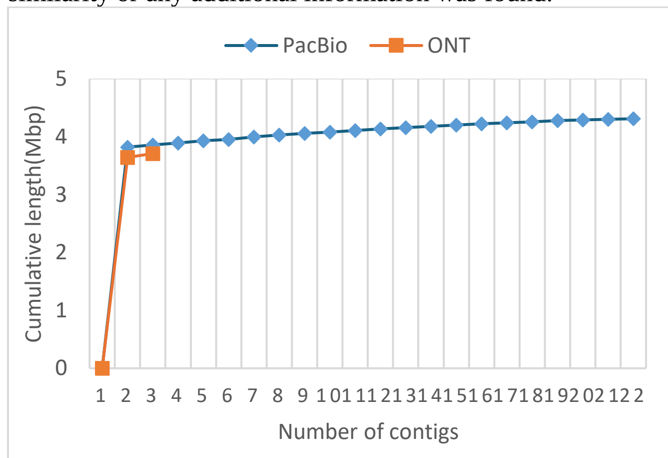



Fig. 2. Quast analysis of PacBio vs ONT sample

### B. Annotation

After annotating sequencing data from PacBio and ONT, the major structural components into which the genome was divided were coding sequences (CDS), tRNA, rRNA, and tRNA. Their distribution can be seen in table 1 below.

TABLE I. Structural annotation

| *Structural Category* | *PacBio* | *ONT* |
|---|---|---|
| Coding sequence(CDS) | 4150 | 3736 |
| rRNA | 6 | 32 |
| tRNA | 67 | 115 |
| tmRNA | 1 | 1 |
| Repeated regions | 0 | 5 |
| GC % | 70 | 45 |

Out of 4150 CDS identified within PacBio sequencing data, approximately 14% of them were found to be localised in the short contigs regions while in ONT only 2% of the total CDS were found to be in the short contig region thus suggest more notable consistency in the prevalence of coding sequences (CDS) within large contig regions across both platforms despite the differences in short contig content. The GC rich region in Pacbio sample was found to be 75% as opposed to 45% in ONT. The lower GC% in ONT assembly can indicate a higher error rate in Nanopore sequencing technology compared to PacBio, especially in homopolymeric regions. The repeated sequence of ONT might play a role into it as it can complicate the assembly thus causing underrepresentation of GC rich region and would lead to falsely lowered GC content.

higher GC content which is closer to bacterial genome averages. Both technologies have inherent biases. ONT tends to have

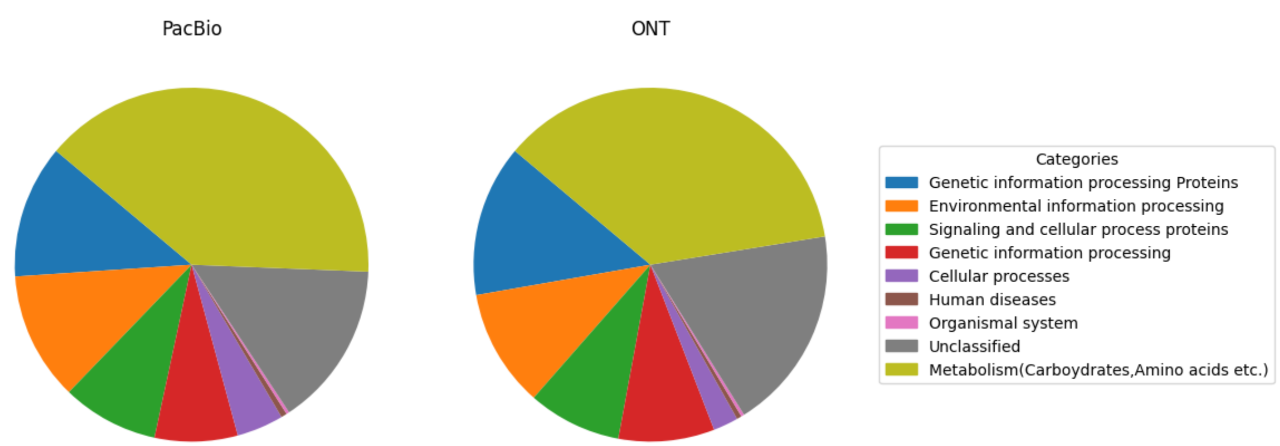



Fig. 3. A pie chart representation of the *Aneurinibacillus* sp. AFn2 and functional properties using PacBio vs ONT.

The functional annotation of both ONT and PacBio showed consistent results with each other with little to no variation amongs them. Most prevalent CDS in PacBio sequence was found to be metabolism which was around 40% and majority of contribution was from carbohydrate and amino acid metabolism, followed by genetic information processing protein which is around 12% and 15% of them were found to be unclassified. While In ONT, metabolism contributed around 36%, majority of which were again carbohydrate and amino acid metabolism, followed by genetic information processing protein which is 14% and around 19% was found to be unclassified from the total CDS as shown in the figure 3.

## IV. Conclusion

The comparative analysis done in this research shows some distinct outcomes in terms of assembly metrics and genomic features. PacBio gives more fragmented genomic assembly with 21 contigs while ONT provides only 2 adjoining contigs. Furthermore, PacBio sequencing is identifying more tRNA and rRNA than ONT which reflects the differences in sequence capture or assembly efficiency between technologies, but the coding sequences (CDS) showed the very distinct distribution across contigs size. While the total CDS is higher in PacBio dataset but they show almost similar number in the large contigs region in both the technologies. This consistency suggests that the core genomic architecture, particularly in terms of larger contigs containing CDS, is robustly captured by both sequencing methods. The functional analysis at the end shows almost similar kind of distribution of gene in metabolism and genetics processing protein with giving little edge to PacBio technology and higher percentage of unclassified gene in ONT can suggest possible differences in gene prediction and annotation accuracy. The findings underscore the importance of selecting the appropriate sequencing technology based on the specific goals of genomic studies. While ONT may provide simple genome reconstruction with the help of better contiguity thus reducing genome complexity, PacBio may also provide advantages in terms of genomic detail and accuracy, which is evident by its difficulty with homopolymeric regions, which can affect base calling and lead to inaccuracies in regions with specific nucleotide repeats. The differences stated are crucial for bacterial functionality and metabolism, especially in environments where genomic information is important. Further research can help us elaborate the balance between depth and assembly as well as functional implications of observed genomic differences to optimize the use of these technologies in microbial studies.

Our study contributes to the evolving field of genomic sequencing by providing insights into the strengths and limitations of PacBio and ONT platforms in assembling and annotating bacterial genomes, which can guide researchers in selecting the most appropriate technology for their scientific inquiries.


## Acknowledgment

This work was supported by Inter Faculty Junior Project 2024-FCH/FEKT-J-24-8516


## References


[1] Stephens, Z.D., et al., *Big data: astronomical or genomical?* PLoS biology, 2015. **13**(7): p. e1002195.

[2] Jain, M., et al., *The Oxford Nanopore MinION: delivery of nanopore sequencing to the genomics community.* Genome biology, 2016. **17**: p. 1-11.

[3] Kchouk, M., J.-F. Gibrat, and M. Elloumi, *Generations of sequencing technologies: from first to next generation.* Biology and Medicine, 2017. **9**(3).

[4] Jain, M., et al., *Nanopore sequencing and assembly of a human genome with ultra-long reads.* Nature biotechnology, 2018. **36**(4): p. 338-345.

[5] Rhoads, A. and K.F. Au, *PacBio sequencing and its applications.* Genomics, Proteomics and Bioinformatics, 2015. **13**(5): p. 278-289.

[6] Subhash, Y., H.-j. Kim, and S.-S. Lee, *Aneurinibacillus sediminis sp. nov., isolated from lagoon sediments.* International Journal of Systematic and Evolutionary Microbiology, 2017. **67**(8): p. 2544-2548.

[7] Pernicova, I., et al., *Application of osmotic challenge for enrichment of microbial consortia in polyhydroxyalkanoates producing thermophilic*

*and thermotolerant bacteria and their subsequent isolation.* International journal of biological macromolecules, 2020. **144**: p. 698-704.

[8] Mozejko-Ciesielska, J., K. Szacherska, and P. Marciniak, *Pseudomonas species as producers of eco-friendly polyhydroxyalkanoates.* Journal of Polymers and the Environment, 2019. **27**: p. 1151-1166.

[9] Janigová, I., I. Lacík, and I. Chodák, *Thermal degradation of plasticized poly (3-hydroxybutyrate) investigated by DSC.* Polymer Degradation and Stability, 2002. **77**(1): p. 35-41.

[10] Pernicova, I., et al., *Introducing the newly isolated bacterium Aneurinibacillus sp. H1 as an auspicious thermophilic producer of various polyhydroxyalkanoates (PHA) copolymers–1. Isolation and Characterization of the Bacterium.* Polymers, 2020. **12**(6): p. 1235.

[11] Wick, R.R., et al., *Unicycler: resolving bacterial genome assemblies from short and long sequencing reads.* PLoS computational biology, 2017. **13**(6): p. e1005595.

[12] Gurevich, A., et al., *QUAST: quality assessment tool for genome assemblies.* Bioinformatics, 2013. **29**(8): p. 1072-1075.

[13] Shang, J., et al., *PhaBOX: a web server for identifying and characterizing phage contigs in metagenomic data.* Bioinformatics Advances, 2023. **3**(1): p. vbad101.

[14] Seemann, T., *Prokka: rapid prokaryotic genome annotation.* Bioinformatics, 2014. **30**(14): p. 2068-2069.

[15] 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): p. 726-731.

# *In silico* Analysis of Rutin from Ruta chalepensis.L for NF-X1 Inhibition in Cervical Cancer: Insights from HPTLC studies


Laxmipriya Rajasekaran[1]
*Department of Biomedical Engineering, Faculty of Electrical Engineering and ommunnication,*
*Brno University of Technology*
Brno, Czech Republic
253932@vut.cz

Vaishali Pankaj[1]
*Department of Biomedical Engineering, Faculty of Electrical Engineering and Communication*
*Brno University of Technology*
Brno, Czech Republic
234084@vut.cz

Inderjeet Bhogal[1]
*Department of Biomedical Engineering, Faculty of Electrical Engineering and Communication*
*Brno University of Technology*
Brno, Czech Republic
234190@vut.cz

Darina Cejkova[1]
*Department of Biomedical Engineering, Faculty of Electrical Engineering and Communication*
*Brno University of Technology*
Brno, Czech Republic
cejkovad@vut.cz



***Abstract*- Bioactive compounds play a pivotal role in the majority of pharmacological activities. Rutin, in particular, exhibits great potential in combination with antitumor drugs and various herbal formulations, enhancing permeability and regulating apoptosis induction. Earlier studies revealed the presence of alkaloids, flavonoids, phenols, quinines, and steroids among other compounds. Moreover, bioactive compounds were previously identified using GC-MS, with a total of 32 compounds identified within the plant sample *Ruta chalepensis* L, exhibiting several beneficial effects on leaf extract carried out through antioxidant properties. The leaves of *Ruta chalepensis* L were previously studied to evaluate the biological constituents and phytochemical screening. In this study, the biologically active compound was performed using high-performance thin-layer chromatography (HPTLC) plates silica gel 60 F 254, with rutin as a standard compound and methanol as a sample solvent. In cervical cancer, NF-X1 is transcriptional repressor which is highly expressed caused by HPV-associated drug compounds, was predicted to bind to rutin compound through *in-silico* modeling. In conclusion, it was found Rutin could be a potential modulator using molecular docking and simulation studies and has a potential therapeutic role in cervical cancer treatment.**



***Key words*:** *Ruta chalepensis*. L, Rutin, HPTLC, NF-X1, *in silico*


## I. Introduction

Cervical cancer (CC) is the fourth most common pathology in women worldwide and presents a high impact in developing countries due to limited financial resources as well as difficulties in monitoring and access to health services. [1] The discovery and development of novel multi-targeted agents to attenuate the dysregulated signaling in cancer is of great importance. In recent decades, phytochemicals from dietary and medicinal plants have been successfully introduced as alternative anticancer agents due to their ability to modulate numerous oncogenic and onco suppressive signaling pathways (2).

Plant-derived natural products have attained great attention in drug discovery programs [3]. Growing evidence demonstrates that cytostatic effects of natural products are derived from their potential in modulating oxidative stress, inflammation, autophagy and apoptosis, thereby leading to the prevention/reduction of their associated toxicity [4]. The key flavonoid found in fringed rue (*Ruta chalepensis* L.) is rutin that has wide variety of medicinal applications [5]. Several differentially expressed genes and transcripts relating to rutin biosynthesis were identified in leaves comparing with roots or stems comparing with roots. All genes known to be involved in rutin biosynthesis showed up-regulation in leaves as compared with roots [4].

The available evidence can be divided into three classes of *in vitro*, *in vivo*, and clinical studies, contributing to up-to-date information on the available mass of knowledge in the field of rutin. Although the effectiveness of rutin on cancer has been documented, the comparison of rutin remained unclear [6]. Most of the current anticancer evidence of rutin is focused on *in vitro* models of cancer, with very limited *in vivo* studies. [1].

NFX-123 is splice variant isoform of the transcriptional repressor NF-X1 gene. It is highly expressed in cervical cancers caused by HPV, and NF-X1 is a protein partner with the HPV oncoprotein E6 [7]. Together, NFX1-123 and E6 affect cellular growth, longevity, and differentiation. The expression status of NFX1-123 in cancers beyond cervical and head and neck cancers, and its potential as therapeutic target, have not been investigated. NFX1 are expressed in epithelial cells, and they both were found to bind to the 16E6 oncoprotein and have association with cervical cancers [8].

In this study, the quantitative determination is done for the biologically active phenolic compound rutin using high-performance thin-layer chromatography (HPTLC). *In silico* methods utilized to identify the binding activity of NF-X1 receptor to rutin compound through molecular docking, molecular dynamic simulations and binding free energy calculations for treatment of cervical cancer.

## II. Material and Methods

*Experimental Preparation*

The leaves of *Ruta chalepensis* L were collected from the hilly region of Udhagamandalam, The Nilgiris, Tamil Nadu, India. The standard of rutin (Sigma Aldrich Chemical Private Ltd, Bangalore) was used without any pretreatment and with no modification.

*Preparation of Sample and Standard solution*

Standard solution containing 1mg $mL^{-1}$ of rutin used for calibration studies. The collected plant leaves were dried under the shade of sunlight for 2-3 days and powered using mortar and pestle. Then powder was extracted with methanol (4x10ml). The sample was vortexed for 2 minutes and left for 1day at room temperature. The extract (20mg) was taken and dissolved in methanol (1Ml) and filtered through filter for further HPTLC study.

*TLC – scanner*

A CAMAG (E. Merck KG, Darmstadt, Germany) HPTLC system equipped with an automatic TLC sampler, TLC scanner 3 (WINCATS version 2.01.02) visualized under UV cabinet and twin trough glass tank (10x10cm) was used for the analysis. Plates were developed in an Automatic TLC sampler (ADC2) with toluene/ethyl acetate/formic acid (6:4:0.8) as the mobile phase(70mm). The saturation time of the chamber were optimized to 5mins at $60^0$C.

*Illumination remission*

The spectra were captured by the Illumination instrument (CAMAG Visualizer_170503) at 80% digital mode and the exposure is at 116.61 ms at the pathlength of 1cm with the 715Pxl x 715Pxl (0.13 mm/ Pxl) fig.1. The HPTLC had a very strong absorption spectrum and it captured as a default correction at 254nm remission. Comparatively the exposure raised to 1039.83ms with the default correction type by 366nm remission.

*Quantification and Calibration*

The standard solution was applied on the HPTLC for the calibration cures and it is observed Track MWL TrackSc4 at all measured wavelengths.

*Binding Site prediction and Molecular Docking*

The prediction of potential active site of the target protein is an important aspect in drug discovery. The alphafold structure of NF-X1 was downloaded from Alphafold protein structure database with ID AF-Q12986-F1. The active site of homology model Alpha Fold structure NF-X1 was analyzed using Sitemap of Schrodinger 2023-3 [9]. SiteMap is an accurate and potential binding site prediction tool, provides site score which is used to identify effective binding sites on the surface of the protein. Sitemap results evaluation locates binding sites showing hydrogen bond donors and acceptors, hydrophobic and hydrophilic regions. Induced-fit Docking (IFD) was performed on the different sites generated from Sitemap to reveal the interaction between rutin compound and NF-X1 protein. IFD is based on rigid receptor docking to provide the best docking conformations. IFD studies were performed using Glide [10] module (extra precision) with twenty poses generated for each site and the best scoring pose for selected on the basis of docking score.

*Molecular dynamic Simulations and Binding Energy Calculations*

MD simulations study was performed for best docked score site model and the conformational changes during the protein-ligand binding were investigated. MD simulations was performed for 100ns using Desmond Schrodinger [11]. Simulation experiment is completed in three stages: system builder, Energy minimization and Molecular dynamics via NPT ensemble run. The pressure is adjusted to 1 bar and temperature up to 300K. The simulation results were analyzed using a simulation interaction diagram. The binding free energy of the docked complex was computed using prime MM-GBSA (Molecular Mechanics-Generalized Born Surface Area) [12]. MM-GBSA computes the difference between the free energy of protein, ligand and complex in the system.

## III. Results and Discussion

*Quantification in Pharmaceutical Preparations*

A HPTLC method was developed for fast and simple quantitative determination of rutin in pharmaceutical preparations (13). The chromatography with methanol solvent as sample 2 matched that of the rutin standard obtained with the Camag TLC Scanner. The chosen HPTLC conditions provided which appeared as a spot in visible light.

*Method Validation and Optimization*

Moreover, the compounds were identified based on the Rf values, with both the rutin standard and methanol solvent having Rf values of 0.03 and a window size of about 1.000. The minimum threshold height starts with 10AU and maximum peak threshold height till 990AU with the track start position 5.0mm and end position with 75.0mm with automatic display.

*Application Position and Quantitative Evaluation*

The rutin compound was initially identified at the position 65.0mm with the volume of 3.0µl, with the band length upto 8.0mm measured 2 tracks with dosage volume up to 150nl/s. Quantitative evaluation of the plate was performed in the reflectance/absorbance mode at 75.5mm, under specific conditions: slit width of 4.00 x 0.03mm, micro scanning speed of 20 mm $s^{-1}$, and data resolution of 100µm $step^{-1}$.

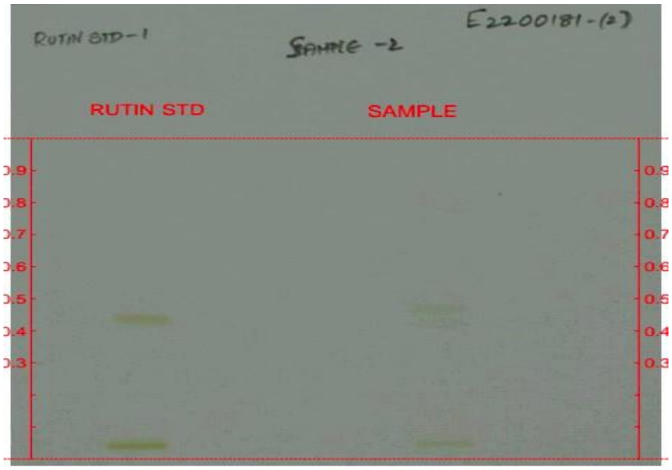



Fig. 1. In the application position Y :Rutin and Sample (Methanol solvent) through the CAMAG Visualizer:170503

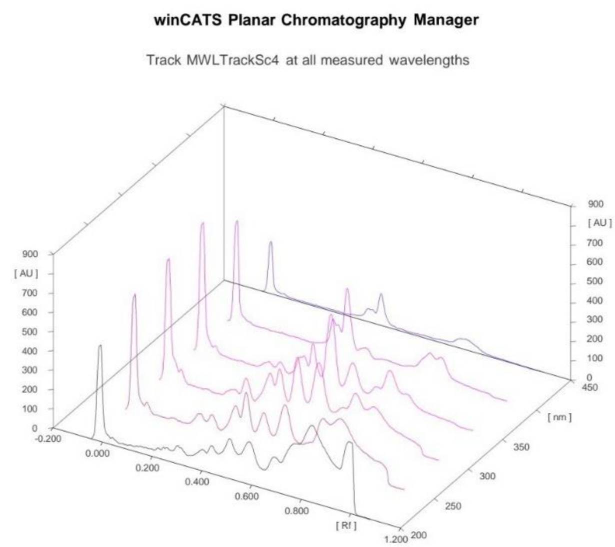



Fig. 2. Dimensional HPTLC chromatogram of standard rutin and methanol solvent

*Binding site prediction and Molecular Docking*

The potential binding site for the predicted alphafold structure NF-X1 protein was done using default settings from Schrodinger's Sitemap tool. The results provided with top 5 binding sites along with its site scores. Sitemap examines the entire protein to locate binding sites, usually in the case of the predicted protein structures. The Site score greater than 1 is a suggestive of a promising binding site. The results provided site score ranging from 1 to 0.95. To assess the robustness of the results provided by the sitemap on the different sites were performed using Induced-fit docking from Schrodinger. The IFD studies performed revealed that site1 has higher binding affinity when compared with other sites. The docking scores and site scores for the top 5 sites predicted are provided in Table 1. The docking scores suggests that site 1 has more stable binding of the compound Rutin with the protein structure. The interacting amino acids involved in the binding site1 with rutin compound is shown in Figure 3.

TABLE I. SITESCORE AND DOCKING SCORES FOR THE NF-X1

| Model | Site score | Docking score | Interacting residues |
|---|---|---|---|
| NF-X1-site1 | 0.99 | -16.21 | Trp 372, Ser 412, Ala 413, His 414, Val 415, Asn 417, Thr 420 |
| NF-X1-site2 | 0.95 | -15.66 | Glu 887, Gln 889, Ser 905, Lys 938, Gln 940, Gln 942, Ala 943 |
| NF-X1-site3 | 1.05 | -15.15 | Asn 334, His 335, Leu 337, Arg 433 |
| NF-X1-site4 | 0.98 | -14.24 | Asn 334, His 335, Leu 337, Thr 352, Arg 433, Trp 739 |
| NF-X1-site5 | 0.99 | -9.99 | Tyr 908, Ser 915, Lys 919, Leu 934, Glu 946 |

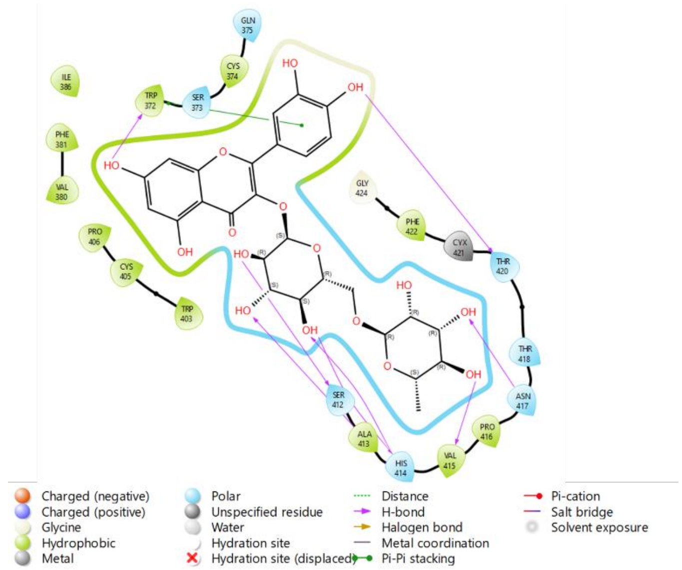



Fig. 3. Protein-ligand 2D interaction of NF-X1_site1-rutin complex

*Molecular Dynamics Simulations*

The molecular dynamics simulations for the NF-X1_site1 docked complex is performed to study the behavior of atoms and molecules over the simulation 100ns run. MD results were analyzed using structural changes in the biomolecular system. Trajectory analysis is done using conformational changes such as protein-folding, domain movements and ligand binding events. These structural changes are analyzed using root-mean square deviation (RMSD), root mean square fluctuations (RMSF) and protein-ligand histogram. RMSD defines the structural changes in the protein structure upon ligand binding. Protein-ligand RMSD also indicates that the simulation has been equilibrated throughout the 100ns run. The optimal range of RMSD is 1-4 Å, which in this case is seen to be exceeded as figure 4. The RMSD plot indicates that protein-ligand binding has formed interaction with the initial 2ns and fluctuated till 80ns and was stabilized after it.

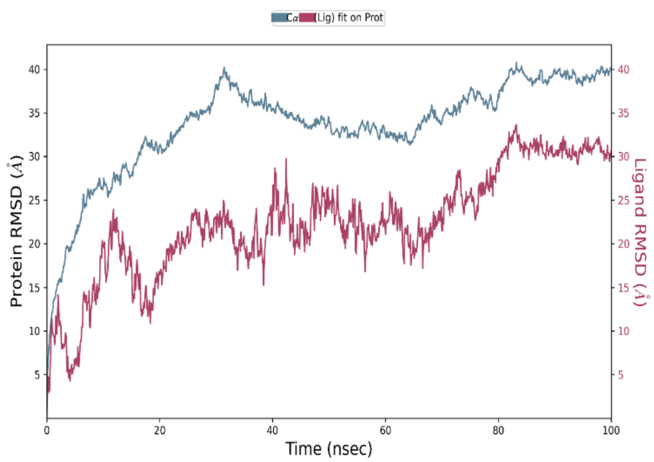


Fig. 4. Protein-ligand RMSD for the NF-X1_site1-Rutin complex

*Binding free energy calculations using MM-GBSA*

The prime MM-GBSA binding free energy calculations after MD simulation reveals the binding free energy (ΔGBind) as the difference between the free energy of protein-ligand complex. These methods provide insights into the contributions of different energy terms to binding energy such as coloumbic interactions (-18.94 Kcal/mol), HBonds (-1.66 Kcal/mol), Lipophilic (-5.59 Kcal/mol), solvation energy (22.77 Kcal/mol) and Van der Waals (-30.89 Kcal/mol) interactions. The ΔGBind for site1 calculates is -33.9 Kcal/mol. The results indicate that the rutin binding to NF-X1 protein shows stable binding complex.

## IV. Conclusion

In conclusion, our thorough investigation into the therapeutic capabilities of *Ruta chalepensis* L. extracts enhanced with rutin for cervical cancer has unveiled encouraging prospects. Employing HPTLC and in silico analysis of NF-X1, we identified the potential binding of rutin and the protein through molecular docking and simulation studies. These results emphasize the significance of natural compounds like rutin in presenting promising therapeutic paths. Furthermore, they underscore the importance of integrating diverse methodologies in cancer research to unveil novel opportunities for patient care and treatment outcomes.

## Acknowledgment

This research work has been carried out with the support of Centre of Bioscience and Nanoscience Research, Tamil Nadu, India for HPTLC work. We gratefully acknowledge the financial support provided by the FEKT-S-23-8196 project from Brno University of Technology.

## References


[1] Nouri, Z.; Fakhri, S.; Nouri, K.; Wallace, C.E.; Farzaei, M.H.; Bishayee, A. Targeting Multiple Signaling Pathways in Cancer: The Rutin Therapeutic Approach. *Cancers* **2020**, *12*, 2276. https://doi.org/10.3390/cancers12082276

[2] Gomes, D.; Silvestre, S.; Duarte, A.P.; Venuti, A.; Soares, C.P.; Passarinha, L.; Sousa, Â. In Silico Approaches: A Way to Unveil Novel Therapeutic Drugs for Cervical Cancer Management. *Pharmaceuticals* **2021**, *14*, 741. https://doi.org/10.3390/ph14080741

[3] Cragg, G.M.; Pezzuto, J.M. Natural products as a vital source for the discovery of cancer chemotherapeutic and chemopreventive agents. Med. Princ. Pract. 2016, 25, 41–59. [CrossRef]

[4] Abdel-Salam, Eslam M., Mohammad Faisal, Abdulrahman A. Alatar, Ahmed A. Qahtan, and Perwez Alam. "Genome-wide transcriptome variation landscape in Ruta chalepensis organs revealed potential genes responsible for rutin biosynthesis." *Journal of Biotechnology* 325 (2021): 43-56.

[5] Loizzo, Monica Rosa, T. Falco, M. Bonesi, V. Sicari, Rosa Tundis, and M. Bruno. *"Ruta chalepensis L.* (Rutaceae) leaf extract: chemical composition, antioxidant and hypoglicaemic activities." *Natural product research* 32, no. 5 (2018): 521-528.

[6] MoballeghNasery et al., 2020; Prasannan et al., 2012; Rajendran et al., 2012; Ramachandran et al., 2012; Sethi, Ahn, Sung, & Aggarwal, 2008; Siveen et al., 2014b; Siveen et al., 2014a; Tan et al., 2010

[7] Chintala S, Dankoski MA, Anbarasu A, Ramaiah S, Miryala SK, Katzenellenbogen RA. NFX1-123: A potential therapeutic target in cervical cancer. J Med Virol. 2023 Jun;95(6):e28856. doi: 10.1002/jmv.28856. PMID: 37288708; PMCID: PMC10264143.

[8] Chintala S, Levan J, Robinson K, Quist K, Katzenellenbogen RA. Genes Regulated by HPV 16 E6 and High Expression of NFX1-123 in Cervical Cancers. Onco Targets Ther. 2020 Jun 26;13:6143-6156. doi: 10.2147/OTT.S251926. PMID: 32617009; PMCID: PMC7326398.

[9] Schrödinger Release 2024-1: SiteMap, Schrödinger, LLC, New York, NY, 2024

[10] Schrödinger Release 2024-1: Induced Fit Docking protocol; Glide, Schrödinger, LLC, New York, NY, 2024; Prime, Schrödinger, LLC, New York, NY, 2024.

[11] Schrödinger Release 2024-1: Desmond Molecular Dynamics System, D. E. Shaw Research, New York, NY, 2024. Maestro-Desmond Interoperability Tools, Schrödinger, New York, NY, 2024.

[12] Genheden, S., & Ryde, U. (2015). The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities. Expert opinion on drug discovery, 10(5), 449-461

[13] Soponar F, Moţ AC, Sârbu C. High-performance thin-layer chromatography and three-dimensional image analysis for the determination of rutin in pharmaceutical preparations. J AOAC Int. 2010 May-Jun;93(3):804-10. PMID: 20629380.

# LSTM-Based Autoencoders in Online Handwriting Data Augmentation and Preprocessing

1st Michal Gavenciak
*Department of Telecommunications*
*Brno University of Technology*
Brno, Czech Republic
0000-0001-8835-1056

***Abstract*—On-line handwriting analysis is a research field that is among others used in assessment of handwriting difficulties (HD), which can be manifestations of degenerative brain diseases such as Parkinson's disease in the elderly, or developmental dysgraphia in children. Using advanced modelling approaches or artificial intelligence is often difficult because of the limited data availability in both demographic cohorts. In this article, a data processing approach, using LSTM-based autoencoders, is described as a way of augmenting the database with semi-synthetic data or preprocessing the data to improve the performance of feature-based classification. The proposed method has led to a 3 percentage point increase in classification accuracy when compared to baseline. While the improvement is marginal, it highlights another possible area of research to improve the efficacy of automated HD assessment.**

***Index Terms*—Handwriting difficulties, XGBoost, LSTM, autoencoder**

## I. Introduction

On-line handwriting analysis is becoming an increasingly more common research topic. There are several domains which can use these analytical methods, such as security, and writer identification. Furthermore there is increasing evidence that handwriting analysis can be used to assess diseases/disorders, such as Parkinson's disease in the elderly [1], or developmental dysgraphia in children population [2]. The analysis is typically carried using some sort of artificial intelligence from the domains of neural networks, or decision trees, such as the commonly used XGBoost [3]–[6]. What these approaches typically have in common is, that they require a substantial amount of data on which these models can be trained. However, real-life data are often difficult to gather, especially among the cohorts in which these handwriting difficulties are usually found. Gathering data and handwriting samples from both children and the elderly can require extra effort and prove more time-demanding. These efforts were recently also made more difficult due to the coronavirus pandemic. One of the ways that can improve data scarcity is introducing synthetic data, which were either manually simulated, e.g. a person simulating a disorder with her/his non-dominant hand, or generated computationally. However, such synthetic data might not reflect the nature of the real-life examples. Another approach to increasing the volume of data is data augmentation, which is a process where existing samples are somehow altered and introduced back into the dataset as unique entries with the same annotation as the original sample. An example of a simple data augmentation technique can be adding noise to the handwriting signal. This approach is not without flaws either, as introducing a copy of a sample into the training dataset, even if altered, can lead to overtraining of the resulting model. The augmentation process could also alter the characteristics of the handwriting which correlate with the assessment goal, bringing the overall performance down.

This article will describe the data augmentation process of an online-handwriting database and the effect that the introduction of these semi-synthetic samples into the database had on the classification accuracy of an XGBoost classifier when identifying school children with handwriting difficulties. The input into the XGBoost classifier consists of a set of features that are first extracted from the online/handwriting.

Two uses of the augmented data will be considered:

- Using the augmented samples as unique entries in the process of classifier training, then observing if the model performance increased in classification of unprocessed samples.
- Using the augmented samples to extract the same features and appending them as additional features, that is, using the data augmentor as a filter on the handwriting signal.

## II. Methodology

A rough overview of the used methodology can be seen in Figure 1 and is further described below.

### A. Dataset

The experiment was carried out using a database of online-handwriting, which consists of 106 samples. The samples were collected using a digitising tablet (Wacom Intuos Pro L PHT-80). The participants were shown a template of the Archimedean spiral (see Figure 3) and were asked to copy it using a Wacom Inking Pen on a paper fixed to the tablet. This approach allows sampling the handwriting in real-time, capturing it as a function of time, while also giving the participants the usual haptic feedback of writing on paper, rather than directly on the display of a tablet. The captured properties are the x and y coordinates of the pen, timestamp,

This study was supported by a project of the Technology Agency of the Czech Republic no. TL03000287 (Software for advanced diagnosis of graphomotor disabilities).

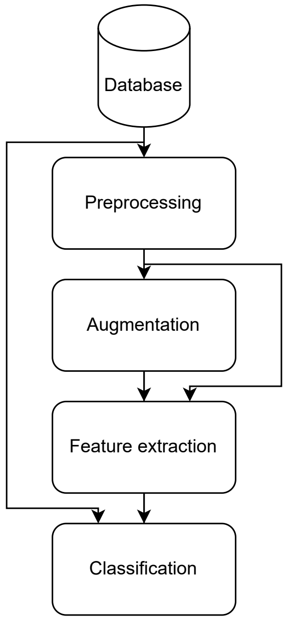



Fig. 1. Flowchart diagram of the proposed methodology

pressure applied to the pen-tip, azimuth, and tilt of the pen. Additionally, the tablet can track the movement of the pen up to 1.5 cm above the surface and indicates whether the pen is or is not touching the tablet. The sampling frequency is 150 Hz. A diagram of the capturing process can be seen in Figure 2.

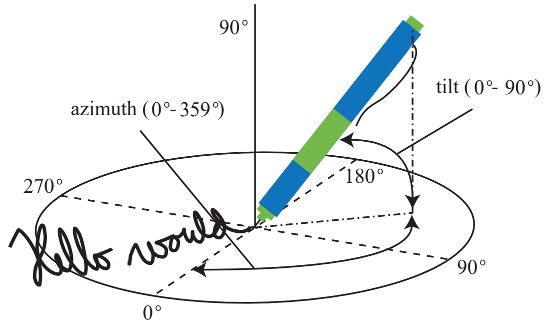



Fig. 2. Diagram of the handwriting sampling process.

The participants were children attending 3rd and 4th grade of primary schools in the Czech Republic, where 66 of the participants were boys and 40 girls. The demographic characteristics of the participants can be seen in Table I. The children were all assigned the Handwriting Deficiency Score (HDC) [7], which is a scoring method combining the Handwriting Proficiency Questionnaire for Children (also known as HPSQ–C) [8], [9] with an assessment provided by an expert remedial teacher in the range of [0; 3]; 3 denoting severely deficient handwriting and 0 denoting intact handwriting. Additionaly, to allow the use of a classification model, the HDC scores were binarised with a threshold (HDC–T) to split the dataset into intact and deficient handwriting. The resulting scoring can be seen in Table II. Examples of how the Archimedean spiral was completed by a child in the negative and positive group can be seen in Figure 4.

TABLE I. Demographic data.

| | | | **3rd grade** | | |
|---|---|---|---|---|---|
| Gender | N | DD | Age [y] | HDC | HDC>1 |
| Boys | 17 | 3 | 8.75±0.56 | 1.06±1.02 | 5 |
| Girls | 16 | 2 | 8.91±0.59 | 0.75±0.86 | 4 |
| Total | 33 | 5 | 8.83±0.57 | 0.91±0.95 | 9 |
| | | | **4th grade** | | |
| Gender | N | DD | Age [y] | HDC | HDC>1 |
| Boys | 49 | 8 | 9.90±0.53 | 1.22±1.06 | 20 |
| Girls | 24 | 0 | 9.90±0.51 | 0.33±0.70 | 3 |
| Total | 73 | 8 | 9.90±0.52 | 0.93±1.04 | 23 |

[1] N – Number of samples; DD – developmental dysgraphia; y – years; HDC – Handwriting Deficiency Criterion

TABLE II. Handwriting disabilities criterion.

| OEE | OEE(t) | HPSQ–C(t) | HDC | HDC–T |
|---|---|---|---|---|
| 0 | 0 | 0 | 0 | 0 |
| 1 | 0 | 0 | 0 | |
| 2 | 0 | 0 | 0 | |
| 2 | 0 | 1 | 0 | |
| 3 | 1 | 0 | 1 | 0 |
| 3 | 1 | 1 | 2 | 1 |
| 4 | 1 | 0 | 2 | |
| 4 | 1 | 1 | 3 | 1 |

[1] OEE – overall expert evaluation; OEE(t) – overall expert evaluation (thresholded); HPSQ–C(t) – child's self-evaluation (thresholded); HDC – new handwriting disabilities criterion; HTC-T – handwriting disabilities criterion (thresholded).

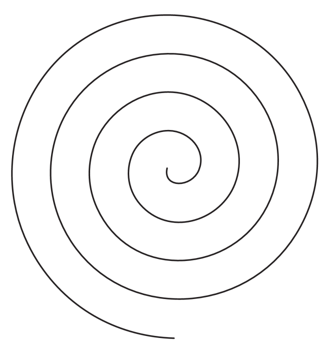


Fig. 3. Template of the Archimedean spiral drawing task (real width = 750 mm, real height = 750 mm).

The project this study is a part of was approved by the Ethics Committee of the Masaryk University and parents of all children enrolled into this study signed an informed consent form. Throughout the whole study, the Ethical Principles of

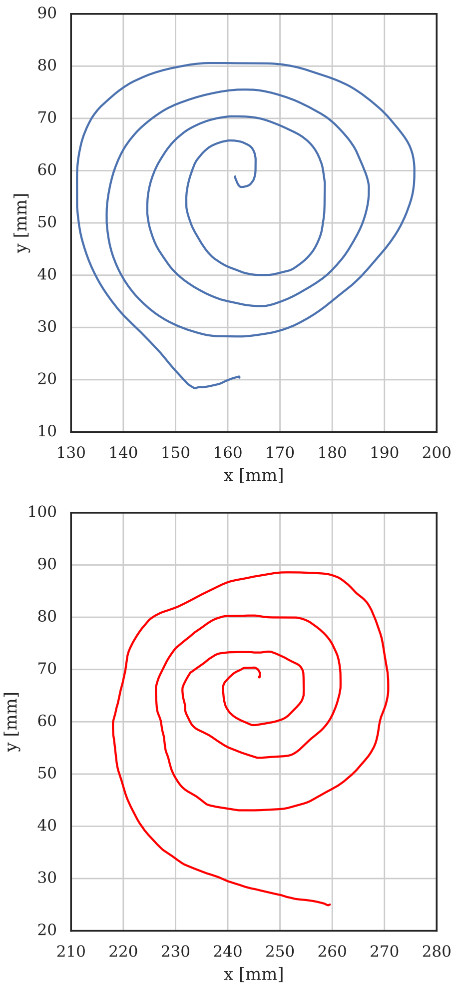



Fig. 4. Example of a spiral sample performed by a child in the HDC-T negative group (top in blue) and by a child in the HDC-T positive group (bottom in red).

Psychologists and Code of Conduct released by the American Psychological Association were followed [10].

### B. Augmentation

Before augmentation, all data samples were preprocessed. First, the information about pressure, tilt and azimuth were removed from the files. Subsequently, above-surface (also called in-air) movement was also removed along the boolean value indicating whether the inking pen is in contact with the tablet. Only the x and y coordinates along with the timestamp were left in the files. The timestamps in all the files were additionally normalised so that they started at the 0 value.

The time-data were then grouped into overlapping windows with the length of 3 samples, producing a 3-by-3 matrix. The samples in the window were always subsequent and ordered chronologically. One window never contained data-points from two different handwriting products. As this was an introductory study, the window size was selected experimentatively. During preliminary testing, it was discovered, that smaller window sizes produced more consistent and predictable results and thus made the tuning of the pipeline more efficient.

These windows were then used to train an LSTM model, which was set up as an autoencoder. While autoencoders are typically fully symmetrical, two dropout layers were introduced in the encoder portion of the model, to increase the transformative effects of the model, as the goal of this experiment was to alter the data. In other words, the LSTM model was used as a lossy compression algorithm and several combinations of the layers' settings were tested. A diagram of the autoencoder can be seen in Figure 5. The layer size progression was kept symmetrical, with the input and output (Encoder_1 and Decoder_3) layers having 8 units, the middle layers (Encoder_2 and Decoder_2) having 6 units and the inner-most, bottleneck layers, having 3 units. Dropouts were set in two different scenarios; in single-dropout approach, Dropout_1 was set to 0.3 and Dropout_2 was left inactive, while in a double-dropout scenario, both Dropout_1 and Dropout_2 were set to 0.1.

This model was then used to produce an augmented handwriting sample for each original unprocessed handwriting sample. To ensure that the model was not trained on a sample it is then asked to reproduce, the augmentation was done in a fashion resembling the leave-one-out cross-validation method, where the model was trained on N-1 samples (these were also divided into train/test sets) and then used to encode/decode the one sample that was held out from the train/test sets, meaning that 106 models were trained and used for augmentation for each scenario (double- and single-dropout).

An example of augmented handwriting samples (same samples as in Figure 4) can be seen in Figure 6.

### C. Feature extraction

After the augmentation, the resulting datasets were used to compute features describing the character of the sampled handwriting. The features used consisted of velocity, acceleration, jerk, and spiral height and width. Of these metrics, statistics of mean, std, median, inter-quartile range, etc. were calculated. Features that were missing more than 10 % of values, were discarded. Features than were missing less than 10 % were filled with the feature's mean value.

### D. Classification

Classification was carried out by the XGBoost classifier for each of the 4 final scenarios, where each scenario denotes that the feature matrix was obtained with a different augmentation approach. These scenarios are: original data, single-dropout, double-dropout, and training-set extension. Before classification, the confounding factors of sex, and grade were controlled for (using linear regression). The XGBoost classifier was selected as it generally offers more interpretability of the model than deep learning networks, through the possibility of inspecting the model's feature importance. While interpretabilty of the model was not among the core interests of this study, the general approach was kept in line with previous papers in this domain [5], [6].

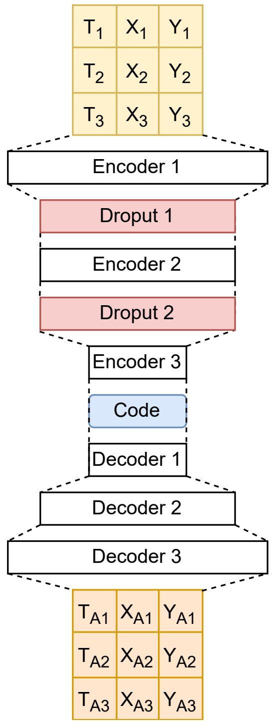



Fig. 5. Schematic of the autoencoder used to augment handwriting data. Input on top; $T$, $X$, and $Y$ are timestamp and XY coordinates in the input matrix; $T_A$, $X_A$, $Y_A$ are timestamp and XY coordinates in the output matrix.

Training of the models was carried out in a two-round manner, in order to find the best performant set of features. First, the model was trained with the whole feature set for prediction of the child's HDC–T score, using Bayesian hyperparameter tuning with stratified 5-fold cross-validation (5 repetitions). The best performing model's hyperparameters and feature-importance coefficients are then used to retrain the model in the second round.

In the second round, the model is trained using leave-one-out crossvalidation in a for loop, where the size of the feature set increases with each round from 1 to 70, where the first N features with the highest feature-importance score are selected.

The model in the training-set extension scenario was trained using the augmented data as unique entries that were added to the training sets during each stage of the model training.

## III. Results

The results for the four scenarios can be seen Table III. The best performing model was trained in the single-dropout scenario, while using features from the augmented data only. It reached a balanced accuracy score (after decision threshold adjustment) BA = 66 % (sensitivity – SN = 69 %, specificity – SP = 63 %), which is an improvement of 3 percentage points above the baseline model. The worst performing model was trained in the train-set extension scenario and reached BA (after decision threshold adjustment) of only 59 %.

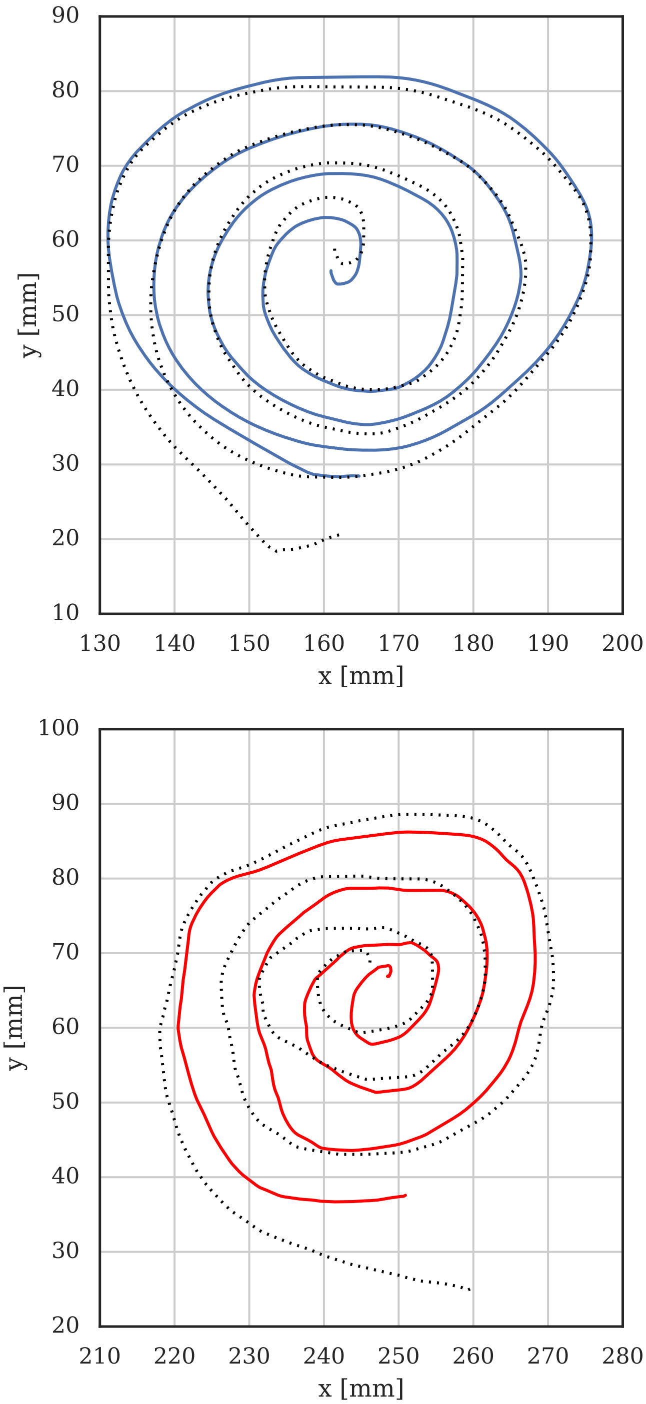



Fig. 6. Example of a spiral sample performed by a child in the HDC-T negative group (top in blue) and by a child in the HDC-T positive group (bottom in red); both samples augmented with the single-dropout approach. Both plots contain a visualisation of the unaltered input signal (dotted in black).

## IV. Discussion and Conclusion

While the best classification accuracy reached in the experiment was only 66 %, it was reached with a model based on the augmented dataset and it improved the performance above

TABLE III. Classification results for the XGBoost classification model.

| Scen | BA | SN | SP | $BA_p$ | $SN_p$ | $SP_p$ | N |
|---|---|---|---|---|---|---|---|
| BSL | 0.48 | 0.22 | 0.74 | 0.63 | 0.75 | 0.51 | 4 |
| **SD** | **0.47** | **0.03** | 0.92 | **0.66** | **0.69** | **0.63** | **36** |
| DD | 0.58 | 0.38 | 0.79 | 0.63 | 0.81 | 0.44 | 4 |
| TSE | 0.43 | 0.06 | 0.79 | 0.59 | 0.91 | 0.27 | 17 |

[1] Scen – scenario; BA – balanced accuracy; SN – sensitivity; SP – specificity; $BA_p$ – balanced accuracy after decision threshold adjustment; $SN_p$ – sensitivity after decision threshold adjustment; $SP_p$ – specificity after decision threshold adjustment; N – number of features used to train the model; BSL – baseline; SD – single-dropout; DD – double dropout; TSE – train-set extension;

the baseline by three percentage points. It needs to be said, that the only model, where the augmented samples were used to extend the dataset (train-set extension scenario), performed the worst, reaching accuracy 4 percentage points lower than the baseline.

These results suggest, that the approach of using an autoencoder to alter the input data can improve classification accuracy, however not in the way of introducing semi-synthetic samples into the database itself, but rather as a preprocessing subsystem of a larger assessment pipeline. While the improvement of the best performing classifier over the baseline scenario is rather marginal, the results show that the data altered with an LSTM autoencoder still retain the characteristics necessary to classify the samples based on the HDC–T scale. Even in the case of the DD scenario, where the overall performance matched the baseline, it is possible that some key features were enhanced with the approach, while others were degraded. These enhanced features could be identified using the XGBoost model's feature importance metrics. The next steps after this study then could include identifying the features that are positively affected by the LSTM preprocessing, finding an autoencoder architecture that maximizes this effect and then combining the best performing features from the altered data with the baseline featureset, either effectively extending it, or replacing the original ones.

The overall low performance of the TSE model might suggest, that the handwriting signal was far too altered up to a point where too many of the features extracted from it did no longer align with the HDC–T annotation. Combined with the fact, that the best performant model on the other hand was in fact trained from the augmented data alone, might suggest, that while most of the features were altered in the direction of misaligning with the baseline labels, some features were in fact enhanced and when used alone, the XGBoost classifier was able to identify them and build a more robust decision tree structure.

The classification performance results reached in this study are far from clinical significance. The current state-of-the-art methods reach much higher accuracies, however, the point of the experiment was to provide a proof-of-concept to a data augmentation or preprocessing method, which could lead to an increase in classification performance even in the state-of-the-art featurisation approaches.

This study has limitations mostly related to the extent of the feature set, which consists mostly of very rudimental metrics and statistics of the handwriting signal and thus might not reflect how such a preprocessing method would affect more complex features. The size of the dataset would in most studies be considered a limitation as well, however the point of the paper itself was to suggest ways to increase model effectivity in small-dataset situations.

## ACKNOWLEDGMENTS

The author would like to thank his supervisor doc. Ing. Jiří Mekyska, Ph.D. for his input and guidance.

## REFERENCES

[1] M. Thomas, A. Lenka, and P. Kumar Pal, "Handwriting analysis in parkinson's disease: current status and future directions," *Movement disorders clinical practice*, vol. 4, no. 6, pp. 806–818, 2017.

[2] J. Mekyska, M. Faundez-Zanuy, Z. Mzourek, Z. Galaz, Z. Smekal, and S. Rosenblum, "Identification and rating of developmental dysgraphia by handwriting analysis," *IEEE Transactions on Human-Machine Systems*, vol. 47, no. 2, pp. 235–248, 2016.

[3] Z. Galaz, P. Drotar, J. Mekyska, M. Gazda, J. Mucha, V. Zvoncak, Z. Smekal, M. Faundez-Zanuy, R. Castrillon, J. R. Orozco-Arroyave, S. Rapcsak, T. Kincses, L. Brabenec, and I. Rektorova, "Comparison of CNN-learned vs. handcrafted features for detection of Parkinson's disease dysgraphia in a multilingual dataset," *Frontiers in Neuroinformatics*, vol. 16, 2022. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fninf.2022.877139

[4] S. Rosenblum and G. Dror, "Identifying developmental dysgraphia characteristics utilizing handwriting classification methods," *IEEE Transactions on Human-Machine Systems*, vol. 47, no. 2, pp. 293–298, 2017.

[5] J. Mucha, J. Mekyska, Z. Galaz, M. Faundez-Zanuy, V. Zvoncak, K. Safarova, T. Urbanek, J. M. Havigerova, J. Bednarova, and Z. Smekal, "Analysis of various fractional order derivatives approaches in assessment of graphomotor difficulties," *IEEE Access*, vol. 8, pp. 218 234–218 244, 2020.

[6] M. Gavenciak, V. Zvoncak, J. Mekyska, K. Safarova, L. Cunek, T. Urbanek, J. M. Havigerova, J. Bednarova, Z. Galaz, and J. Mucha, "Exploring the contribution of isochrony-based features to computerized assessment of handwriting disabilities," in *2022 45th International Conference on Telecommunications and Signal Processing (TSP)*, 2022, pp. 355–359.

[7] J. Mekyska, K. Safarova, T. Urbanek, J. Bednarova, V. Zvoncak, J. M. Havigerova, L. Cunek, Z. Galaz, J. Mucha, C. Klauszova *et al.*, "Graphomotor and handwriting disabilities rating scale (ghdrs): Towards complex and objective assessment," *PsyArXiv*, 2023.

[8] S. Rosenblum and L. Gafni-Lachter, "Handwriting proficiency screening questionnaire for children (HPSQ–C): development, reliability, and validity," *The American Journal of Occupational Therapy*, vol. 69, no. 3, pp. 6 903 220 030p1–6 903 220 030p9, 2015.

[9] K. Šafárová, J. Mekyska, V. Zvončák, Z. Galáž, P. Francová, B. Čechová, B. Losenická, Z. Smékal, T. Urbánek, J. M. Havigerová *et al.*, "Psychometric properties of screening questionnaires for children with handwriting issues," *Frontiers in Psychology*, vol. 10, p. 2937, 2020.

[10] A. P. Association *et al.*, "Ethical principles of psychologists and code of conduct," *American psychologist*, vol. 57, no. 12, pp. 1060–1073, 2002.

# FPGA-Based Sound Acquisition Prototype from MEMS Microphone Array

1st Matej Liska
*Department of Telecommunications*
*Brno University of Technology*
Brno, Czech Republic
xliska19@vut.cz

2nd Jiří Tomešek
*Department of Telecommunications*
*Brno University of Technology*
Brno, Czech Republic
xtomes07@vut.cz

***Abstract*—This paper presents a comprehensive exploration of an Field Programmable Gate Array (FPGA) prototype designed for efficient data acquisition from Micro Electro Mechanical Systems (MEMS) microphones through a Time Division Multiplexed (TDM) interface. The primary focus lies in the seamless integration of software and hardware elements, showcasing the implementation of IP cores for controlling TDM peripherals, generating clock signals, and facilitating communication with the Future Technology Devices International (FTDI) chip. The paper underscores the flexibility of the device design, which can be seamlessly deployed across various FPGA boards, thereby mitigating dependency on specific hardware models. Leveraging a large number of GPIO pins on the FPGA board, the PCB design incorporates multiple TDM buses and efficient communication channels with the FTDI chip, streamlining data transfer to the PC. This work not only presents a functional prototype but also lays the groundwork for future advancements in FPGA-based sound acquisition systems, promising enhanced versatility and efficiency in audio signal processing applications.**

***Index Terms*—FPGA, TDM, Microphone Array, MEMS, FT232H, FTDI**

## I. Introduction

In the realm of audio signal processing, the pursuit of efficient and versatile sound acquisition systems remains paramount. The advent of Field-Programmable Gate Arrays has opened a promising avenue for creating sophisticated and adaptable solutions to audio acquisition problems. This introduction sets the stage for exploring the innovative domain of FPGA-based devices for sound acquisition from a microphone array. At the intersection of cutting-edge hardware and signal processing techniques, these FPGA prototypes offer a glimpse into the future of audio capture technology. By integrating FPGA technology with a microphone array, this prototype has the potential to revolutionize a variety of applications, from spatial audio recording to noise cancellation and more. This article provides a basic overview of an FPGA-based audio acquisition prototype, explores its underlying principles, and offers solutions for both software and hardware development in audio engineering.

Large microphone arrays, such as [1] , were built in the early 2000s to evaluate various speech enhancement algorithms aimed at conference rooms. Small microphone arrays that use MEMS technology are now ubiquitous in many consumer devices, such as laptops, webcams, or smartphones, as a speech recognition aid [2] due to improved recognition compared to a single omnidirectional microphone [3].

FPGA technology has changed the way microphone arrays are used for certain acoustic applications. The latest FPGA-based architectures are able to fully incorporate all computational tasks from creating clock modules to sampling data and interfacing with a PC to acquire captured data. However, many FPGA-based architectures face similar challenges when interconnecting and processing multiple data streams from microphone arrays.

A more detailed description of the related work, in which the acoustic sensor and the FPGA board were chosen, is described in the Section II. In Section III, the main principles of implementing IP cores for clocking structures, data sampling operations, and the state machine for controlling the entire TDM peripheral [6] are described. Additionally, the alternative transfer of high-frequency signals from the FPGA to the MATLAB workspace on the PC using FT232H converter is discussed.

## II. Related Work

The task of solving this work was to create a prototype that would be able to acquire and process digital microphone array data from the TDM interface into the MATLAB workspace on PC. The solution of the work follows the previous article *FPGA Design with TDM Module for Real-Time Sound Acquisition from MEMS Microphone Array* [4]. The designed device contained ICS-52000 MEMS microphones and used a TDM interface capable of transmitting data from up to 16 connected microphones on a single line. The selected Digilent Arty S7-25 [9] FPGA board offers numerous GPIO pins that can be utilized to create multiple TDM buses, enabling the implementation of audio data acquisition from microphone arrays. The possibility of connecting to MATLAB workspace for audio recording using the HDL Verifier [8] and Data Capture tool was also proposed. During the implementation of this proposed prototype, it was necessary to make changes to the design to achieve the desired functionality. These changes were made to optimize the FPGA, ensure sufficient speeds on the peripherals, and create an adequate number of connectors and selected peripherals to connect individual microphones to the FPGA.

## III. Software Section

Based on the previously submitted design [4], IP cores were implemented to generate clock signals for the TDM module. Additionally, a prototype of the final state machine (FSM) for controlling the TDM module according to the required number of connected microphones was created. Initially, the prototype allowed for the connection of 2, 4, 8, and 16 microphones by setting switches on the FPGA board. Later iterations expanded this capability to include control of the FPGA board from the user interface on the PC. However, control using hardware switches remained integral to the FSM algorithm - specified in section III-A.

Another aspect of the software prototype involved the acquisition and storage of data obtained from the TDM interface is specified in III-B. Processes, including the utilization of First In, First Out (FIFO) memory, were employed for this purpose.

The presented design also suggested transferring FPGA data to a PC using a USB peripheral and an IP core called HDL Verifier from MathWorks via the Data Capture Tool. However, this design couldn’t be implemented due to limited transmission capabilities. Instead, we opted to utilize GPIO peripheral with FT232H converter - section III-C.

### A. Implementation of Clock Signals

As described in [4], the main clock signal on our FPGA board has a frequency of 100 MHz. By utilizing the Clocking Wizard IP core, this clock was converted to a frequency of 98.304 MHz, which is divisible to generate timing frequencies to control the TDM interface of the MEMS microphone array. The Clocking Wizard IP core utilizes MMCM or PLL blocks for its functionality. In the case of our FPGA board, there are 3 MMCM blocks and 3 PLL blocks. Therefore, it is not necessary to use them for every clock signal generation; instead, clock dividers should be employed for multiples. The conversion of the 100 MHz clock signal utilized two MMCM blocks to create a cascade, resulting in the clock signal being converted to a frequency of 12.288 MHz. Subsequently, it was multiplied by a factor of 8 to obtain the resulting precise clock signal. However, this clock signal cannot be used externally, as it may lead to internal damage to the blocks and cause oscillations. Therefore, the resulting clock signal is fed into frequency dividers, which then convert it to output GPIO pins for TDM interfaces.

For the sake of simplicity, a module was created in which these two clock operations are connected. The module is shown in Fig. 1.

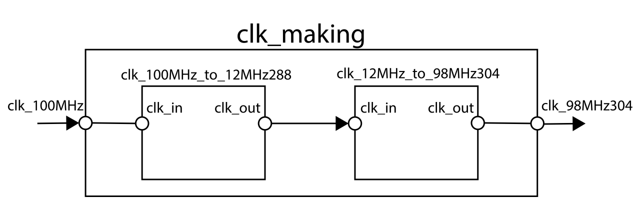



Fig. 1. MMCM Cascade Block.

The resulting SCK clock signals are generated using clock dividers based on the number of connected microphones and FSM settings. Additionally, the synchronization signal for frame switching - WS, which is set to a frequency of 48 kHz, is generated in the same manner.

Prototype block connection of the output interface for TDM line control is shown in Fig. 2.

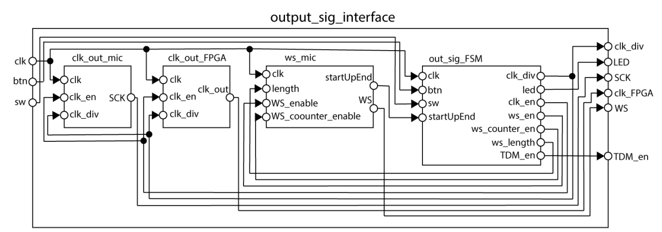



Fig. 2. TDM Output Interface.

### B. Data Acquisition and Storage

Each microphone has an allocated frame of a fixed length within which to transmit data. The frames are 32 bits long, with 24 bits representing data from the microphone and the remaining 8 bits filled with unspecified data [5]. The data are in two’s complement format and is transmitted on the bus in order from the most significant bit to the least significant one [5]. Therefore, during implementation, it is crucial to capture the first 24 bits, while the remaining 8 bits can be omitted as they contain unnecessary data.

The implementation of the TDM interface is divided into 5 processes. The first two processes simply wait for non-zero data from TDM peripheral, as determined in [5]. The third process subsequently loads 32-bit data into variables. The last two processes handle the removal of the unnecessary 8 bits, creating a 24-bit frame, and then sending it to a variable for insertion into the FIFO memory.

This code for final process is shown in Listing 1.

```
data_to_out : process (clk) begin
   if rising_edge (clk) then
      if valid_data = true then
         if counter = 25 and first = ’0’ then
            data_to_output <= data_reg;
            data_ready_s <= ’1’;
            first <= ’1’;
         elsif counter /= 25 then
            first <= ’0’;
         else
            data_ready_s <= ’0’;
         end if;
      end if;
   end if;
end process data_to_out;

data_out <= data_to_output;
data_ready <= data_ready_s;
```

Listing 1. Data-to-Out Process.

Following the creation of the 24-bit frame, the next step involves inserting the data into the FIFO memory block. The input port *full* serves as a signal indicating that the memory is full, for modules sending data to the input port *din*. The data output from the TDM module is connected to the *din* port. Data is written to the memory only if the *wr_en* input port is set to 1. Otherwise, the data will not be written to the memory and will be lost. Output data is available on the *dout* port, hence a data input of FTDI IP core is connected to this port.

The fifo memory block is shown in Fig. 3.


Fig. 3. FIFO Block Ports.

Within the module, there is a process synchronized with the rising edge of the clock signal. The condition ensuring the control of the *read_enable* signal is as follows: Reading from memory should occur when the FTDI core is ready, indicated by the *data_ready* signal being set to logic 1. Additionally, the FIFO must not be empty, indicated by the *fifo_empty* signal being set to logic 0. In this scenario, the *read_enable_s* signal is set to logic 1. Otherwise, it is set to logic 0. Finally, this signal is routed to the *read_enable* output port, which is then connected to the corresponding FIFO memory port.

This process is shown in Listing 2.

```
  signal read_enable_s : std_logic := '0';
begin
  process (clk) begin
    if rising_edge(clk) then
      if fifo_empty = '0' and data_ready = '1'
   then
        read_enable_s <= '1';
      else\Omega
        read_enable_s <= '0';
      end if;
    end if;
  end process;
  read_enable <= read_enable_s;
```

Listing 2. FIFO Process.

### C. FTDI IP core & MATLAB Integration

As previously mentioned, the attempt to implement the block connecting the FPGA board with MATLAB workspace on PC through the HDL Verifier IP core [8], as recommended by MathWorks and mentioned as design in [4], was unsuccessful. This was primarily due to the USB and JTAG interface's maximum transfer rate of 12 Mbps, which resulted in overflowing FIFO buffers on the FPGA board due to the low transfer speed.

An alternative solution was considered, involving the use of an integrated SPI to UART controller on the FPGA board, specifically the Arty S7-25. However, the manufacturer did not connect the output pins of the IC chip SPI for high-speed transfer (USB SS), thereby imposing further limitations on speed.

The final solution involved utilizing converter specifically FT232H chip from FTDI chip company. Initially, implementation and functionality were verified successfully using the UM232H-B [10] development board. This connection significantly increased the transmission speed up to 480 Mbps, which proved to be adequate for accommodating more than two TDM peripherals with 16 connected microphones each.

On the FPGA board, it was necessary to implement a module for data transfer via the GPIO peripheral to the FT232 chip. This resulted in the creation of the FTDI module, wherein, besides the input data, a 60 MHz clock from the oscillator on the FPGA board was provided. Subsequently, it was extended by a clock with a shift created using the Clocking Wizard IP core. The shifted clocks implemented waiting for the sending buffer to fill before outputting to the register itself.

The actual transmission of information via FTDI converter to the output of the USB port in the UART peripheral was achieved using the so-called three-state pins. This involves setting the pins to the input and output positions. Therefore, it was necessary to implement the logic for controlling this data transfer, from the PC to FPGA and from FPGA to PC, ensuring that two-way transmission does not occur simultaneously. This was accomplished by setting the output of the internal FPGA controller to the Hi-Z state when loading control data from the PC - MATLAB workspace, and setting the input of the controller to receive data from the pin. Conversely, during the reverse transfer, the input port is set to the Hi-Z state when sending data from the FPGA.

The Hi-Z state is set by inserting the "Z" attribute on the corresponding port. This process is shown on Listing 3 .

```
process(clk_ftdi) begin
  if rising_edge(clk_ftdi) then
    if pres_st = st_Read then
      if rxf = '0' then
        data <= (others => 'Z');
        data_from_pc_to_fpga_fifo <= data;
      end if;
    elsif pres_st = st_Write then
      if txe = '0' and write_en_next_s = '1'
      and write_en_s = '1' then
      data_from_pc_to_fpga_fifo <= (others => 'Z');
      data <= data_from_fpga_to_PC_from_fifo;
      end if;
    end if;
  end if;
end process;
```

Listing 3. Read/Write Data Process.

The final step in the software section involved implementing the library for receiving and transmitting data from the FTDI chip in the MATLAB environment to communicate with the FPGA. The *ftd2xx.h* library and driver [11] were used for this purpose. Separate codes were created for reading and writing data from the FPGA to the PC and vice versa. For individual installations and implementations, we followed the manufacturer's instructions and sought guidance from MATLAB forums.

## IV. Hardware Section

This section was based on the initial proposal mentioned in [4]. In this design, a connection issue arose at high frequencies due to the inadequate choice of connectors. To address this, the connectors were replaced with USB-C connections, which are better suited for high-frequency signals. Another challenge was the transmission from the microphone array to the FPGA control unit. Since we were dealing with high-frequency signals over longer distances, it was necessary to find a suitable connection solution. This issue is discussed more comprehensively in the IV-A section, which also presents the design of the printed circuit board (PCB). The section, IV-B, elaborates on the final design implemented using a PCB shield on the FPGA, thus establishing the connection between the microphone array and the FPGA control unit, including the FTDI chip with a USB connector for PC connection.

### A. LVDS Connection

When searching for a suitable connection between the microphone array and the control unit, it was essential to implement a device capable of transmitting signals over longer distances without signal degradation and with sufficient speed and minimal latency. Initially, the use of a buffer was proposed, but it resulted in information loss. Subsequently, options such as Ethernet cables were considered, but this was rejected due to size limitations of the RJ45 connector within the microphone array, and the transmission speed using RS232 was insufficient. Finally, the decision was made to implement transmission using Low Voltage Differential Signaling (LVDS) peripheral.

LVDS is a high-speed ($< 155.5$ Mbps), low-power interface standard that effectively addresses bottleneck issues across a wide range of applications [12]. To accommodate LVDS transmission, a printed circuit board with a selected chip was designed. In our case, the transmission is limited and requires a characteristic impedance of $100\,\Omega$. This can only be achieved by using $100\,\Omega$, preferably LVDS cables. Typically, LVDS cables involve a large number of differential pairs and large connectors. However, this was circumvented by using HDMI cables, which have the same impedance but offer the advantage of micro-HDMI connectors.

The final design of the prototype incorporated the connection of an LVDS transceiver. Based on the ICS-52000 TDM interface [5], the transceiver needed to transmit two signals on the FPGA side and receive one, and vice versa on the microphone array side. For this purpose, the SN65LVDS049 transceiver was selected, capable of reaching speeds of up to 400 Mbps [13].

### B. FPGA Shield

The resulting designed shield was intended to incorporate an FT232H chip, which facilitated the simulation transfer between the FPGA board and the MATLAB workspace on the PC, a process that was successfully verified. Additionally, the shield included necessary connector connections, as well as two LVDS transceivers SN65LVDS049 to enable the connection of two TDM microphone field interfaces simuntaniously.

This design was based on the development board PCB design [14], which was expanded with LVDS transceivers and adapted for power supply, optionally between the FPGA board or PC, using a USB connector. The resulting connection layout was aligned with the pins on the Arty S7-25 FPGA board.

The design of the PCB involved several components to ensure efficient and reliable operation. One element was the incorporation of a voltage regulator, such as the LM1117, to provide stable and regulated power supply to the various components on the board. The LM1117 regulator offers low dropout voltage and excellent line and load regulation [15], making it ideal for powering the FPGA board and associated peripherals with connected microphones.

In addition to the voltage regulator, the PCB featured two HDMI connectors specifically tailored for TDM peripherals. These connectors were strategically positioned to accommodate easy connection and communication between the microphone array and the FPGA board. HDMI connectors were chosen for their robustness, high-speed data transmission capabilities, and widespread availability. Another critical aspect of the PCB design was the selection of a suitable crystal oscillator. A high-quality crystal oscillator with precise frequency stability was essential for clock synchronization and data integrity. The chosen oscillator ensured accurate timing for signal processing and communication.

Furthermore, careful attention was paid to the layout and routing of traces on the PCB to minimize signal interference and ensure signal integrity. Proper grounding techniques and signal isolation were implemented to mitigate noise and distortion. This comprehensive approach ensured the successful integration and operation of the FPGA-based sound acquisition prototype.

## V. Conclusion

This paper builds upon the FPGA design outlined in [4], which acquires of data from MEMS microphones via a TDM interface.

The main contribution of this work lies in the effective implementation of the software and hardware aspects of the prototype. By describing IP cores for controlling TDM peripherals, generating clock signals, and communicating with the FTDI chip, this document presents a versatile device design applicable to a wide variety of FPGA boards, reducing dependence on specific hardware models. Utilizing the GPIO pins on the FPGA board, the PCB design was created with multiple TDM buses and communication channels with the FTDI chip, allowing data to be transferred to the computer.

The developed prototype provides a solution for capturing and processing sound waves from multiple microphone arrays in digital audio format. Importantly, it enables seamless integration with the MATLAB workspace, allowing for sound capture with minimal delay. This integration opens up possibilities for various subsequent processing of the captured data, enhancing the versatility and utility of the proposed solution.

The final system architecture of the prototype, which is shown in Fig. 4

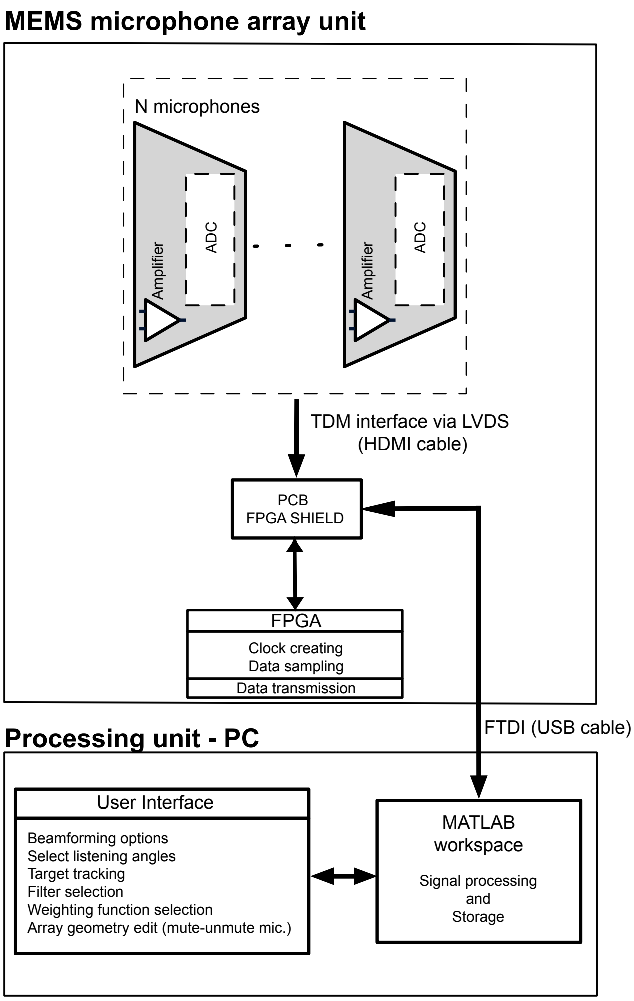



Fig. 4. Prototype architecture.

## REFERENCES


[1] E. Weinstein, K. Steele, A. Agarwal, J. Glass, "Loud: A 1020-Node Modular Microphone Array and Beamformer for Intelligent Computing Spaces," Technical Report, Massachusetts Institute of Technology: Cambridge, MA, USA, 2004.

[2] J. Tiete, F. Domínguez, B. da Silva, A. Touhafi, K. Steenhaut, "MEMS microphones for wireless applications. In Wireless MEMS Networks and Applications," in Elsevier: New York, USA, 2017, pp. 177–195.

[3] D.C. Moore, I.A. McCowan, "Microphone array speech recognition: Experiments on overlapping speech in meetings," in *Proceedings of the 2003 IEEE International Conference on (ICASSP'03) IEEE Acoustics, Speech, and Signal Processing*, Hong Kong, China, 6–10 April 2003, Vol. 5, p. 497.

[4] M. Liska, "FPGA Design with TDM Module for Real-Time Sound Acquisition from MEMS Microphone Array," in *Proceedings I of the 29th Conference STUDENT EEICT 2023.* Brno University of Technology, Faculty of Electrical Engineering and Communication, 2023, pp. 399-403.

[5] Invensense, Inc., "ICS-52000: Low-Noise Microphone with TDM Digital Output," DS-000121 data-sheet, rev. 1.3, April 2017.

[6] Cirrus Logic, "Time Division Multiplexed Audio Interface: A Tutorial," AN301 User Manual, rev. 1, Sept. 2006.

[7] Xilinx, Inc., "Clocking Wizard v6.0," PG065 LogiCORE IP Product Guide, April 2022, pp. 52-59.

[8] The MathWorks, Inc., "HDL Verifier™ Support Package for Xilinx® FPGA Boards," R2022b User's Guide, 2014-2022.

[9] Digilent, Inc., "Arty S7™," Reference Manual, rev. Oct. 2019. [Online]. Available: https://bit.ly/41OoL6r

[10] Future Technology Devices International Ltd., "FT232H Single Channel Hi-Speed UART to Multipurpose UART/FIFO IC," FT_000288 datasheet, rev 2.0, Nov. 2019. [Online]. Available: https://ftdichip.com/wp-content/uploads/2020/07/DS_FT232H.pdf

[11] Future Technology Devices International Ltd., "Instructions on Including the D2XX Driver in a VS Express Project," TN_153 Technical Note, rev. 1.1, June 2018. [Online]. Available: https://ftdichip.com/wp-content/uploads/2020/08/TN_153-Instructions-on-Including-the-D2XX-Driver-in-a-VS-Express-2013-Project.pdf

[12] National Semiconductor, "An Overview of LVDS Technology," Application Note 971, July 1998.

[13] Texas Instruments Inc., "SN65LVDS049 Dual-LVDS Differential Drivers and Receivers," SLLS575A data-sheet, revised in July 2015. [Online.] Available: https://bitly.ws/3fmK3

[14] Future Technology Devices International Ltd., "UM232H-B USB to Serial/Parallel Break-Out Module," FT_000721 data-sheet, rev. 1.0, Oct. 2010.

[15] Texas Instruments Inc., "LM1117 800-mA, Low-Dropout Linear Regulator," SNOS412Q data-sheet, revised in Jan. 2023.

# Application of Auditory Masking based Speech Denoising in Automotive Environments

Jan Malucha
*Department of Radio Electronics*
*Brno University of Technology*
Brno, Czech Republic
203286@vutbr.cz

***Abstract*—This paper presents an application experiment of denoising speech in the automotive field. An algorithm based on the auditory masking phenomenon was programmed and deployed for this purpose. Synthetic composite recordings of speech and vehice interior noise were used for three types of vehicles equipped with internal combustion engines - truck, jeep and sports car. The final results after denoising process are evaluated by four speech quality metrics. The trend of quality improvement depending on the SNR of the input noisy signal is examined. A possibility of using denoised speech signals for further speech features analysis is briefly discussed.**

***Index Terms*—speech enhancement, auditory masking, automotive, denoising**

## I. Introduction

Vibration and noise are undoubtedly among the most critical variables in the development and optimization of automobiles. Modern passenger cars nowadays achieve very low noise emissions and with the advent of electric mobility, vehicles moving at lower speeds are inaudible to the extent of incorporating loudspeakers and synthetic sound for the sake of traffic safety.

Despite the fact that the future of transport in the light of electric vehicles may appear very quiet, imperfections in the electromobility system, technology and infrastructure still leave the door open for many applications of internal combustion engines, especially in sectors requiring high vehicle power; acoustic comfort in their design must logically give way to other functional parameters, whether it is high speed, load capacity or reliability in the most demanding operating conditions. Namely, we can mention the defence or construction industry - military and large trucks can still hardly rely on electric drives. Also, much of the popularity of automobile racing still revolves around vehicles with internal combustion engines up to this day. The stereotypical image that has developed around these industries is then always accompanied by high noise emissions, as evidenced by the numerous well-publicised anti-noise protests against military bases or motor racing [1].

One of the most significant specifics of vehicle operation in these conditions is the degradation of communication due to their noise, both within and outside the crew of a single vehicle. The ability of a military vehicle crew to communicate actively is a matter of life and death. Truck and lorry drivers also often use radios to communicate with each other, as do motorcyclists and car racers for navigation.

Noise inside vehicles enters the communication channel of radios or intercoms through the built-in microphone and, depending on the intensity, degrades the transmitted speech signal. At the same time, as it degrades the listening intelligibility and quality, it also degrades the comfort of the crew, which can have a significant negative impact on both human performance and information comprehension and thus can lead to adverse consequences. Although special anti-noise microphones (called gradient microphones) have been developed that are able to provide favourable SNR, the problem still largely persists. This is also evidenced by the relatively long and extensive research of means for digital noise cancellation and speech enhancement at the signal level.

There is a wide range of adaptive algorithms and methods for noise suppression - the concept of optimal Wiener filtering [2], methods based on statistical noise estimation, Euclidean subspace methods [3]. In recent years, machine learning methods that formulate the speech enhancement problem as a classification task using a data-driven approach [4] [5] have become widespread. In the early 1980s, spectral subtraction algorithms began to be developed specifically for military applications [6]. To date, there have been more publications on this family of algorithms and their modifications than on any other family [3]. The algorithms are simple, robust and reliable. There are a number of variants and modifications, ranging from the oldest variants including e.g. nonlinear spectral subtraction [7], optimal MMSE estimation of short-term spectral amplitude [8], Berouti's algorithm [9], to newer methods such as the adaptive spectral subtraction method based on the auditory masking principle [10].

In this paper, we present the results of an application of a noise suppression and speech enhancement algorithm based on auditory masking. Tests were conducted in the relevant environments of a truck, a military off-road vehicle and a racing car interiors. It aims to offer the user a denoised listening experience of voice communication while maintaining a good compromise between quality and intelligibility.

Research described in this paper was supported by Brno University of Technology, specific research grant number FEKT-S-23-8191.

## II. Vehicle Noises

The authors of [11] and [12] give a detailed list of possible sources of vibration in the car mechanism. Namely, they list the combustion system, crank train system (piston, pin, rod, crankshaft, crankshaft damper, flywheel / flexplate), camdrive system, valvetrain system (impulsive noise at idle, engine shanking forces at idle, engine noise and vibration under acceleration), fuel delivery system, air flow system and boost system. Each of these systems contributes to the overall sound emission of the vehicle and may be dependent on driving conditions.

For our analysis, we use interior noise recordings of three vehicles traveling at approximately uniform speeds on an open asphalt road, namely a truck, a jeep, and a sports car. By comparing the spectrograms of the vehicle noise recordings against the spectrogram of speech in Fig. 1, we can estimate the frequency bands in which interference with speech occurs.

The spectrograms were calculated by algorithms integrated in Audacity using a Hann window of 2048 samples; the y-axis is calibrated in Mel scale and labeled with the corresponding frequency values in Hz. The duration of each recording is approximately 1.8 seconds at a sampling rate of 8000 Hz.

Speech is characterized by a typical spectral waveform dependent on the instantaneous phonetic event. The Czech words "šifra, šálek" were selected for the occurrence of specific consonants and vowels, including fricatives. Thus, it is possible to observe both areas of noise character at the beginnings of words and voiced areas characterized by distinctive formant pattern.

The spectral characteristic of the noise of the truck shows a marked dominance at frequencies up to 500 Hz; not much energy is concentrated in the region above 1000 Hz. In particular, the tonal character of the noise is typical, which can be seen from the clear continuous lines in the spectrogram around 80 and 500 Hz. Similarly, in the jeep, tonal character can be observed to some extent in the continuous lines around 35 and 120 Hz, although it is weaker compared to the character of the truck. In addition, energy distribution can be observed up to 2400 Hz. The sports car noise manifests itself in a broadband manner throughout the region up to 4000 Hz. The energy distribution is very pronounced up to 500 Hz and the tonal character is not apparent.

Measurements [13] have shown that the noise of a tracked or wheeled vehicles can be considered stationary at a uniform speed over an unchanging surface. Such conditions are not to be expected in real deployment for all the time; when the terrain undulates, the surface changes (dry or wet dirt, concrete), during acceleration and deceleration, and therefore when the engine load changes, noise of a highly non-stationary nature is produced, dominating at lower frequencies. A stationarity analysis of the three vehicle interior noise recordings plotted in spectrograms was done in MATLAB and is shown in Fig. 2. Dispersion was calculated from windows of length n starting at the first sample and settling of resulting dispersion curve indicates the gradual attainment of stationarity.

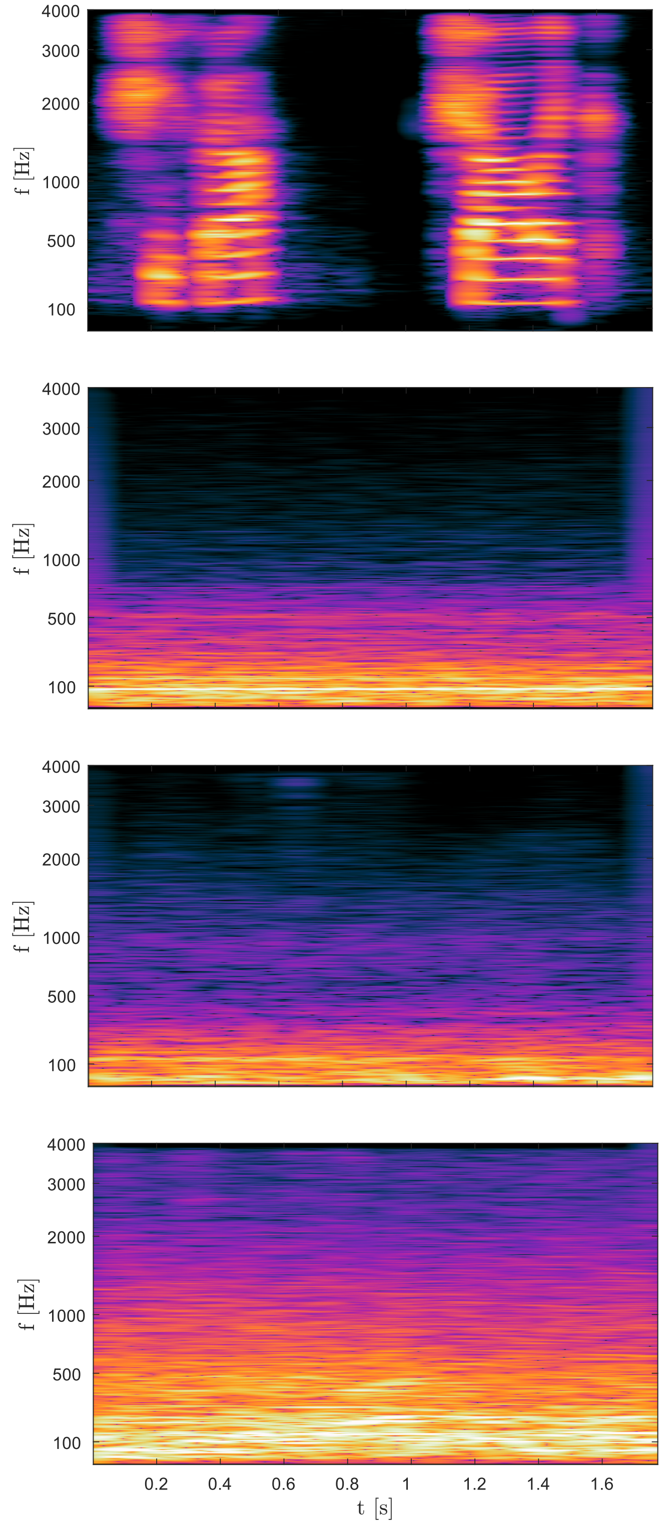



Fig. 1. Spectrogram comparison from top: audio recording of Czech words "šifra, šálek", truck noise, jeep noise, sports car noise.

Although for truck and jeep we can see a gradual stabilisation and the achievement of stationarity, sports car continues to fluctuate. This may be due to the more pronounced sounds of slight ride fluctuation. At the same time, it should be said

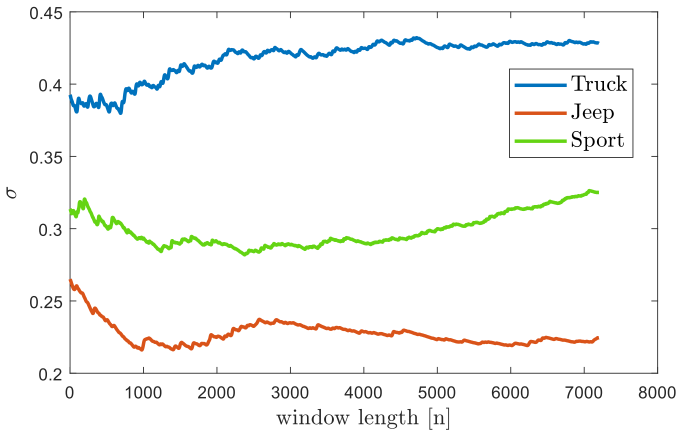



Fig. 2. Comparison of stationarity for three vehicle types.

that stationarity of driving noise is not common, especially for vehicles intended for demanding conditions. In case of interference between car noises and speech, there will inevitably be a deterioration in quality and intelligibility. If we were to assume persistent stationarity, a simple non-adaptive filter would suffice for speech denoising purposes. However, in real conditions, adaptive denoising methods must be used.

## III. The Algorithm

The operational requirements for a speech enhancement algorithm are computational efficiency, maximum robustness and simplicity; the algorithm must also be able to offer a good compromise between the level of denoising and maintaining the quality and intelligibility, namely not distorting the speech contours. A method from the **spectral subtraction** family based on the **principle of auditory masking** was chosen and programmed. Based on objective metrics, this particular algorithm proved to be suitable during military intercom development in cooperation with one of our industrial partners.

The principle is described in great detail in [10] or [14], for the sake of clarity we give a very brief characterization. An additive speech and noise signal has penetrated the communication channel through the microphone:

$$y[n] = x[n] + d[n], \tag{1}$$

where $x[n]$ is the undistorted speech signal, $d[n]$ is the ambient noise, and $y[n]$ speech signal with additive ambient noise. At its core, the general spectral subtraction method is always based on obtaining the desired speech signal estimate by subtracting the noise estimate from the composite signal:

$$\hat{X}[f] = (|Y[f]| - |\hat{D}[f]|)e^{j\theta_y[f]}, \tag{2}$$

where $\hat{X}[f]$ is the **estimate** of the speech signal spectrum, $\hat{D}[f]$ is the **estimate** of the ambient noise spectrum usually obtained in non-speech sound segments identified by a VAD (Voice Activity Detector), and $Y[f]$ is the speech signal with additive ambient noise (capital letters indicate spectral plane). By transforming the expression from the subtraction operation to the multiplication operation, we can obtain an expression that formulates the problem as filtering the spectrum of a composite signal by a time-varying adaptive filter, whose gain function depends on the estimate of the instantaneous noise:

$$\hat{X}[f] = G[f]|Y[f]|, \tag{3}$$

where $G[f]$ is the gain function or the attenuation curve. A typical problem with spectral subtraction algorithms is the so-called musical noise, which arises because the operation of subtracting the estimated noise spectrum $D[f]$ from the noisy signal $Y[f]$ may result in isolated locations along the frequency axis where $D[f] > Y[f]$, and the resulting $X[f]$ thus reaches negative values at these locations; when the resulting function is then inverted to an absolute value, isolated peaks appear in the $X[f]$ spectrum. The listening effect is then a very rapidly fluctuating "melodic" background or residual/musical noise.

The development of the spectral subtraction family of algorithms has for a long time been primarily concerned with minimizing this problem. A generalization of the problem was proposed by Berouti [9]; his approach is to introduce two tunable parameters, $\alpha$ and $\beta$, by which the correct trade-off between additive and residual noise can be adaptively adjusted:

$$G[f] = \begin{cases} (1 - \alpha(\frac{|\hat{D}[f]|}{|Y[f]|})^{\gamma_1})^{\gamma_2} & \text{for } (\frac{|\hat{D}[f]|}{|Y[f]|})^{\gamma_1} < \frac{1}{\alpha+\beta} \\ (\beta(\frac{|\hat{D}[f]|}{|Y[f]|})^{\gamma_1})^{\gamma_2} & \text{otherwise} \end{cases} \tag{4}$$

where $\gamma_1$ and $\gamma_2$ are sharpness exponents. Variants of this algorithm family differ precisely in their approach to setting these two parameters. In the case of this one, the $\alpha$ and $\beta$ parameters are given by auditory masking thresholds.

$$\begin{aligned} \alpha[f] &= F_a[\alpha_{min}, \alpha_{max}, T[f]], \\ \beta[f] &= F_b[\beta_{min}, \beta_{max}, T[f]], \end{aligned} \tag{5}$$

where $T[f]$ is the masking threshold, $\alpha_{min}$, $\alpha_{max}$, $\beta_{min}$ and $\beta_{max}$ are boundary conditions. $F_a$ and $F_b$ are special functions based on knowledge of the biological processes of auditory perception. A description is beyond the scope of this paper, it can be found in detail in [3]. The method workflow is very briefly illustrated by the block diagram in Fig. 3.

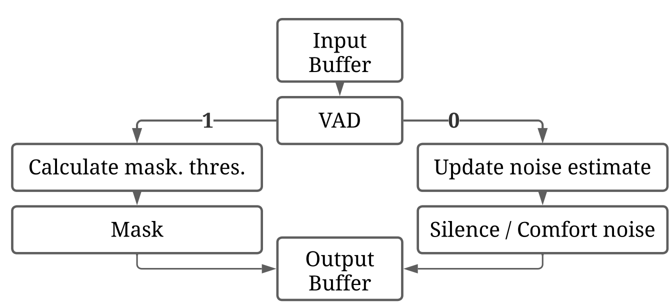



Fig. 3. The method flowchart.

## IV. Experiments and Results

The algorithm has been tested on synthetic composite recordings of speech and vehicle noise, for different SNRs. These were additively composed at selected SNR in MATLAB environment. Three types of interior noise corresponding to each of the vehicle platforms mentioned have been applied - truck, jeep and sports car. The raw noise recordings were taken from database acquired by the authors of [15] (the database was originally intended for training machine learning systems in vehicle type recognition). Furthermore, a short recording of the Czech word ”peníze” (En. ”money”) was extracted from a professionally produced audiobook read by a female voice. As simplicity and speed are required for denoising, a working sample rate of 8000 Hz was chosen to simulate the worst case scenario (in relation to speech quality degradation due to low sampling rate) - all recordings were resampled to this value before the addition. In the next step, the recordings were filtered by the algorithm.

Fig. 4 shows a brief illustrative example of the denoising result for a mixture of speech and truck noise. The top of the three subplots displays the clean speech recording, the middle subplot the noisy speech at SNR equal to 1 (absolute value) and the bottom subplot the denoised recording. A visual inspection will reveal significant changes in the denoised signal shape, which will be reflected in a degradation of quality, but not a loss of intelligibility, when listening.

The quality of the denoising process under different conditions is usually assessed by special metrics. The ones most suitable for our purpose are the so-called objective, i.e. computational methods for assessing the quality of speech recordings, developed as an alternative to subjective methods based on questionnaires of a large number of competent respondents [16]. In this case, the metrics chosen were Segmental SNR (SegSNR), Weighted-spectral slope metric (WSS), Perceptual Evaluation of Speech Quality (PESQ) and the composite method Overall quality score (OQS); all the methods mentioned are described in great detail by Loizou [3]. The results for the vehicles are summarised in the following tables.

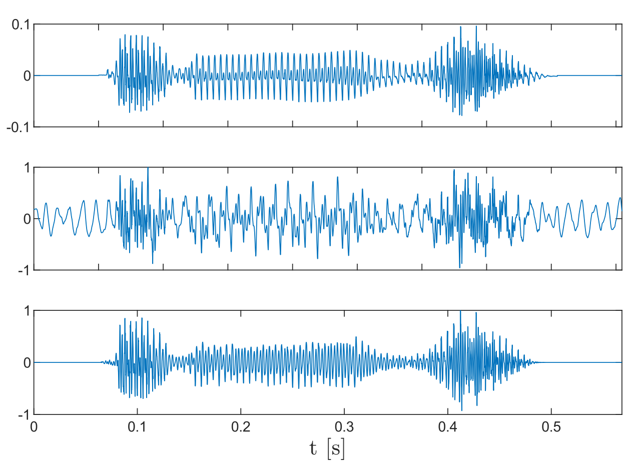



Fig. 4. Comparison of clean, noisy and denoised speech signals under the influence of truck noise.

TABLE I
Quality of Denoised Speech Metrics for Truck

| *SNR* | **OQS** | **SegSNR** | **WSS** | **PESQ** |
|---|---|---|---|---|
| **worst** | 1.00 | -3.09 | 167.32 | 0.88 |
| **0.5** | 2.46 | 5.95 | 66.13 | 2.09 |
| **1** | 3.77 | 8.37 | 32.01 | 3.24 |
| **2** | 3.89 | 10.41 | 21.91 | 3.30 |
| **4** | 4.12 | 11.93 | 13.59 | 3.47 |
| **10** | 5.00 | 19.08 | 1.07 | 4.37 |
| **ideal** | 5.00 | 35.00 | 0.00 | 4.50 |

TABLE II
Quality of Denoised Speech Metrics for Jeep

| *SNR* | **OQS** | **SegSNR** | **WSS** | **PESQ** |
|---|---|---|---|---|
| **worst** | 1.00 | -0.97 | 164.46 | 0.45 |
| **0.5** | 2.46 | 5.81 | 81.13 | 2.33 |
| **1** | 3.85 | 7.17 | 50.28 | 3.55 |
| **2** | 3.86 | 9.80 | 25.88 | 3.29 |
| **4** | 4.08 | 11.73 | 14.75 | 3.40 |
| **10** | 5.00 | 18.86 | 1.33 | 4.35 |
| **ideal** | 5.00 | 35.00 | 0.00 | 4.50 |

TABLE III
Quality of Denoised Speech Metrics for Sports Car

| *SNR* | **OQS** | **SegSNR** | **WSS** | **PESQ** |
|---|---|---|---|---|
| **worst** | 1.00 | -0.02 | 140.23 | 0.99 |
| **0.5** | 1.05 | 1.34 | 130.07 | 1.13 |
| **1** | 1.47 | 3.67 | 111.23 | 1.66 |
| **2** | 3.22 | 6.51 | 64.77 | 2.98 |
| **4** | 4.11 | 11.34 | 16.03 | 3.41 |
| **10** | 4.95 | 17.06 | 2.96 | 4.20 |
| **ideal** | 5.00 | 35.00 | 0.00 | 4.50 |

In practice, objective metrics are the output of computational algorithms comparing a reference clean speech signal with a denoised one. To clarify the scale of a given metric, values for the worst case (comparing the reference signal against the raw noise) and the best case (comparing the reference signal against itself) of denoising were included in the tables. Based on this information, a gradual convergence of the denoising quality towards the ideal state for increasing SNR can be observed.

The general phenomenon for all metrics is a very similar trend of improving de-noised signal quality with increasing SNR for truck and jeep. For SNR values equal to 10, the metrics are then very close to the ideal state with their values, which is visible especially for OQS, WSS and PESQ.

Based on the results, it appears that the non-stationarity of the sports car noise manifests itself in a worse de-noising quality, especially for SNR of the input noisy signal up to 4 (absolute value), from where the de-noising quality is approximately on par with other vehicles - thus non-stationarity does not play a significant role for SNR higher than 4. The broadband noise of a sports car may also play a role here, which then interferes intensively with the higher frequency bands of speech (this is not the case for truck and jeep). A comparison of algorithm performance against other denoising algorithms using both objective and subjective metrics can be found, e.g., in [3].

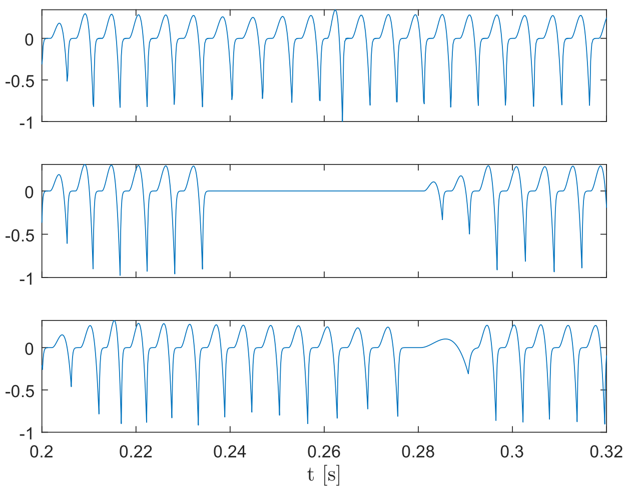



Fig. 5. Comparison of clean, noisy and denoised glottal pulse train derivative.

Voice recordings from military vehicle intercoms could further serve research purposes in the field of glottal pulse extraction and subsequent mining of specific information from speech, such as stress and emotional state. For example, acute stress can be investigated both from the glottal pulse themselves [17] and from their first derivative [18]. Analysis of glottal pulses has great potential for real-time monitoring of soldiers in highly demanding tasks. However, the distortion caused by the noise inside these vehicles as well as by the subsequent denoising must be taken into account. Fig. 5 compares the derivative of glottal pulses train extracted from a segment of clean speech (top), its version noised by jeep sound (middle) for SNR equal to 1 and its denoised version (bottom). At this level, a significant distortion of the pulse train can be seen in the denoised version. Therefore, in order to provide conclusive data for stress detection, it is necessary to ensure good SNR values, possibly by using anti-noise microphones.

## V. Conclusion

In this paper, we presented an application experiment of denoising speech signals distorted by vehicle interior noises.

The denoising process eliminates ambient noise, but depending on the SNR, there is more or less distortion of the signal, which is manifested by a reduction in listening quality (without any loss of intelligibility). The specific degree of denoised speech distortion was evaluated by four objective / computational metrics of speech quality. For the case of truck and jeep, whose noise recordings are stationary, the trend of denoised speech quality is very similar, and for an SNR equal to 10, the resulting quality is close to ideal. For the case of non-stationary sports car noise, the trend of quality increase depending on SNR is quite different - for SNR lower than 4, the denoising quality against truck and jeep is lower, but above this value the trend is similar to stationary noises. It follows that from certain values of the signal-to-noise ratio, stationarity does not play a significant role. Tests have shown that the distortion of the recording after denoising can have a major effect on glottal pulses, which are a well-known feature for mining information from speech.

In future work, we will first test and optimize the algorithm for the presence of non-stationary noise [19] that can occur in vehicles. Our goal is to find an efficient denoising method which not only improves the intelligibility of noisy speech, but at the same time does not distort the shape of extracted glottal pulses. Such an algorithm will be used in our further research as a pre-processing of the speech signal in recognition of stress based on glottal pulses. To the best of our knowledge, no publication has yet addressed the effect of denoising on the waveform of extracted glottal pulses.

## References


[1] ČT24 (2023, November). Hlukem proti hluku. Lidé z Mostu demonstrovali proti tamnímu Autodromu. [Online]. Available: https://ct24.ceskatelevize.cz/clanek/regiony/hlukem-proti-hluku-lide-z-mostu-demonstrovali-proti-tamnimu-autodromu-562

[2] J. Chen, J. Benesty, Y. Huang, and S. Doclo, "New insights into the noise reduction Wiener filter," in *IEEE Transactions on audio, speech, and language processing*, vol. 14, no. 4, pp. 1218-1234, 2006.

[3] P. C. Loizou, *Speech Enhancement: Theory and Practice*. CRC press, 2013.

[4] D. Wang, J. Chen, "Supervised speech separation based on deep learning: An overview," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 26, no. 10, pp. 1702-1726, 2018.

[5] Y. Wang, D. Wang, "Cocktail party processing via structured prediction," *Advances in Neural Information Processing Systems 25*, 2012.

[6] S. Boll, "Suppression of acoustic noise in speech using spectral subtraction," *IEEE Transactions on acoustics, speech, and signal processing*, vol. 27, no. 2, pp. 113-120, 1979.

[7] P. Lockwood, J. Boudy, "Experiments with a nonlinear spectral subtractor (NSS), hidden Markov models and the projection, for robust speech recognition in cars," *Speech communication*, vol. 11, no. 2-3, pp. 215-228, 1992.

[8] Y. Ephraim, D. Malah, "Speech enhancement using a minimum-mean square error short-time spectral amplitude estimator," *IEEE Transactions on acoustics, speech, and signal processing*, vol. 32, no. 6, pp. 1109-1121, 1984.

[9] M. Berouti, et al., "Enhancement of speech corrupted by acoustic noise," *in ICASSP'79. IEEE International Conference on Acoustics, Speech, and Signal Processing*, IEEE, vol. 4, 1979, pp. 208-211.

[10] N. Virag, "Single channel speech enhancement based on masking properties of the human auditory system," *IEEE Transactions on speech and audio processing*, vol. 7, no. 2, pp. 126-137, 1999.

[11] M. S. Qatu, "Recent research on vehicle noise and vibration," *International Journal of Vehicle Noise and Vibration*, vol. 8, no. 4, pp. 289-301, 2012.

[12] M. S. Qatu, M. K. Abdelhamid, J. Pang, and G. Sheng, "Overview of automotive noise and vibration," *International Journal of Vehicle Noise and Vibration*, vol. 5, no. 1-2, pp. 1-35, 2009.

[13] J. Hovorka, "Kombinované vícepásmové adaptivní zvýraznění řeči," Ph.D. Dissertation, Brno University of Technology, 2016.

[14] B. Scharf, "Fundamentals of auditory masking," *Audiology*, vol. 10, no. 1, pp. 30-40, 1971.

[15] E. Akbal, T. Tuncer, and S. Dogan, "Vehicle interior sound classification based on local quintet magnitude pattern and iterative neighborhood component analysis," *Applied Artificial Intelligence*, vol. 36, no. 1, 2022.

[16] Y. Hu, P. C. Loizou, "Subjective comparison of speech enhancement algorithms," *IEEE International Conference on Acoustics Speech and Signal Processing Proceedings*, Toulouse, 2006, pp. 1153-1156.

[17] M. Stanek, M. Sigmund, "Psychological stress detection in speech using return-to-opening phase ratios in glottis," *Elektronika ir Elektrotechnika*, vol. 21, no. 5, pp. 59-63, 2015.

[18] M. Sigmund, A. Prokes, and Z. Brabec, "Statistical analysis of glottal pulses in speech under psychological stress," *16th European Signal Processing Conference*, Lausanne, 2008, pp. 1-5.

[19] P. Zelinka, M. Sigmund, "Hierarchical classification tree modeling of nonstationary noise for robust speech recognition," *Information Technology and Control*, vol. 39, no. 3, pp. 202-210, 2010.

# Deep prior audio compression

Michal Švento
*Dept. of Telecommunications, FEEC*
*Brno University of Technology*
Brno, Czech Republic
michal.svento@vut.cz

Peter Balušík
*Dept. of Telecommunications, FEEC*
*Brno University of Technology*
Brno, Czech Republic
230531@vut.cz

***Abstract*—Audio compression is still an up-to-date topic because the demand for big data streams is rapidly increasing. Deep learning has brought up new algorithms that decrease bitrates with good perception quality. The novel approach in generative artificial intelligence is to produce new data from prior stored in network parameters, called a deep prior. The deep audio prior framework shows its success in various tasks such as inpainting, declipping, and bandwidth extension, but it has not been tested for compression. In this paper, we test this method with a pre-built network for inpainting. Our idea of compression is based on reducing the number of time-frequency coefficients in the spectrogram while allowing the reconstruction of the original signal with high quality.**

***Index Terms*—audio processing, deep learning, deep audio prior, compression**

## I. Introduction

The problem of signal compression is as old as the communication itself [1]. Despite the increased speed of the transmission channels, user demands also follow this trend. Currently in audio, there is a demand for real-time applications, such as music shows that take place at multiple locations simultaneously that require a fast data stream and are especially sensitive to latency. On the contrary, some applications need better sound quality than low latency. This makes the topic of audio compression attractive and opens up various research directions for improvement.

The operation of a generic codec, which stands for coder–decoder, can be categorised into two key parts. The *encoder* compresses the original data into a smaller size while maintaining the quality of perception as much as possible. The fundamental component of the encoder is *quantiser* where compressed data are effectively encoded for transmission. The *decoder* reconstructs the received signal with minimal deviation from the original unprocessed signal.

The key properties of an audio codec are decoded quality (ability to restore compressed audio with minimal loss), compression efficiency (low bitrate), generation speed, and latency (minimum time to initialise the codec) [2]. Audio codecs are split into two variants: waveform codecs and parametric codecs. The parametric codecs saves storage in underlying features such as the time-frequency (TF) characteristic. Popular TF representations are the Short-time Fourier Transform (STFT), Constant-Q Transform (CQT), or Modified Discrete Cosine Transform (MDCT). These are used by most successful coders. The most popular audio codec is the MPEG layer 3 (mp3) as a part of the video codec [3]. Its efficiency is based mostly on psychoacoustical principles. The ear tract is a mechanical system and is imperfect. The codec suppresses information below the absolute threshold of hearing and exploits frequency masking. The subsequent technologies, Advanced Audio Codec (AAC) [4] and the open-source Vorbis [5], enhance the concepts of the mp3 and result in more efficient storage utilization. These codecs are significant for low bitrate, but the decoded quality is generally worse than waveform codecs.

In contrast, waveform codecs manipulate with raw audio signal. The simplest idea is to downsample waveform data and then interpolate the dropped samples. The first digital codecs were built upon this idea. Algorithms for coder are fast and this technique is still present [6]. The advent of deep learning has also affected the area of codecs. For example, SoundStream [7], Encodec [8] as end-to-end waveform codecs. They used residual vector quantisation to reduce bitrate, while utilising the loss of the HiFi-GAN vocoder [9] to ensure the fidelity of the decoded audio. The novel state-of-the-art combines the strengths of the parametric and waveform codec, APCodec [2].

Deep neural networks also open up quite different ways of approaching problems in signal processing and restoration. Instead of an end-to-end or parametric approach in audio, we can optimise the prior of data hidden in network layers. These methods have shown success in image restoration tasks, e.g. Deep Image Prior – DIP [10]. Additional applications were developed from DIP, including variations with audio, e.g. the Deep audio prior [11] and Deep prior audio inpainting (DPAI) [12]. They have shown success for a variety of different restoration problems. Compression is a task that has not yet been tested with deep prior networks to our best knowledge. The similarity of data (de)compression and inpainting problems motivates the use of DPAI as a decoder of subsampled, i.e. compressed, data [6].

This paper is organised as follows. Section II describes the tools used to solve inpainting problems with the deep prior neural network. Section III contains the setup of the

The work was supported by the Czech Science Foundation (GAČR) Project No. 23-07294S. The authors are grateful to NVIDIA for donation of the Titan XP graphic card, which has been used in this research. PhD study of Michal Švento is supervised by Pavel Rajmic.

experiment and the choice of TF masks. Sections IV and V summarise the results and future work.

## II. Problem formulation

Assume a single channel audio signal $\mathbf{x} \in \mathbb{R}^L$ sampled at sampled at the sampling rate of $f_s$ (Hz) where $L$ is the number of samples. The TF characteristic of the signal is computed using the STFT. Conversely, given the TF characteristic, the time-domain signal can be synthesised using the inverse STFT (iSTFT). Both operations are linear and can be represented as the analysis operator $\mathcal{A}$ (STFT) and the synthesis operator $\mathcal{D}$ (iSTFT). The analysis operator $\mathcal{A} : \mathbb{R}^L \to \mathbb{C}^{M \times N}$ transforms the signal $\mathbf{x}$ into a complex matrix (spectrogram)

$$\mathbf{X} = \mathcal{A}(\mathbf{x}) \in \mathbb{C}^{M \times N}. \tag{1}$$

The number $M$ expresses the number of frequency bins, and $N$ is the number of time frames. The synthesis operator is a mapping $\mathcal{D} : \mathbb{C}^{M \times N} \to \mathbb{R}^L$. With an appropriate setting for the pair of $\mathcal{A}$ and $\mathcal{D}$ the composition of the analysis and synthesis is the identity [13], i.e. $\mathbf{x} = \mathcal{D}(\mathcal{A}(\mathbf{x}))$ for any signal in the time domain $\mathbf{x}$.

For the task of audio inpainting, the damage is usually considered in the time domain. However, there is an alternative approach that simulates loss in the TF domain [12], [14]. For a clean spectrogram $\mathbf{X} \in \mathbb{C}^{M \times N}$, the damaged spectrogram $\widetilde{\mathbf{X}}$ is obtained according to

$$\widetilde{\mathbf{X}} = \mathbf{X} \odot \mathbf{M}, \tag{2}$$

where $\mathbf{M} \in \mathbb{R}^{M \times N}$ is mask matrix and $\odot$ is Hadamard (element-wise) product. The matrix $\mathbf{M}$ is a binary mask with value one when the corresponding complex coefficient is preserved, and zero if its discarded. The authors of [12] used only masks for the inpainting task where the full columns are removed. Our implementation extends these masks beyond these presumptions, and we remove coefficients from the whole coefficient matrix. The exact process for the generation of masks will be described in Section III-B.

### A. Deep Prior

We would like to use the Deep Prior neural network for the audio compression scenario. Most approaches train the network and get the knowledge about parameter distribution from learning. However, a significant amount of data knowledge about the sound is contained within the network architecture even without performing any training of the model parameters. This approach uses a completely untrained generative convolutional neural network (CNN).

We can write inverse audio problem as an energy minimisation task:

$$\widetilde{\mathbf{X}}^* = \arg\min_{\mathbf{X}} E(\mathbf{X} \odot \mathbf{M}; \widetilde{\mathbf{X}}) + R(\mathbf{X})\,, \tag{3}$$

where $E(\cdot)$ is task-dependent data fidelity term. For inpainting (compression) tasks, it is mostly the Mean Square Error (MSE) or $\ell_1$ norm. The fidelity part ensures the solution will not deflect from the reliable part of the spectrogram. The second part $R(\cdot)$ is the regularisation term. It captures generic prior in the spectrogram, and it is usually hard to describe. In deep prior approach, the regularizer $R(\cdot)$ is replaced by the neural network $f_\theta(\cdot)$ with parameters $\theta$ which captures the implicit prior [10]:

$$\theta^* = \arg\min_{\theta} E(f_\theta(\mathbf{Z}); \widetilde{\mathbf{X}}), \quad \mathbf{X}^* = f_{\theta^*}(\mathbf{Z})\,, \tag{4}$$

where $\mathbf{Z} \sim \mathcal{N}(0, 1)$ is an initial random noise realization. The minimiser $\theta^*$ is obtained using well-known optimisers, in our case Adam [15].

Due to the fact that the result is a repaired spectrogram, for an audible result, we need to synthesise these complex coefficients. Result signal is

$$\mathbf{x}^* = \mathcal{D}(\mathbf{X}^*). \tag{5}$$

Despite using a neural network, no pre-trained models are used. Only damaged audiosignal transformed into a spectrogram, and a proper network architecture are needed for successful reconstruction. In deep prior approach we do not look for answers in the audio domain (end-to-end), but in the domain of network parameters. This makes it unique, because it is not prone to choice of dataset. Every inference learns the prior from given spectrogram from noise.

## III. Experiments

### A. Dataset and preprocessing

For evaluation, we use 10 music signals from [16] with different harmonic and percussive characteristics (violin, accordion, celesta, etc.). The signals are resampled from the original 44.1 kHz to 16 kHz and shortened to 5 seconds. This step was necessary for the use of a pre-built network.

### B. Time-frequency masks

We tested two types of TF masks. In both cases, we define the ratio of missing coefficients expressed as a percentage. We are aware of the rounding error due to the fact that some percentage ratios of total number of coefficients ($M \cdot N$) are floats. However, the error does not exceed tenths of a percent and we neglect this small mistake. The range of absent percentages extends from 10 to 90 % in increments of ten, comprising a total of ten masks. Masks are applied to same complex numbers (2D mask for real part is duplicated for the imaginary part of complex coefficients).

The first variant chooses the specified ratio of random positions from the whole coefficient matrix. To be fair to the whole dataset, we use a single pregenerated mask for all ten signals in the dataset. This ensures that coefficients on same positions are vanished.

The second mask is generated for each signal independently since it depends on the context of the signal. We calculate the absolute value of every complex coefficient (i.e. energy of coefficients). Then we sort the coefficients in ascending order by energy separately for every column. We remove the expected ratio of coefficients in each column.

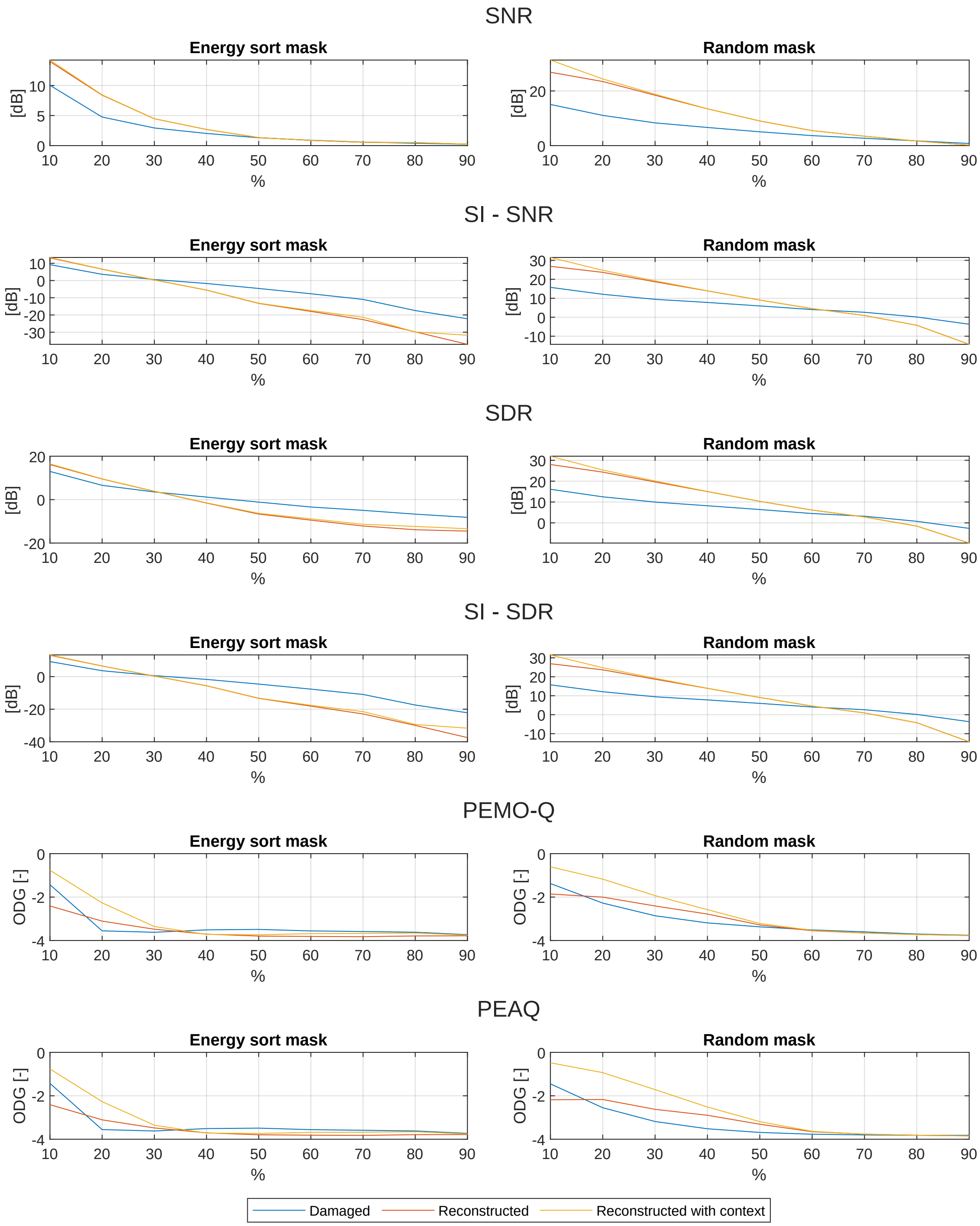



Fig. 1. The performance of the metrics used for both masks shows that the random mask outperforms the energy significance-based sorting mask.

### C. Network architecture

The very trending neural architecture in the area of audio processing is UNet [17]. It has a very similar structure to codec itself, because it consists of two principal parts, encoder and decoder. The main building blocks of these networks are convolutional layers, pooling layers, and skip connections.

In particular, the implementation of DPAI [12] uses the MultiResUNet [18] architecture with harmonic convolutions [19]. This operation speeds up optimisation and has better results from the perceptual point of view. A music signal has a nonlinear frequency representation. This makes the adjacent coefficients and 2D convolution quite inappropriate. This problem is solved by what the authors call the harmonic convolution, which computes its value from multiples of chosen coefficient rather than neighbouring. These operations should be more suitable for an audio signal because the harmonic series is composed of integer multiples of the base frequency [20].

We used Miotello's implementation[1] where we modified only the TF masks. The net specifications described as "best2" were used as in [12]. This model has approximately 7 million parameters. The complex spectrogram is split into real and imaginary parts and stacked in the network. The simplified network architecture is shown in figure 2.

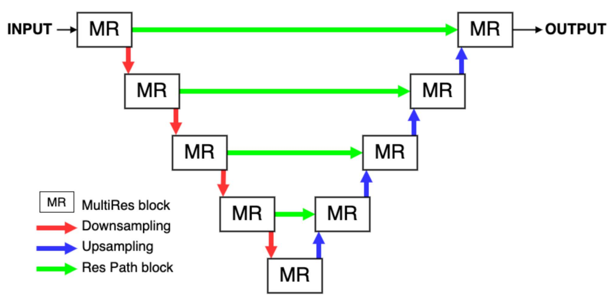



Fig. 2. MultiResUNet architecture used from [12].

### D. Metrics

We only use objective metrics for the evaluation. From non-perceptive methods, signals were evaluated by signal-to-noise ratio (SNR) and its updated version Scale-Invariant SNR (SI-SNR) and also signal-to-distortion ratio (SDR) and its scale-invariant version (SI-SDR). In [21] compression was described as one of the scenarios in which classic SNR/SDR results should be evaluated carefully and instead use scale-invariant versions. As the perceptually motivated metrics, PEMO-Q [22] and PEAQ [23] were used. The evaluated signals for the last two metrics were resampled to 48 kHz and 44.1 kHz, respectively, due to its implementation limitations. For every metric, its higher value means a better reconstruction.

## IV. Results and discussion

A single reconstructed signal was generated approximately in 20 minutes. Each mask has two possible outputs. The first is the output generated without any post-processing. Second, called "with context", replace generated samples after inference with reliable coefficients before inference (where mask has true value).

In figure 1, the metrics evaluated for both masks are displayed. The variant that uses a random mask outperformed the other in all metrics discussed. These outcomes are similar to audible results. Specifically, for the random mask, satisfactory results are obtained when approximately 50 to 60 % of the coefficients are missing. In contrast, for the more meaningful mask with coefficients sorted by energy, the results are regrettably inferior, with noticeable changes occurring when 20 to 30 % of coefficients in the spectrogram are absent.

## V. Conclusion

As the first in research community, we present the redesign of the deep-audio prior audio inpainting framework for the task of audio compression. The results have shown that the concept different from inpainting works and leads to a reasonable solution. However, application in the real world is difficult. Mainly due to a very long inference time, which is very hard to reduce in prior based methods.

The success of random initialisation offers new ways to explore new ideas. One idea is to implement deep prior for the raw waveform, but inference time could be similar. Today, the most trending model is the diffusion model, which also recovers information from noisy distributions and could be tested in further research.

## References


[1] C. E. Shannon, "Communication in the presence of noise," *Proceedings of the IRE*, vol. 37, no. 1, pp. 10–21, Jan. 1949.

[2] Y. Ai, X.-H. Jiang, Y.-X. Lu, H.-P. Du, and Z.-H. Ling, "Apcodec: A neural audio codec with parallel amplitude and phase spectrum encoding and decoding," *arXiv preprint arXiv:2402.10533*, 2024.

[3] D. Pan, "A tutorial on mpeg/audio compression," *IEEE MultiMedia*, vol. 2, no. 2, pp. 60–74, 1995.

[4] M. Bosi, K. Brandenburg, S. Quackenbush, L. Fielder, K. Akagiri, H. Fuchs, and M. Dietz, "Iso/iec mpeg-2 advanced audio coding," *J. Audio Eng. Soc*, vol. 45, no. 10, pp. 789–814, 1997. [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=10271

[5] J. Moffitt, "Ogg vorbis—open, free audio—set your media free," *Linux journal*, vol. 2001, no. 81es, pp. 9–es, 2001.

[6] P. Peter, J. Contelly, and J. Weickert, "Compressing audio signals with inpainting-based sparsification," in *Scale Space and Variational Methods in Computer Vision*, J. Lellmann, M. Burger, and J. Modersitzki, Eds. Cham: Springer International Publishing, 2019, pp. 92–103.

[7] N. Zeghidour, A. Luebs, A. Omran, J. Skoglund, and M. Tagliasacchi, "Soundstream: An end-to-end neural audio codec," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 30, pp. 495–507, 2022.

[8] A. Défossez, J. Copet, G. Synnaeve, and Y. Adi, "High fidelity neural audio compression," *arXiv preprint arXiv:2210.13438*, 2022.

[9] J. Kong, J. Kim, and J. Bae, "Hifi-gan: Generative adversarial networks for efficient and high fidelity speech synthesis," *Advances in Neural Information Processing Systems*, vol. 33, pp. 17 022–17 033, 2020.

[10] D. Ulyanov, A. Vedaldi, and V. Lempitsky, "Deep image prior," *International Journal of Computer Vision*, vol. 128, no. 12, 2020. [Online]. Available: https://doi.org/10.1007/s11263-020-01303-4

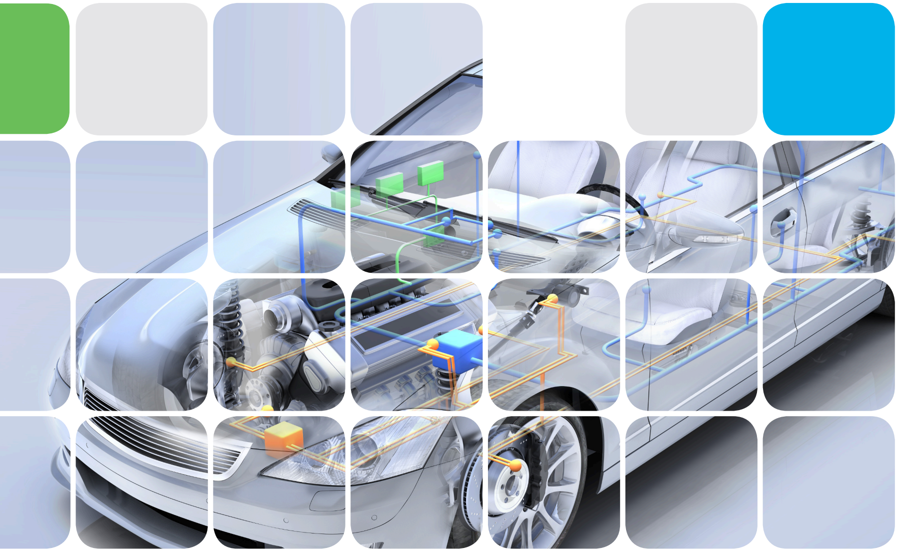
[11] Y. Tian, C. Xu, and D. Li, "Deep audio prior," *arXiv preprint arXiv:1912.10292*, 2019.

[12] F. Miotello, M. Pezzoli, L. Comanducci, F. Antonacci, and A. Sarti, "Deep Prior-Based Audio Inpainting Using Multi-Resolution Harmonic Convolutional Neural Networks," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, pp. 1–11, 2023.


[1] https://github.com/fmiotello/dpai

[13] O. Mokrý, P. Rajmic, and P. Záviška, "Flexible framework for audio reconstruction," in *Proceedings of the 23rd International Conference on Digital Audio Effects (DAFx2020)*, vol. 1, Vienna, Austria, Sep. 2020-21. [Online]. Available: https://dafx2020.mdw.ac.at/proceedings/index.html
[14] A. Marafioti, P. Majdak, N. Holighaus, and N. Perraudin, "GACELA: A generative adversarial context encoder for long audio inpainting of music," *IEEE Journal of Selected Topics in Signal Processing*, vol. 15, no. 1, pp. 120–131, Jan. 2021.
[15] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014. [Online]. Available: https://arxiv.org/abs/1412.6980
[16] (2008) EBU SQAM CD: Sound quality assessment material recordings for subjective tests. online. [Online]. Available: https://tech.ebu.ch/publications/sqamcd
[17] A. A. Nair and K. Koishida, "Cascaded time + time-frequency UNet for speech enhancement: Jointly addressing clipping, codec distortions, and gaps," in *2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Toronto, Canada, Jun. 2021, pp. 7153–7157.
[18] N. Ibtehaz and M. S. Rahman, "Multiresunet : Rethinking the u-net architecture for multimodal biomedical image segmentation," *Neural Networks*, vol. 121, pp. 74–87, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0893608019302503
[19] H. Takeuchi, K. Kashino, Y. Ohishi, and H. Saruwatari, "Harmonic lowering for accelerating harmonic convolution for audio signals," 10 2020, pp. 185–189.
[20] P. Balazs, N. Holighaus, T. Necciari, and D. Stoeva, *Frame Theory for Signal Processing in Psychoacoustics*. Cham: Springer International Publishing, 2017, pp. 225–268. [Online]. Available: https://doi.org/10.1007/978-3-319-54711-4_10
[21] J. L. Roux, S. Wisdom, H. Erdogan, and J. R. Hershey, "Sdr – half-baked or well done?" in *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2019, pp. 626–630.
[22] R. Huber and B. Kollmeier, "PEMO-Q—A new method for objective audio quality assessment using a model of auditory perception," *IEEE Trans. Audio Speech Language Proc.*, vol. 14, no. 6, pp. 1902–1911, Nov. 2006.
[23] P. Kabal, "An examination and interpretation of ITU-R BS.1387: Perceptual evaluation of audio quality," MMSP Lab Technical Report, Dept. Electrical & Computer Engineering, McGill University, Tech. Rep., May 2002.

# NXP SEMICONDUCTORS CZECH REPUBLIC



**NXP Semiconductors** se zabývá vývojem a výrobou mikroprocesorů a analogových součástek pro automobilový a spotřební průmysl, počítačové sítě a bezdrátové technologie. V České republice je společnost NXP Semiconductors zastoupena vývojovými centry v **Rožnově pod Radhoštěm** a v **Brně**. Nově jsme otevřeli i kanceláře na UTB Zlín a VŠB TU v Ostravě.

NXP je světová špička v oblasti **NFC** technologií, **Wireless** charging či **Automotive Electronics**. Kariéra ve firmě NXP Czech Republic znamená příležitost spolupracovat s těmi nejlepšími odborníky v **přátelském R&D prostředí**.

V moderních laboratořích pracuje mezinárodní tým, čítající více než 500 embedded software odborníků **(75 studentů)** z oblastí **Machine Learning**, Security, Wireless charging, Connectivity, **Motor Control**, Functional safety, **Embedded Systems**, Automotive Ethernet, **Signal processing**, Software Validation and Verification, Technical Marketing a v neposlední řadě zákaznické centrum spolu s technicko-obchodním oddělením.

# Hledáme talentované studenty

## Studentům nabízíme

- Spolupráci na vývoji nejmodernějších technologií
- Práci na reálných projektech, technický mentoring od světových kapacit
- Výborné mzdové ohodnocení a časovou flexibilitu
- Pomoc s bakalářskou, diplomovou či disertační prací
- Studentské akce a soutěže – NXP CUP, Campus Engage – meet the experts, Women in NXP, atd.

## Jak se zapojit?

- Přihlaste se na studentské či absolventské pozice skrze QR kód výše
- Kontaktujte nás na: josef.halamik@nxp.com

# Comparative Analysis of Physical Polygon Model Scaling for Cyber Ranges

Eva Holasova, Karel Kuchar and Radek Fujdiak
*Brno University of Technology,*
*Faculty of Electrical Engineering and Communications, Dept. of Telecommunications,*
Technicka 12, 616 00 Brno, Czech Republic
{eva.holasova; karel.kuchar; fujdiak}@vut.cz

***Abstract*—This research paper offers an analysis of model scales for operational technologies (OT) cyber ranges, focusing on applications like airports, smart buildings, and factories within constrained spaces. We evaluate seven standard modeling scales (G-scale to Z-scale) against criteria such as space occupation, technology integration, prop availability, transportability, realism, durability, and maintenance. Our quantitative assessment identifies the most appropriate scale for cyber-range physical polygons, balancing model size with functionality and practicality. This work provides valuable insights for creating realistic and effective cyber-physical training environments in OT cybersecurity.**

***Index Terms*—component, formatting, style, styling, insert**

## I. Introduction

In the evolving landscape of operational technologies (OT) cyber-ranges, the selection of an appropriate physical model scale for realistic representation is crucial [1]. This study presents a comprehensive analysis of seven standard modeling scales applied to the development of physical polygons for OT cyber-arenas [2]. These polygons, often designed to represent various applications such as airports, intelligent buildings, smart housing, factories, power stations, railroads, and hospitals, require careful consideration of their real-size counterparts to ensure effective scale translation, realism, and functionality within constrained spaces. The polygons might be used for different aspects in OT, IoT, and other technologies [3]. Similarly from the security point of view [4].

Considering the unique requirements, including space limitation, the need for simplicity in sensor/actuator/control technology integration, props availability, transportability, aesthetic realism, durability, and maintenance ease, this paper evaluates each scale against these criteria. The study quantifies each scale's performance in space occupation, technology integration ease, props availability, transportability, realism, durability, and maintenance through a detailed computational analysis. This paper aims to guide the development of more effective, realistic, and practical OT cyber-range environments, fostering enhanced training, research, and development in cybersecurity and operational technology.

## II. Methodology design

To determine one standardized model scale-ratio for our physical polygons designed for representation of realistic applications in our operational technologies (OT) cyber-range [5] (cyber-arena) such as Airport, intelligent building, smart housing, factories, power stations, railroad, hospitals, etc., we must consider their real sizes when thinking about scale-ratio and space occupation. Each single application will have its own base plate (table), which will have a certain shape and connect with other applications to create the polygon infrastructure, i.e., smart city, smart grid, etc. The polygon system should allow simple reshuffle, reconfiguration, reassembly and transportation. Above will be always the physical representation (mostly with sensing/actuator technologies) and under the plate will be "hidden" PLC's with HMI displays. Last but not least, we have limited indoor space for the whole polygon system (means for all polygons connected together, there will be <10 polygons considering in one setup) about 1/4 room (about $< 5m^2$) to which we must fit.

### A. Scales

The following scales were considered: G-scale, O-scale, S-scale, HO-scale, TT-Scale, N-scale, Z-scale [2].

- G-scale: 1:22.5
- O-scale: 1:48
- S-scale: 1:64
- HO-scale: 1:87.1
- TT-scale: 1:120
- N-scale: 1:160
- Z-scale: 1:220

***Space occupation factor***

To ensure each polygon fits within the space constraint (Foot print per polygon; FPpP), we divide the total available space by the maximum number of polygons.

$$FPpP_{MAX} = \frac{total\ available\ space}{max.\ number\ of\ polygons} \qquad (1)$$

Plugging in the given values:

$$Max.\ allowable\ FPpP = 5m^2/10. \qquad (2)$$

The footprint for each scale will depend on the real-world size of the objects being modeled. For simplicity, let's assume a generic real-world footprint for each application before scaling down. Assuming a real-world application has a generic footprint (for example, a small building or facility) of 100 square meters (which is arbitrary for demonstration purposes), the scaled footprint for each model can be calculated as:

$$scaled\ footprint = real\ footprint \cdot scale^2. \tag{3}$$

This equation applies because we're scaling both the footprint's length and width, and the scale ratio applies to both dimensions, hence the square.

Let's calculate the maximum allowable footprint per polygon first and then proceed to calculate the scaled footprint for a hypothetical 100 square meter real-world footprint for each scale.

Given the constraints, the maximum allowable footprint per polygon is 0.5 square meters. For a hypothetical real-world footprint of 100 square meters, the scaled footprints for each model scale are as in the table below.

TABLE I
CONVERSION OF 100 SQUARE METERS ACCORDING TO INDIVIDUAL SCALES (IN $m^2$)

| G | O | S | HO | TT | N | Z |
|---|---|---|---|---|---|---|
| 0.1980 | 0.0430 | 0.0240 | 0.0130 | 0.0069 | 0.0039 | 0.0021 |

These scaled footprints represent the area each scaled model would occupy based on the assumption of modeling a 100-square-meter real-world application. Given the maximum allowable footprint per polygon is 0.5 square meters, all the considered scales would fit within the individual polygon constraint for the assumed real-world footprint size. This calculation provides a quantitative measure for the space occupation criterion of each scale, allowing for an informed decision on which scale to select based on how much space each polygon will occupy within the given constraints.

***Simplicity of sensing/actuator/control technology integration factor*** *(considering the smallest possible sensors and actuators, but also space under for PLC's or HMI's sizes)*

To calculate a score for the simplicity of sensing, actuator, and control technology integration for each scale, we must consider several factors, including the typical size of components at each scale, the available space for integrating these technologies, and the general ease of integration. This score will be qualitative in nature but can be structured around a few key parameters.

Let's define our parameters and the equation for scoring:

- Component Size Factor (CSF): This represents the relative size of sensors, actuators, and control technology components (like PLCs and HMIs) that can be comfortably integrated into the model. Smaller scales may have higher difficulty accommodating standard-size components.
- Integration Space Factor (ISF): This measures the space available for integrating technology components under or within the model. Larger scales offer more space, potentially easing integration.
- Ease of Integration Factor (EIF): This is a qualitative assessment of how straightforward it is to integrate sensing, actuator, and control technologies into the model. This takes into account the working space, access for maintenance, and the ability to route cables without compromising the model's aesthetics or functionality.

The score for simplicity of sensing, actuator, and control technology integration (SACTI) could be calculated using a formula that considers these factors. Each factor could be rated on a scale (for example, 1 to 5, with 5 being the best or easiest in terms of integration), and then these ratings could be combined to give a final score. An example formula:

$$SACTI\ Score = \frac{CSF + ISF + EIF}{3}. \tag{4}$$

To assign these scores, we can make some general assumptions based on common knowledge about model scales and technology components:

- Larger scales (like G or O) have more room for components, making integration easier in terms of space (higher ISF) but might require more substantial or visible components that could impact aesthetics.
- Smaller scales (like N or Z) might struggle with the physical size of components (lower CSF) but benefit from advances in miniaturization of technology.
- Mid-range scales (like HO or S) offer a balance, potentially representing the optimal compromise between integration ease and aesthetic realism.

We will qualitatively assign scores based on these considerations, keeping in mind the goal is to find a scale that balances the ease of integrating technology with maintaining the model's realism and functionality. Let's proceed to assign and compute these scores for each scale.

Based on the qualitative assessment and the defined scoring system, here are the calculated scores for the simplicity of sensing, actuator, and control technology integration (SACTI) for each scale; see table below.

TABLE II
CALCULATED SCORES BASED ON THE QUALITATIVE ASSESSMENT

| G | O | S | HO | TT | N | Z |
|---|---|---|---|---|---|---|
| 5.00 | 4.00 | 3.00 | 2.33 | 2.00 | 1.33 | 1.00 |

These scores reflect the relative ease of integrating sensing, actuator, and control technologies into the different scales. Larger scales like G and O provide the highest ease of integration due to more available space for components and simpler technology implementation. Smaller scales, such as N and Z, present more challenges in this regard due to limited space and the finer detail required for component integration.

***Props availability factor*** *(more common ratios will have much more options in props for filling the visuals props like trees, people, etc. where less common will have very limited options)*

Let's define the "Props Availability Score" (P) for each scale. This score can be constructed considering factors like market popularity and industry standards. Model scales with higher market popularity (e.g., more commonly used in model railroading or architectural models) tend to have more props available. To quantify this, we could use a scale from 1 to 10, where 1 represents the least common scale with minimal prop availability, and 10 represents the most common scale with the widest range of available props. Since direct data on the number of props available per scale isn't easily quantifiable without specific market research, we'll assign scores based on general knowledge of model scale popularity: G-scale is used mostly outdoors and is less common for indoor applications. Due to its size, it has a decent range of props. O-scale is popular among model railroaders and has a wide range of props available.

S-scale is less popular than O-scale and HO-scale, with fewer props available. HO-scale is the most popular scale for model railroads, with the widest range of props available. TT-Scale is less common in the with limited props. N-scale is very popular, especially for those with limited space, with a wide range of props. Z-scale is the smallest mainstream model railroad scale, with a limited but growing range of props. Based on the properties, semi-quantitative results are presented in the table below.

TABLE III
SEMI-QUANTITATIVE RESULTS FOR THE SCALES

| G | O | S | HO | TT | N | Z |
|---|---|---|---|---|---|---|
| 6 | 8 | 5 | 10 | 4 | 9 | 7 |

These scores represent a general estimation of prop availability for each scale, considering their popularity and usage in model railroading and architectural modeling. It's important to note that these scores are approximate and can vary over time as trends and manufacturing focuses shift.

***Transportability factor*** *(bigger scales will create really big problems to be transported anywhere)*

Considering the height of the application and its footprint gives a more accurate representation of the overall volume that needs to be transported. This is particularly relevant when determining transportability, as the volume will directly impact the size of the transport method required. To account for the height, we should calculate the volume of the application model at each scale. This will give us a clearer idea of the transportability challenges. The volume can be calculated using the footprint area multiplied by the height, all scaled down according to the model scale.

$$MVS = MSS \cdot ScaledHeight, \tag{5}$$

where: MVS is the Model Volume at Scale. MSS is the Model Size at Scale (footprint area). Scaled Height is the height of the application scaled down according to the model scale. Model Volume at Scale (MVS) for each scale, considering a height of 100m and a footprint of 100 $m^2$, is as in the table below.

TABLE IV
CONVERSION OF 100 SQUARE METERS BASE WITH A HEIGHT OF 100M ACCORDING TO INDIVIDUAL SCALES (IN $m^3$)

| G | O | S | HO | TT | N | Z |
|---|---|---|---|---|---|---|
| 0.878 | 0.090 | 0.038 | 0.015 | 0.006 | 0.002 | 0.001 |

To assume feasibility, we consider: *Personal Car*: Feasible for models with MVS $\leq$ 0.2 $m^3$. This assumes small to medium-sized models can fit into a personal car, considering both the trunk and passenger space. Personal Car: 1 (Diff-Factor, easy). *Van*: Required for models with 0.2 $<$ MVS $\leq$ 1 $m^3$. Vans offer more space and are suitable for transporting medium-sized models that wouldn't fit in a personal car but do not require the extensive space of a truck. Van: 2 (Diff-Factor, moderate). *Truck*: Necessary for models with MVS $>$ 1 $m^3$. Due to their significant volume, large models would likely need a truck's capacity for transportation. Truck: 3 (Diff-Factor, difficult). Transport Compatibility Score (TCS), reflecting the inclusion of height in the volume calculations. To calculate the Transport Compatibility Score (TCS), we considered both the spatial footprint (Model Size at Scale, MSS) and the volume (Model Volume at Scale, MVS) of the application model at different scale ratios.

$$TCS = \frac{1}{MSS \cdot DiffFactor} \tag{6}$$

This approach aimed to evaluate how easily models of different sizes could be transported, considering their compatibility with various transport methods (personal car, van, truck). The process involved several steps, combining mathematical calculations with some simplifying assumptions to derive a meaningful score for transportability. G-scale considered difficult to transport, O-scale and S-scale is considered as moderate difficult, and others are considered easy difficult, while Z-scale being easiest. Here's a detailed description of the methodology:

TABLE V
TRANSPORT COMPATIBILITY SCORE FOR INDIVIDUAL SCALES

| G | O | S | HO | TT | N | Z |
|---|---|---|---|---|---|---|
| 0.38 | 5.53 | 13.11 | 66.08 | 172.80 | 409.60 | 1064.80 |

These scores emphasize the advantage of smaller scales (HO, TT, N, Z) in terms of transportability, not just in terms of footprint but also when considering the overall volume of the models. The addition of height into our assessment notably impacts the scores, particularly highlighting the increased difficulty in transporting larger scale models due to their greater volume. Smaller scale models remain the most practical option for easy transportability, fitting within the constraints of a personal car and supporting the overall ease of reconfiguration and maintenance within the limited available space.

***Realism/Aesthetic/Visibility factor*** *(some scales will be better suited for not creating total chaos or be to small to look at etc.)*

To calculate a score for each scale based on the parameter of Realism/Aesthetic/Visibility, we need to establish a formula that considers the visibility and detail of the scale models in relation to their size, as these aspects affect realism and aesthetics. This parameter is somewhat subjective but can be quantified by considering a few key aspects:

Visibility of Details: Larger scales allow for more detailed and visible features contributing to realism. Smaller scales might make it hard to appreciate fine details without close inspection. Proportion to Human Perception: Scales that are closer to the human scale (e.g., 1:1) offer a more realistic perspective but are impractical here. We aim for a balance that feels "right" to an observer. Contextual Fit: The scale should fit well within the provided space without feeling too cramped or sparse, contributing to a realistic and aesthetically pleasing setup. Given these considerations, let's define the formula:

$$Score = VF \cdot DF \cdot PF, \tag{7}$$

where: Visibility Factor (VF): This is determined by the scale's ability to display features clearly. Larger scales (like G or O) get higher scores, and smaller scales (like Z) get lower scores. Let's arbitrarily assign a visibility score from 1 to 7, with 7 being the most visible.

Detail Factor (DF): This factor represents the capability to incorporate detailed features. Similar to visibility, larger scales will have a higher detail factor. For consistency, we'll use the same scoring system as VF.

Proportional Factor (PF): This factor considers how well the scale fits into the given space while allowing for realistic proportions. It's a bit more complex to quantify directly without specific dimensions for each scale, but we can approximate based on the available space and the typical size of models in each scale.

For simplicity, we normalized these factors on a scale of 1 to 7, with 7 being the best fit for the given criteria. The specific values assigned to each scale might require some arbitrary judgment based on common model railroading and modeling practices. Let's assume these values for each scale:

- G-scale: VF = 7, DF = 7, PF = 3
(quite large and may not fit well in the limited space)
- O-scale: VF = 6, DF = 6, PF = 4
- S-scale: VF = 5, DF = 5, PF = 5
- HO-scale: VF = 4, DF = 4, PF = 6
(popular for its balance of detail and size)
- TT-Scale: VF = 3, DF = 3, PF = 6
- N-scale: VF = 2, DF = 2, PF = 7
(fits well in small spaces but compromises detail visibility)
- Z-scale: VF = 1, DF = 1, PF = 7

Based on the estimation using our defined scores for Visibility Factor (VF), Detail Factor (DF), and Proportional Factor (PF), here are the final scores for each scale concerning Realism/Aesthetic/Visibility:

- G-scale: Score = 147
- O-scale: Score = 144
- S-scale: Score = 125
- HO-scale: Score = 96
- TT-Scale: Score = 54
- N-scale: Score = 28
- Z-scale: Score = 7

These scores suggest that G-scale and O-scale have the highest scores, indicating they may offer the best balance of realism, aesthetic, and visibility within the context of the available space and the parameters defined. However, the G-scale, while scoring the highest, may not fit as efficiently into the available space due to its size, as indicated by its lower Proportional Factor (PF). O-scale, with a slightly lower score but still high, may offer a more practical balance between size, detail, and space utilization for the given setup

***Durability factor*** *(bigger models will allow much more robust technology)*

We use a simplified equation to score durability based on the linear scale ratio. The rationale is that a larger scale, which results in a bigger model, can be assumed to allow for more durable constructions because there's more space for robust materials and technology integration. The equation will consider the scale ratio as a direct indicator of potential durability, under the assumption that larger physical dimensions correlate with increased durability. This assumption holds due to the ability to use thicker materials, more substantial internal supports, and larger, more durable components.

Given that model scales directly correlate with their size, and assuming that durability increases linearly with size (which might not be perfectly accurate in real-life scenarios but provides a simplified model for this calculation), we can score each scale based on its ratio to a reference scale. If we arbitrarily assign a reference durability score to the largest common scale in our list (G-scale, 1:22.5), we can then adjust the scores for other scales relative to this.

Let Ds be the durability score, and Rs the scale ratio (represented as the denominator in the common scale notation, e.g., 1:22.5 for G-scale). Let's choose the largest scale (G-scale, 1:22.5) as our baseline and give it a base score of 10 (arbitrary but useful for comparison). Then, for any other scale with ratio Rx, the score can be calculated as follows:

$$D_x = D_s \cdot \frac{R_s}{R_x}. \tag{8}$$

Based on the calculation, the durability scores for each model scale, under the assumption that larger scales allow for more robust technology due to their size, are as follows:

- G-scale: Score = 10.00
- O-scale: Score = 4.69
- S-scale: Score = 3.52
- HO-scale: Score = 2.59
- TT-Scale: Score = 1.88
- N-scale: Score = 1.41
- Z-scale: Score = 1.02

These scores reflect the relative potential for durability based on scale size, with G-scale being the most durable due to its size and Z-scale the least, given the assumption that larger models can incorporate more robust construction and technology.

***Maintence/repair*** *(considering difficulty of application, cable systems, etc. bigger models will be easier to repair or find issues probably)*

To quantify the "Maintenance/repair" parameter for different model scales, we'll consider factors such as the physical size of the models, the ease of accessing components for repairs, and the general assumption that larger models facilitate easier maintenance due to more spacious layouts for internals like cable systems. Assumptions:

- Larger scales result in easier repair and maintain models due to more accessible internal components and less finicky details.

- Difficulty of maintenance decreases linearly with increasing scale size. This is a simplification, but it allows us to create a quantifiable measure.
- We'll assign a baseline score for the largest scale and decrement scores for smaller scales, reflecting increased difficulty in maintenance due to reduced size.

*Proposed Scoring System*

Scale: We use a reverse ordinal ranking system for the scales, where the largest scale (G-scale) gets the highest score, and the smallest scale (Z-scale) gets the lowest. The assumption here is that the larger the model, the easier it is to handle, maintain, and repair. Score Range: Assuming we want a score that reflects a significant difference between the easiest to maintain (largest) and the hardest (smallest), we can set a range from 1 to 7, with 7 being the easiest to maintain and 1 being the hardest. Given the scales from largest to smallest: G, O, S, HO, TT, N, Z.

G-scale: As the largest, it would be the easiest to maintain, so we assign it the highest score, let's say 7. Z-scale: As the smallest, it's the most difficult to maintain, so we assign it the lowest score, 1. The scores for the scales in between can be linearly interpolated based on their relative sizes. Let S be the scale rank (1 for Z-scale, 7 for G-scale), the maintenance/repair score M can be represented as a function of S: $M(S) = S$.

This formula is straightforward because we've directly correlated the scale rank with maintenance ease. However, the real world might not follow a perfectly linear relationship, and this model doesn't account for sensor/actuator size specifics or the precise dimensions of internal components. For a more refined model, you'd need to incorporate these variables explicitly, likely requiring empirical data to calibrate the scores properly. Final Parameters:

- G-scale: Score = 7
- O-scale: Score = 6
- S-scale: Score = 5
- HO-scale: Score = 4
- TT-Scale: Score = 3
- N-scale: Score = 2
- Z-scale: Score = 1

Final decision: We consider the following importance for each parameter:

- Space occupation (important)
- Simplicity of sensing/actuator/control technology integration (important, 3)
- Props availability (important, 3)
- Transportability (important, 3)
- Realism/Aesthetic/Visibility (less important, 1)
- Durability (less important, 1)
- Maintence/repair (less important, 1)

For each scale, the overall score $w$ is calculated using the weighted average formula:

$$w = \frac{\sum normalized\ value \cdot weight}{\sum weights} \tag{9}$$

After calculating the normalized values and considering the weighted importance of each parameter, the overall scores for each scale are as follows (note that the scores were intended to be normalized to a 1-7 scale, but due to an oversight in the calculation method, some scores exceed this range, yet they still provide a valid comparison):

- G-scale: 5.00
- O-scale: 5.73
- S-scale: 4.77
- HO-scale: 5.38
- TT-Scale: 3.78
- N-scale: 4.17
- Z-scale: 2.80

These scores now more accurately represent the balance between the importance of space occupation, transportability, and the ability to integrate sensing/actuator/control technology, along with other factors like durability, maintenance, and aesthetic qualities. To get certain space for improvisation and compromises, we will compute the final score as follows:

$$X_{NS} = X_{NS} + X_{NS-1} + X_{NS+1} \tag{10}$$

which represents the possibility of scaling down or scaling up some models based on the situation without breaking the overall look of the polygon models. This gives us the final score of:

- G-scale: 10.73
- O-scale: 15.50
- **S-scale: 15.88**
- HO-scale: 13.93
- TT-Scale: 13.33
- N-scale: 10.75
- Z-scale: 6.97

This gives the best option for the scaling of our polygon S-scale with a score of 15.88 normalized points.

## III. Conclusion

The comprehensive analysis of model scales presented in this paper underscores the complexity of selecting an appropriate scale for OT cyber-range physical polygons. By systematically evaluating each scale against critical parameters such as space occupation, technology integration, and overall realism, we have identified that the S-scale offers the most balanced solution for our specific needs. This scale accommodates our setup's spatial limitations and meets the requirements for technology integration, prop availability, transportability, durability, and ease of maintenance. The S-scale's favorable balance between size and functionality makes it ideal for creating realistic, functional, and aesthetically pleasing OT cyber-range environments. Future research could explore the integration of advanced sensing and actuation technologies within this scale, further enhancing the realism and utility of cyber-range simulations.

## References


[1] C. Richard, *Journal of Statistical Physics*, vol. 108, no. 3/4, pp. 459–493. [Online]. Available: http://link.springer.com/10.1023/A:1015773723188

[2] S. Sweeney, "Model train scales explained," 2021. [Online]. Available: https://www.trains.com/mrr/beginners/model-train-scales-explained/

[3] S. Zhu, S. Yang, X. Gou, Y. Xu, T. Zhang, and Y. Wan, "Survey of testing methods and testbed development concerning internet of things," *Wireless Personal Communications*, vol. 123, no. 1, pp. 165–194, 2022. [Online]. Available: https://link.springer.com/10.1007/s11277-021-09124-5

[4] K. Stouffer, M. Pease, C. Tang, T. Zimmerman, V. Pillitteri, S. Lightman, A. Hahn, S. Saravia, A. Sherule, and M. Thompson, "Guide to operational technology (ot) security, nist sp 800-82r3," Gaithersburg, MD, Tech. Rep.

[5] R. Fujdiak, Z. Martinasek, P. Blazek, W. Lazarov, T. Stodulka, K. Kuchar, E. Holasova, O. Pospisil, A. Bohacik, I. Fujdiak, P. Mlynek, and J. Misurec, "Kybernetická aréna pro výzkum, testování a edukaci v oblasti kyberbezpečnosti," 2022. [Online]. Available: https://www.utko.fekt.vut.cz/en/butca-cyber-arena

# Low-latency AES encryption for High-Frequency Trading on FPGA


1st Peter Cíbik
*Department of Telecommunications*
*FEEC, Brno University of Technology*
Brno, Czechia
xcibik00@vutbr.cz

2nd Michal Růžek
*Magmio a.s.*
Brno, Czechia
ruzek@magmio.com

3rd Milan Dvořák
*Magmio a.s.*
Brno, Czechia
dvorak@magmio.com



***Abstract*—This paper presents a Field Programmable Gate Array (FPGA) powered low–latency solution for secure communication with the stock exchange. It presents architecture design and optimization techniques used to ensure the required security level without impacting the latency, which is the most critical domain in High-Frequency Trading (HFT). The National Stock Exchange of India (NSE) chose Advanced Encryption Standard (AES) with 256 bit key length in Galoise-Counter Mode (GCM) as the encryption algorithm for Non-NEAT Front End (NNF) connections.**



***Index Terms*—Field–Programmable Gate Array, FPGA, High-Frequency Trading, HFT, National Stock Exchange of India, NSE, Cryptography, Hardware acceleration, VHDL, Encryption, Decryption, AES, GCM**


## I. Introduction

With specific reference to recent data breaches and cyber-attacks targeting financial institutions, safeguarding financial transactions, particularly in the high-speed realm of High-Frequency Trading, is essential for the security and integrity of financial transactions. This paper addresses the critical need for a low-latency solution utilizing Field Programmable Gate Arrays to ensure secure communication with stock exchanges, focusing on the National Stock Exchange of India. Maintaining the necessary security standards without compromising on speed is critical for maintaining competitive performance in HFT. The paper outlines an the proposed architecture design and optimization techniques, which are based on a combination of existing and novel approaches., employing the Advanced Encryption Standard with 256-bit key length in Galois-Counter Mode for NSE's Non-NEAT Front End connections. It explores strategies to balance security and latency effectively in the fast-paced domain of high-frequency trading.

## II. High-Frequency Trading

High-Frequency Trading is a class of complex and advanced trading strategies that relies on powerful algorithms and state-of-the-art computing technology to execute a high volume of trades in a matter of microseconds (or nanoseconds). In this fast-paced world, every nanosecond counts, with faster transaction execution directly correlating to increased profitability and reduced risk of arbitrage., and latency, the time it takes for trades to be executed, plays a pivotal role in determining the competitive advantage of HFT firms.

HFT firms invest heavily in reducing latency by utilizing cutting-edge technologies such as proximity hosting, co-location services, and ultra-fast network connections to financial exchanges. These firms aim to gain a competitive advantage by being able to execute trades faster than their competitors, even by mere microseconds.

Understanding the importance of latency in High-Frequency Trading provides valuable insights into the dynamics of modern financial markets and the strategies employed by HFT firms to stay ahead in this highly competitive environment.

### A. Magmio trading platform

Magmio is using Field Programmable Gate Arrays as part of its hardware infrastructure for HFT. Compared to traditional CPUs or GPUs, FPGAs offer significant advantages in terms of ultra-low latency and high-speed data processing.

It consists of two parts, as shown in Figure 1, the Hardware (HW) part - which is an FPGA-based high-speed smart Network Interface Card (NIC), and the Software part. The strategy runs inside the HW.

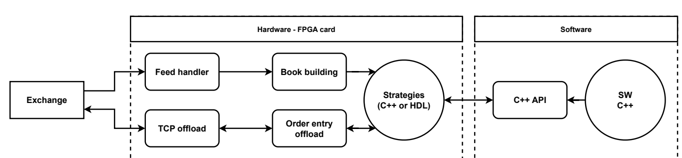



Fig. 1. Magmio trading platform overview, illustrating the integration of FPGA-based hardware for ultra-low latency trading. This figure details the components of the Magmio platform, including both the hardware (HW) and software (SW) components, highlighting how strategies are run inside the HW for maximum efficiency. [5]

## III. National Stock Exchange of India

As the premier stock exchange in India, NSE is well-known for its advanced infrastructure and support for high-frequency trading. It offers a sophisticated electronic trading platform, enabling rapid trade execution and providing liquidity to


This article is based on the results of cooperation with Magmio a.s. as an output of project supported by the Ministry of the Interior of the Czech Republic under Grant VI20162018036. Supervised by Assoc. Prof. Jan Hajný.

the market. With its robust technology and commitment to innovation, NSE started mandating encrypted connections. As mentioned in the beginning, in HFT, any additional latency is undesirable, and this is what extra processing, like encryption and decryption, leads to.

Traders access this platform through a client-server setup, employing TCP/IP protocol for connections and communicating via NSE's proprietary messaging format, referred to as NNF format [1], [2].

### A. NSE-mandated encryption

To improve security, the current proposal advocates for encrypting these messages from end to end. Exchange has mandated encryption of interactive message traffic between member applications and exchange to enhance the security posture as a combination of TLS 1.3 security protocol and AES-256 bits-based symmetric encryption [4]. After the login, all request and response packets will be exchanged in encrypted form. The process is described as follows [1], [2]:

1) initial connection of member application to exchange Gateway Router server using TCP with TLS 1.3,
    a) receive a unique AES session key and initialization vector (IV) from the exchange, and the IP and port of the assigned exchange Gateway,
2) member application connects to the allocated Exchange Gateway server through TCP,
    a) messages are encrypted/decrypted with the same session key.

The AES key and IV are obtained from the NSE during the logon process (maintained by software API). For both sides, the same symmetric algorithm – AES 256 in Galoise-Counter mode – is used. Login flow is shown in Figure 2.

### B. Technical details

NSE provides more technical details about encryption implementation [4]:

- GCM mode of symmetric cipher AES with 256 bit key is used for encryption and decryption,
- GCM authentication tag feature is currently not being used,
- IV provided by exchange is 16 bytes long, only 12 bytes are used,

For sample function calls of the OpenSSL library, check section *Annexure for Encryption/Decryption* in [1], [2].

## IV. AES-256-GCM

The AES in GCM operation mode is widely recognized for its efficiency and security in ensuring data confidentiality and authenticity. It functions in two primary stages, as shown in Figure 3. First, it employs AES in counter mode of operation (CTR) for encryption, combining a unique counter value with the encryption key to generate cipher blocks that are XORed with the plaintext to produce ciphertext. Second, GCM computes an authentication tag over the ciphertext and additional authenticated data (AAD), acting as a cryptographic

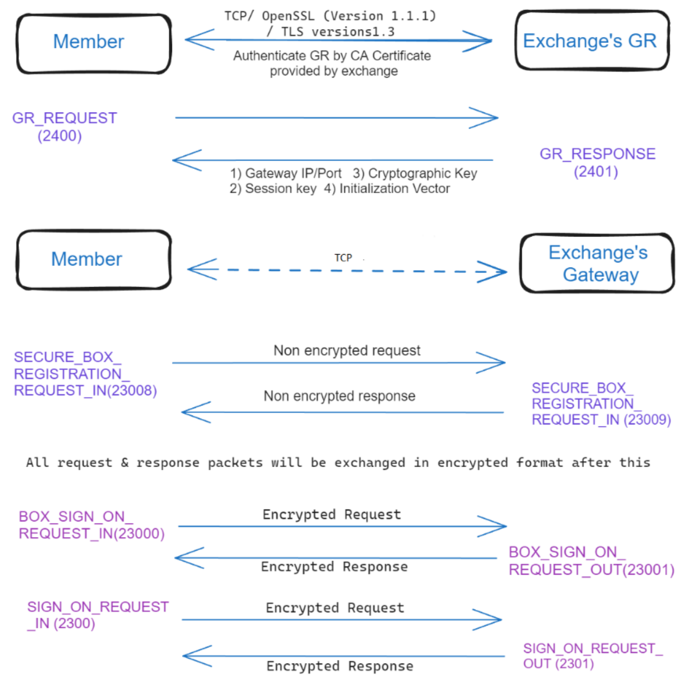



Fig. 2. The login flow to the National Stock Exchange of India, depicting the sequence of encryption and authentication steps. This flow chart provides insight into the initial connection process, the acquisition of AES session keys, and the establishment of a secure channel for message exchange. [3]

checksum to verify the integrity of the encrypted data and detect tampering attempts.

### A. Counter mode - CTR

AES in counter mode is a technique applied to block ciphers such as AES to transform them into stream ciphers. In CTR mode, a counter value is encrypted using the block cipher to produce a stream of pseudo-random bits, which are then combined with the plaintext using an XOR operation to create the ciphertext, as shown in Figure 4. With each subsequent block, the counter is incremented, ensuring that every block of plaintext is encrypted using a distinct keystream. This process guarantees both confidentiality and uniqueness for each encrypted block.

## V. Implementation Design and Details

Our extended architecture consists of two parts: decryption of response from the exchange, which is done in software (SW) using OpenSSL toolkit (RX flow), and encryption of requests that are sent to the exchange using offloading inside hardware (TX flow). The integration points are shown in Figure 5. We can leave the RX flow in software to save FPGA resources since we are primarily targeting trading strategies for which the latency of the exchange responses isn't critical. Even so, we are currently working also on the RX flow hardware acceleration, to support the decoding of exchange responses directly in the FPGA for trading strategies that require it.

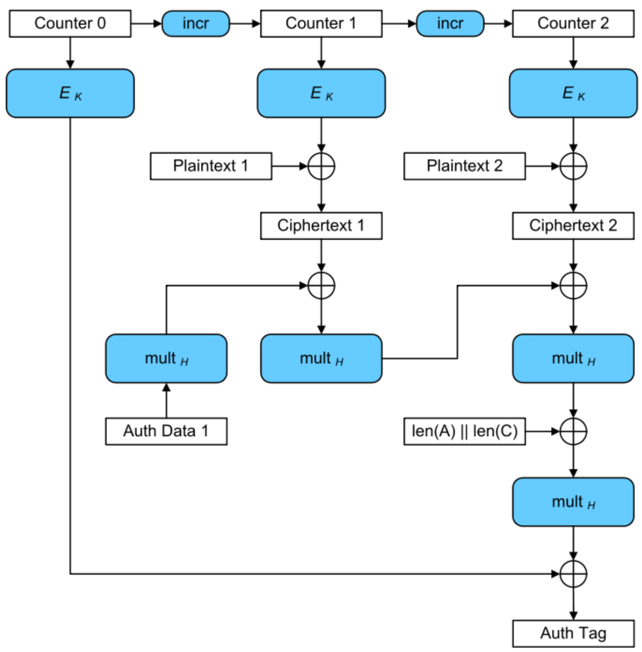



Fig. 3. AES GCM mode of operation [6]

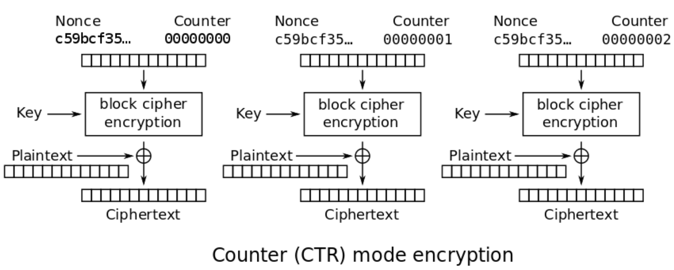



Fig. 4. AES CTR mode of operation [7]

Both flows are part of the same connection but in different directions, so they use separate AES counters. However, the key and IV negotiated during the connection establishment are the same.

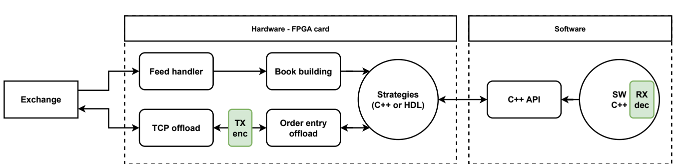



Fig. 5. Extended Magmio trading platform

Since the exchange does not use the GCM authentication tag as described in the section III-B, we can use AES in CTR mode of operation, described in IV-A. The architecture of CTR mode allows us to generate and store encrypted counter values ahead of time before they're needed to encrypt message data. Because TCP sessions work as a stream of data that is not aligned to AES block size, we need to implement advanced shifting and synchronisation to be able to provide the correct counter value and encrypt the data stream regardless of the transmitted message size.

Our approach includes a FIFO followed by an advanced shift register storing the encrypted counter values. AES encryption core works in a loop to process and encrypt the free-running counter values and store them into the FIFO and the shift register.

During the NSE login flow, the metadata (key and IV) required for the encryption is obtained and used to configure the AES encryption core. The initialization is finished before the first encrypted message (*BOX_SIGN_ON*) is required to be sent. The AES encrypted counter is running asynchronously/in parallel with the sending of encrypted messages to the network. This ensures that a sufficient number of new AES blocks will be ready in the shift register once the current message is transmitted, so we can immediately encrypt and transmit another message.

The final encryption process of the message is done by XORing the plain data of the message with the output of the shift register of the same width. This way, the whole process of encrypting the data consists of just one quick logic XOR operation over the given data width, as shown in Figure 6. The extra XOR operation might add no latency at all, depending on the timing margin of the whole FPGA design.

The size of the FIFO and the shift register is not unlimited, but it is set based on the requirements of the maximum length of the messages sent. The exact values are sixteen blocks for FIFO and ten blocks for shift register.

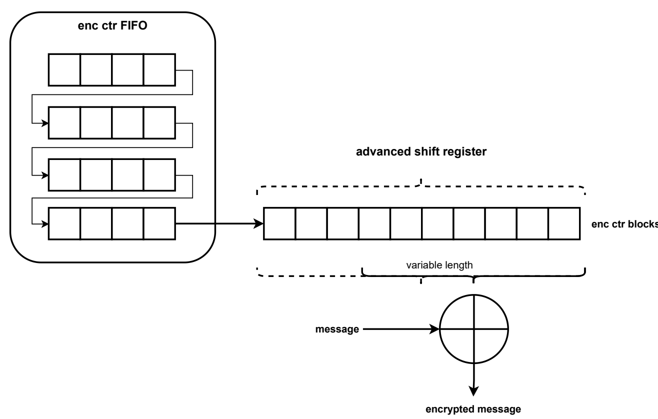



Fig. 6. FIFO and shreg illustration

### A. *aes_ctr_top*

The top component of the TX flow encryption part is *aes_ctr_top*, with an interface consisting of a generic parameter for setting the number of parallel AES blocks available in the output stream, an output stream interface for encrypted counter data, and feedback for information about the number of bytes used for the latest message encryption. There's also a bus for communication with software (MI32). Via this bus, the metadata required for encryption (like key and IV) are configured.

The top component provides interconnection between instantiated submodules such as *aes_ctr_fifo* (the free-running AES counter) and the generic number of *aes_block_shift* modules. The high-level architecture is shown in Figure 7.

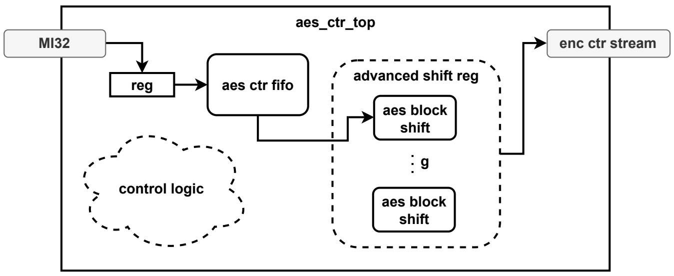



Fig. 7. *aes_ctr_top* high-level architecture

### B. aes_ctr_fifo

This component is a wrapper for the AES encryption core component providing free-running encrypted counter values. It incorporates a key expansion component together with the AES encryption core and generic FIFO component, as you can see in Figure 8. Only one AES encryption core component processing 128 bit blocks is used. It starts generating encrypted counter values once the AES IV and key are set. The encrypted counter output is stored into FIFO. The module is running asynchronously with processed data, trying to keep the encrypted counter FIFO always full. The FIFO output is connected to one of the *aes_block_shift* modules in the top level, as described in subsection V-A.

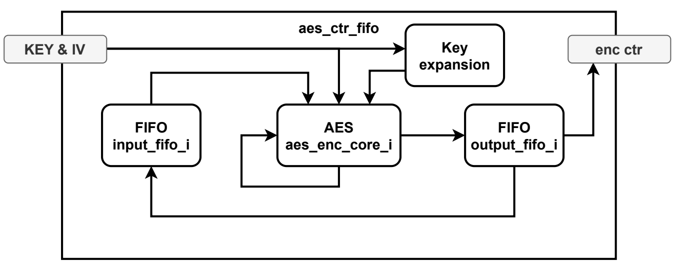



Fig. 8. *aes_ctr_fifo* component architecture

### C. Core components

We have incorporated fully-pipelined key expansion and AES core components, facilitating encryption in every cycle even amidst encryption key alterations. Additionally, these components offer flexibility in adjusting the number of registered rounds through a generic parameter *ENC_PIPELINES*, thereby balancing area and timing costs. The number of clock cycles for output based on the generic setting is shown in the Table I. Most processes within each round are easily implemented in hardware, primarily consisting of permutation and XOR operations. However, the SubBytes operation, being nonlinear, poses a critical challenge. To address this, we have utilized a substitution S-BOX table, efficiently stored in block RAM.

TABLE I
*ENC_PIPELINES* VALUE AND LATENCY

| ENC_PIPELINES | Latency [clock cycles] |
|---|---|
| 1 | 15 |
| 2 | 8 |
| 3 | 6 |
| 4 | 5 |
| 6 | 4 |

In our current architecture, we are using a setting of *ENC_PIPELINES* $= 1$ for better timing. A higher latency of the AES block doesn't impact the effective encryption latency thanks to the CTR mode, as explained in V.

### D. aes_block_shift

This component provides dynamic shifting of 16 B AES encrypted counter blocks by a given number of bytes. It internally uses generic MUXes to combine data from this block with data from another instance of the module. A high-level diagram of generic components and interconnection with the interface is shown in Figure 9.

Multiple instances of the *aes_block_shift* module are chained together in the top module, as described in the subsection V-A, forming a structure similar to a big shift register. In one clock cycle, it's able to move the whole 16 B AES counter block from one *aes_block_shift* instance to another.

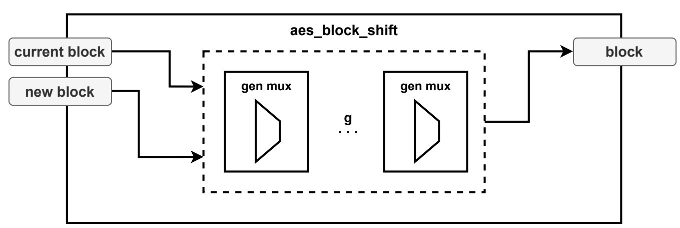



Fig. 9. *aes_block_shift* component diagram

To illustrate how the component works on the top level, consider a message of length 120 bytes being encrypted. Once the plain data are XOR'ed with the encrypted counter stream output, as shown in Figure 6, the feedback interface of the top module is used to inform the module about 120 bytes being read out from the shift register formed by *aes_block_shift* instances. In the following $\lceil 120/16 \rceil = 8$ clock cycles, the AES counter blocks are shifted between *aes_block_shift* instances while new blocks are being sourced from the *aes_ctr_fifo* to the upper *aes_block_shift* instance. After the 8 clock cycles, the whole shift register contains valid blocks. Since the NSE exchange uses 10G Ethernet, the sending of an encrypted message to the network takes more than 8 clock cycles, so once the transmission is done, the module is ready to encrypt another message immediately.

## VI. Performance Evaluation

To assess the impact of the proposed AES-GCM encryption on the Magmio trading platform, we conducted a series of benchmarks comparing system performance before and after optimization.

The latency benchmark measures the time taken for transactions to be executed with and without the proposed encryption. The goal is to quantify the overhead introduced by encryption and evaluate the effectiveness of our optimization techniques in minimizing this overhead.

## VII. Conclusion

This paper presents a solution leveraging FPGA technology to achieve swift AES encryption suitable for High-Frequency Trading applications. It specifically addresses the imperative of secure communication with stock exchanges, notably highlighting the National Stock Exchange of India. Emphasizing the crucial balance between maintaining rigorous security standards and maximizing speed, the paper explores inventive architectural designs and hardware acceleration techniques.

In addition to discussing AES in GCM mode, the paper also delves into the utilization of Counter mode due to the absence of an authentication tag requirement from the exchange. In this mode, data is encrypted in a stream-like manner, employing encrypted counter values that can be precomputed. This strategic choice enables efficient encryption with a latency of at most 1 clock cycle, which is a key performance metric of HFT systems.

By providing practical insights into managing the delicate equilibrium between security and low latency, this paper offers valuable guidance for the development of secure and high-performance trading systems within the dynamic landscape of HFT.

## References


[1] *Trimmed Protocol for Non-NEAT Front End (NNF): Futures and Options Trading System*. Version 9.38. 2023. Online: https://nsearchives.nseindia.com/web/sites/default/files/inline-files/TP_FO_Trimmed_NNF_PROTOCOL_9.38_0.pdf

[2] *Trimmed Protocol for Non-NEAT Front End (NNF): Capital Market Trading System*. Version 5.5. 2023. Online: https://nsearchives.nseindia.com/web/sites/default/files/inline-files/TP_CM_Trimmed_NNF_PROTOCOL_5.5.pdf

[3] *Members FAQ's on: "Encryption of Interactive Messages"*. Version 1.2. 2023. Online: https://nsearchives.nseindia.com/web/sites/default/files/inline-files/Annexure%204%20-%20FAQs%20related%20to%20Encrypted%20Data%20Flow.pdf

[4] *Download Ref No: NSE/MSD/59984*. 2023. Online: https://nsearchives.nseindia.com/web/sites/default/files/inline-files/MSD59984.zip

[5] *High-frequency trading - magmio: Product overview*. Online: https://www.magmio.com/product

[6] gcm - IV in Galois Counter Mode - Cryptography Stack Exchange. Online: https://crypto.stackexchange.com/questions/56188/iv-in-galois-counter-mode

[7] Cipher Block Modes - Practical Cryptography for Developers - CTR (Counter) Block Mode. Online: https://cryptobook.nakov.com/symmetric-key-ciphers/cipher-block-modes#ctr-counter-block-mode

# Security Analysis of a Commercial Quantum Key Distribution System

Ondrej Klicnik
*Department of Telecommunications*
*Brno University of Technology*
Brno, Czech Republic
ORCID 0000-0001-5903-7877

Petr Munster
*Department of Telecommunications*
*Brno University of Technology*
Brno, Czech Republic
ORCID 0000-0002-4651-8353

***Abstract*—This research paper investigates the practical challenges in implementing Quantum Key Distribution (QKD) systems, focusing on bridging the gap between theoretical security guarantees and the complexities of real-world deployment. It emphasizes the importance of addressing system vulnerabilities, enhancing protocol integrity, and ensuring robust security measures to advance the reliability and effectiveness of quantum communication technologies.**

***Index Terms*—attacks, classical algorithms, quantum key distribution, security analysis**

## I. Introduction

Although quantum cryptography theoretically offers unconditional security, it is important to remember that system is not just the Quantum Key Distribution (QKD) protocol itself. The practical system is not nearly as perfect as the mathematical model. In practice, this means that it consists of components with only limited capabilities and can thus be exploited for attacks. Furthermore, a practical system depends on a number of classical algorithms whose security can only be computational [1].

Since the analysis was performed on ID Quantique's commercial and proprietary Clavis$^3$ system, it is not always possible to obtain information about how its non-quantum components work internally. For example, how keys are stored in the Key Management System (KMS) and then securely deleted again. Thus, this paper is only concerned with the accessible and known classical components of this system [1].

## II. Clavis$^3$ system

Clavis$^3$ is a quantum key distribution system from the Swiss company ID Quantique (IDQ). It is a basic short-range platform with both three-state and four-state versions of the Coherent One-way (COW) protocol. Since it is a research platform, the advantage is mainly the ability to monitor and modify the basic parameters of the QKD system using command line tools [2].

The main part of the QKD Network (QKDN) topology are two QKD servers. However, the two devices are different and have different purposes. While Alice serves mainly as a quantum signal transmitter, Bob is the receiving end of the pulses. In the Figure 1 below, the most important parts of both devices are highlighted. The quantum channel interface is fitted with green Ferrule Connector / Angled Physical Contact (FC/APC). For proper operation of the device, a minimum of 10 dB attenuation is required on the quantum channel. For this reason, it may be necessary to add an attenuation cell immediately after the connector. According to the manufacturer, the maximum possible attenuation on the path is 14 dB. However, measurements have shown that both higher and significantly lower path attenuation typically result in lower key delivery rates. However, the system remains operational [2].

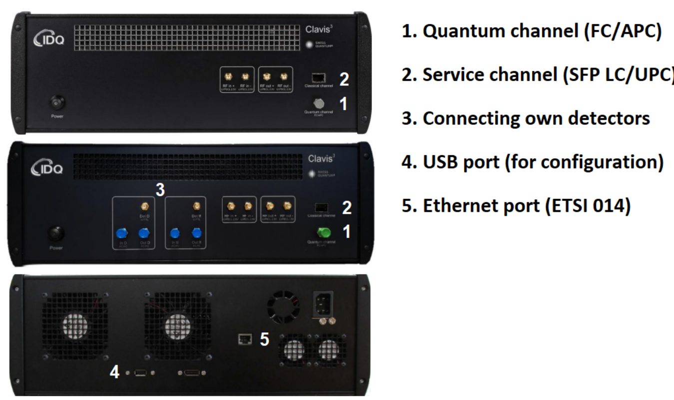



Fig. 1. The difference between Alice and Bob QKD servers.

The system provides slots for Small Form-factor Pluggable (SFP) modules for connecting service channels. In this case, a pair of blue Lucent Connectors / Ultra Physical Contact (LC/UPC) is used for duplex communication. However, this channel can theoretically run on any media. It does not have to be optical media only. What distinguishes Bob from Alice at first glance is the presence of ports for connecting custom data and monitoring detectors [2].

The back panel is identical for both devices and contains an Ethernet port for the network communication, Key Management link and links connected to consumers (encryptors). A Universal Serial Bus (USB) port is also present, whose main function is access for uploading Transport Layer Security (TLS) security certificates with which these links are secured [2].

## III. QUANTUM CHANNEL

Selected system parameters are listed in the Table I below. Even though this may not be the recommended solution, the research version also allows some system parameters to be changed. However, only the default parameter values will be considered in this paper. Among the most important is the mean photon number. This follows the Poisson distribution and states how many photons are on average emitted per pulse [3].

$$P(X=x)=\frac{\mu^x e^{-\mu}}{x!} \tag{1}$$

TABLE I
IMPORTANT SYSTEM PARAMETERS CLAVIS$^3$.

| Important system parameters | | | |
|---|---|---|---|
| **Manufacturer** | — | ID Quantique (IDQ) | |
| **QKD protocol** | — | Coherent One-way (COW) | |
| **Pulse generation rate** | $v$ | 1.25 GHz | |
| **Key rate** | — | 1.4 kbps | |
| **Dynamic range** | — | 10–14 dB | |
| **Mean photon number** | $\mu$ | 0.03 | |
| **Quantum channel** | $\lambda$ | 1551.72 nm | DWDM 32 |
| **Service channel I** | — | 1553.33 nm | DWDM 30 |
| **Service channel II** | — | 1554.13 nm | DWDM 29 |

One of the most important parameters is the power of the quantum channel. This can be calculated from the parameters above. First, the energy of a single photon must be calculated. This is a function of the quantum channel wavelength.

$$E=\frac{hc}{\lambda}=1.28\cdot 10^{-19}\ J=0.8\ eV \tag{2}$$

Since the pulse generation rate and the mean photon number are known, the total channel power can now be now calculated using these and the previously obtained photon energy.

$$P=Ev\mu=4.8\cdot 10^{-12}\ W=-83.2\ dBm \tag{3}$$

The outcome shows that the channel only transmits at very low power. Therefore, it requires special treatment. Otherwise, the channel could be absorbed by noise from other lines or affected by nonlinear phenomena. In particular, Raman noise. For this reason, it is necessary to consider the appropriate positioning of the channel and to ensure that it is sufficiently filtered. The calculated values of these parameters can be found in Table II.

TABLE II
THE CALCULATED SYSTEM PARAMETERS CLAVIS$^3$.

| The calculated system parameters | | |
|---|---|---|
| **Photon energy** | $E$ | 0.8 eV |
| **Power of the quantum channel** | $P$ | –83.2 dBm |

## IV. QKD PROTOCOL SECURITY

The first ever proposed QKD protocol is BB84, which is originally based on the necessity of using a single-photon source. These would work in a photon-on-demand fashion. However, such sources are not used in practice. Instead, Weak Coherence Pulses (WCP) are used, with a source laser attenuated to less than the single-photon level. Thus, some pulses contain a photon while others do not. This ratio is determined by the average photon number described earlier [4].

However, the use of these sources carries risks, since in some cases a pulse may contain two or more photons. This is exploited by the so-called Photon Number Splitting (PNS) attack, where eavesdropper keeps one of the photons and forwards the rest. The attacker can then make a measurement over the captured photon without being detected. The protection against this is the so-called decoy states. These are used in practice in many implementations of various protocols [5].

While the security of some protocols is mathematically proven, for other protocols unconditional security is so far only intuitive. The COW protocol, used in the Clavis$^3$ system, is a case in point.

This protocol exists in two basic versions. The original, so-called three-state protocol is theoretically vulnerable to Unambiguous State Discrimination attacks (USD). For this reason, an extended four-state version has been developed, which adds one more type of decoy state to the two signal states and one original decoy state. This protocol is more secure but slightly slower than previous version. The Clavis$^3$ system supports both versions to be used [6].

## V. RANDOM NUMBER GENERATION

Another thing to consider is the randomness of the generated key. If an attacker was somehow able to predict the key, the security of the entire system would be severely compromised. For such a reason, it is necessary to choose the right number generators. The standard Pseudorandom Number Generator (PRNG) or most of the so-called True Random Number Generator (TRNG) based on chaos theory are not suitable candidates. While physically complex, it is still a deterministic operation. And for that reason, it can still be considered somewhat predictable. Similarly, however, truly non-deterministic generators that do not have a uniform distribution may not be suitable [7].

In the field of QKD, the so-called Quantum Random Number Generation (QRNG) is most often used. This can be imagined simplistically as a perfect 50:50 coupler. In this way, a photon passes through one of the paths completely at random. Similar generators (although much more complex) already exist today in chip form and are deployed as standard in QKD systems, including Clavis$^3$ [7].

## VI. ROBUSTNESS AGAINST FIBRE MANIPULATION

The original QKD protocols were based on encoding the key value into two polarization bases. However, in fiber-based systems, the polarization of the transmitted signal can change rapidly, since it is dependent on the fiber geometry. Thus, in the event of a change, the entire quantum channel may be disrupted or slowed down. These problems are then addressed by algorithms that compensate for polarization changes [8].

A convenient alternative are phase-based QKD schemes that are already independent of this problem and changes in the geometry of the fiber do not affect the quantum transfer.

## VII. Classical Algorithms

In this context, classical algorithms can be thought of as mainly distillation functions such as authentication (based on preshared key and hash), error correction process (based on Low-density Parity-check code – LDPC) and subsequent compression (based on hash). Unlike the quantum algorithm, these functions already possess only computational security [9].

For topologies more complex than point-to-point, key forwarding already occurs between trusted devices at the KMS level. One-time pad (OTP) is typically used for this purpose, which is inherently unbreakable when combined with QKD. This is especially necessary because the key is forwarded outside the secure zone (the area near the device where an attack is not expected).

The communication between the QKD system and the key consumer (any cryptographic application – encryptor) within the secure zone is based on the ETSI 014 standard. Here, however, the communication is protected only on the basis of standard TLS certificates, which do not even need to contain Post Quantum Cryptography (PQC) algorithms yet [10].

Configuration using the QNET tool is similar. The other option is to use direct device configuration. This is done using the Secure Shell (SSH) protocol, most implementations of which currently have only the current asymmetric schemes.

## VIII. System Management

The entire Clavis[3] system is based on Linux operating system. However, the network administrator does not communicate directly with the operating system, as he only has a limited QNET shell at his disposal. It can be accessed using the SSH protocol, and offers only a very limited number of commands. A privileged account (idq) is required for more detailed system administration, including modification of most transmission parameters. However, this is only available to the manufacturer, not the end user (owner) of the system.

The whole QKD network (QKDN) can also be managed in a centralized manner using the QNET tool. This is a tool that communicates with nodes using an API based on the REST architecture. A possible disadvantage of this solution is that user login credentials (their names and password hashes) are stored in the *~/.qnet/default* file and are thus freely accessible within the user's environment. The structure of the file can be found in Figure 2. The tool works in such a way that when it is first started, the corresponding QNET tool user is configured, whose login credentials are stored in the file mentioned above. This user is then always automatically used when communicating with the QKD servers. It is therefore necessary to take into account that access to the central configuration of the entire QKDN is based on the security of the Linux user in question.


```
idq@idq-virtual-machine:~/.qnet$ cat default
{
  "ConfigWebApi": {
    "Url": "https://localhost:443",
    "AuthenticationScheme": "basic",
    "OAuthTokenServiceUrl": null,
    "Username": "Admin",
    "Password": "+os2TWkkgBrc4mHNxU4uKnRnCm9BYPH7ed6H43zxXHI="
  },
  "AdminWebApi": {
    "Port": 8443,
    "AuthenticationScheme": "basic",
    "Username": "kms",
    "Password": "tkiQuiMaWWcojdcQXSh0827sYTVHHme+7MDP8BG7Ab8="
  },
  "IdentityServer": {
    "Url": null,
    "AuthenticationScheme": "basic",
    "OAuthTokenServiceUrl": null,
    "Username": null,
    "Password": null
  },
  "Staging": false,
  "OutputFormat": null,
  "Language": 0,
  "QncWebApis": [
    {
      "name": "QNCA",
      "uid": "4a739570de2848d89cd75432dd730a6c",
      "url": "https://192.168.10.102:8443/",
      "scheme": "basic",
      "username": "kms",
      "password": "vQxh6hY1kvnkFNV5cr2cudsjjjm9jg3VW0IB0+tiG0c=",
      "lastSignIn": "2023-05-13T01:10:20.0436855Z"
    },
    {
      "name": "QNCB",
      "uid": "11ef6b184cd04648bc63c900993faf7c",
      "url": "https://192.168.10.107:8443/",
      "scheme": "basic",
      "username": "kms",
      "password": "gmCX0TLbDET67ORQcBax9pka739jB4y4VRSOereuFDg=",
      "lastSignIn": "2023-05-13T01:16:25.8412444Z"
    }
  ]
```


Fig. 2. A listing of a file containing information about users and QKD servers.

Both of these options are for device configuration only and do not offer any access to the keys stored in the Key Management System. Thus, until a software vulnerability is found in these tools, this part of the system can be considered secure.

## IX. Classical Channels

Although it is theoretically possible to use any transmission medium for the classical channel, the current state of the system only allows optical links. At the same time, it must be taken into account that the length of the quantum and service channel can differ by a maximum of 30 km. The permitted difference between the two service channels is not precisely specified by the manufacturer, but should only differ minimally. Otherwise, unexpected communication failures may occur. The classical channel also does not communicate using a standard set of protocols and thus cannot be easily routed over common IP networks.

Practically, the service channel is implemented using a set of two SFP modules, whose selected parameters are listed in Table III below.

The differences in length and properties of the SFPs used can be exploited for both temporary, and permanent DoS attacks. An attacker can target both the quantum and service channels.

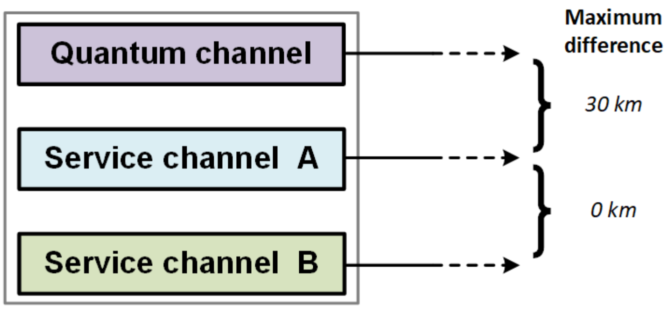



Fig. 3. Maximum allowed difference between channels of the QKD system.

TABLE III
PARAMETERS OF FINISAR AND SKYLANE SFP MODULES [11], [12].

| Selected parameters of SFP | | |
|---|---|---|
| **Manufacturer** | Finisar | Skylane |
| **Model** | FWLF1632xx | SPDTU080100D |
| **Source** | | |
| **Rate** | 2.7 Gbps | 2.7 Gbps |
| **Wavelength (CH29)** | 1554.13 nm | 1554.13 nm |
| **Wavelength (CH30)** | 1553.33 nm | 1553.33 nm |
| **Output power (measured)** | –1 to 1 dBm | –3 dBm |
| **Detector** | | |
| **Input power** | –28 to –9 dBm | –24 to –7 dBm |

- Channel blinding – An attacker sends a high power signal to the channel that obscures the useful signal or saturates the photodetector.
- Damage to the detector – If the transmitted power exceeds the maximum the detector may be permanently damaged.
- Channel attenuation – As power decreases, the error rate of the service channel may become out of sync. Similarly the quantum channel can also be disabled.
- Channel length difference – An attacker can redirect one of the channels to a route to a different length channel and thereby disrupt communication.

## X. PRE-SHARED KEY

While the logical protocol only addresses the key exchange itself, practical QKD systems require other security mechanisms to be involved. Thus, one of the biggest drawbacks of quantum key distribution is the need for authentication of both parties, which is not addressed by these protocols. The latter is most often addressed by means of pre-shared secrets. This paradoxically creates a system for key exchange that already needs a preshared key to function.

It is thus clear that much of the security lies in the ability to keep the key a secret. Moreover, it is gradually renewed from newly generated keys. These keys are referred to as Initialization Secret Keys (ISK) and are used each time a system connection is established (re-authentication). If the system initialization is not successful, the used ISK cannot be reused. For this reason, devices maintain a set of multiple keys. Thus, if the keys are exhausted before the key exchange connection is established, new keys must be manually added to the system. This can occur, for example, if authentication is successful but the quantum or classical channel subsequently fails to synchronize [13].

In this case, the problem of how to transfer the new key to the remote device arises. Although it is now possible to upload the key over the network, the most common method is to transfer the key on a USB drive. This can be considered secure, but hardly usable if the two drives are far apart.

## XI. ATTACK SIMULATION

In addition to the QKD servers mentioned above, it is also possible to plug an attack simulator into the quantum channel. The device or Eve does not implement any specific attack. Rather, it is an attack simulator whose action produces similar effects in the system as a real attack.


Fig. 4. Eva's quantum channel eavesdropping device.

The idea behind it is an adjustable coupler that separates a small part of the pulses and delays them by a fixed number of positions. By having the pulses arrive at the wrong time, the error rate increases and the transmission rate decreases. In contrast, changes in interferometric visibility are minimal. Eve is equipped with a rotary potentiometer that is used to adjust the amount of delayed pulses. As the amount of pulses grows, so does the QBER. The system responds by increasing compression ratio, which reduces the key rate. The quantum channel is connected to the Eve using FC/APC connectors.

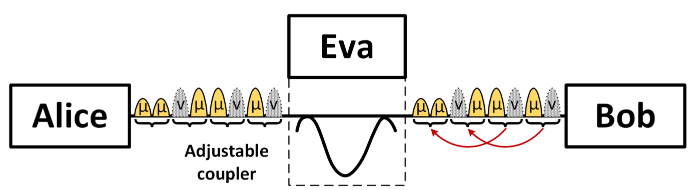



Fig. 5. The principle of the Eva device.

The principle of the simulation is to rotate the potentiometer by 5° each time, thus gradually increasing the amount of delayed light. The total path attenuation is around 11 dB. The module itself has an insertion loss of around 3.5 dB. According to the manufacturer's instructions, there should be a noticeable increase in the error rate and a decrease in the key rate until about 20°. Furthermore, the system should stop generating keys when the QBER reaches a threshold of about 6 %. This occurs when the potentiometer is set to 35°. The potentiometer can only be set to a maximum of 60° otherwise there is a risk of damage to the module. The minimum possible visibility limit is considered to be 95 %. The measurement results for each angle can be found in Table IV.

TABLE IV
IMPACT OF THE ATTACK ON THE CLAVIS$^3$ SYSTEM.

| Angle | Key rate | QBER | Visibility |
|---|---|---|---|
| 0° | 1.78 kbps | 2.28 % | 98.50 % |
| 5° | 1.67 kbps | 2.29 % | 98.40 % |
| 10° | 1.60 kbps | 2.41 % | 98.50 % |
| 15° | 1.62 kbps | 2.40 % | 98.30 % |
| 20° | 1.61 kbps | 2.68 % | 98.50 % |
| 25° | 1.51 kbps | 3.12 % | 98.20 % |
| 30° | 0.30 kbps | 4.91 % | 97.00 % |
| 35° | 0 kbps | 5.96 % | 97.40 % |
| 60° | 0 kbps | 29.20 % | 87.50 % |

### A. Results

The delivered module corresponds in its characteristics to the parameters described by the manufacturer. The simulation also confirms the assumed 6% limit for safe transmission. The measurements show that the attack has a much lower impact on the visibility than on the error rate. This may be due to the fact that the mutual coherence of the two compared pulses is violated in fewer cases. However, as the last row of the Table IV shows, the attack also results in a reduction in visibility for larger amounts of delayed light. Except for this last (extreme) measurement, however, visibility never dropped below 95 %.

## XII. CONCLUSION

In conclusion, this paper has critically examined the Clavis$^3$ Quantum Key Distribution system by ID Quantique, highlighting the disparity between theoretical security promises and practical vulnerabilities. Through comprehensive analysis, it was found that while QKD offers a solid foundation for secure communication, the practical deployment is still full of challenges. These include susceptibility to various attacks, the necessity for robust random number generation, and effective management of pre-shared keys. The document also includes a demonstration of the attack attempt and captures the system's response to it (higher QBER and lower key rate).


## ACKNOWLEDGMENT

This work is supported by the Ministry of the Interior of the Czech Republic, program IMPAKT1, under grant VJ01010008, project Network Cybersecurity in the Post-Quantum Era.


## REFERENCES


[1] Pereira, F. R. A., & Amelia, M. (2024). Implementation security in quantum key Distribution. Advanced Quantum Technologies. https://doi.org/10.1002/qute.202300380

[2] Products. (2023, March 2). ID Quantique. https://www.idquantique.com/quantum-safe-security/products/

[3] Engle, R. D., Mailloux, L. O., Grimaila, M. R., Hodson, D. D., McLaughlin, C. V., & Baumgartner, G. (2017). Implementing the decoy state protocol in a practically oriented Quantum Key Distribution system-level model. The Journal of Defense Modeling and Simulation, 16(1), 27–44. https://doi.org/10.1177/1548512917698053

[4] Bennett, C. H., & Brassard, G. (2014). Quantum cryptography: Public key distribution and coin tossing. Theoretical Computer Science, 560, 7–11. https://doi.org/10.1016/j.tcs.2014.05.025

[5] Gaidash, A., Egorov, V. I., & Gleim, A. (2016). Revealing of photon-number splitting attack on quantum key distribution system by photon-number resolving devices. Journal of Physics: Conference Series, 735, 012072. https://doi.org/10.1088/1742-6596/735/1/012072

[6] Gao, R., Xie, Y., Gu, J., Liu, W., Weng, C., Li, B., Yin, H., & Chen, Z. (2022). Simple security proof of coherent-one-way quantum key distribution. Optics Express, 30(13), 23783. https://doi.org/10.1364/oe.461669

[7] Quside. (2023, November 1). Quantum Random Number Generator (QRNG) - explained. Quside. https://quside.com/quantum-random-number-generators-why-how-where/

[8] Wu, G. (2006, June 13). Stable polarization-encoded quantum key distribution in fiber. arXiv.org. https://arxiv.org/abs/quant-ph/0606108

[9] Constantin, J., Houlmann, R., Preyss, N., Walenta, N., Zbinden, H., Junod, P., & Burg, A. (2015). An FPGA-Based 4 Mbps secret key distillation engine for quantum key distribution systems. Journal of Signal Processing Systems, 86(1), 1–15. https://doi.org/10.1007/s11265-015-1086-1

[10] Dahmen-Lhuissier, S. (n.d.). QKD. ETSI. https://www.etsi.org/committee/1430-qkd

[11] Electronic Components Distributor - Mouser Electronics Europe. (n.d.). Mouser Electronics Europe. https://mouser.com/

[12] Skylane Portal. (n.d.). https://portal.skylaneoptics.com/

[13] Wang, LJ., Zhang, KY., Wang, JY. et al. Experimental authentication of quantum key distribution with post-quantum cryptography. npj Quantum Inf 7, 67 (2021). https://doi.org/10.1038/s41534-021-00400-7

# Design of Broadband over Powerline Modem

1st Michal Mahút
*Department of Telecommunications*
*Brno University of Technology*
Brno, Czech Republic
xmahut00@vut.cz

2nd Lukáš Benešl
*Department of Telecommunications*
*Brno University of Technology*
Brno, Czech Republic
xbenes44@vut.cz

3rd Martin Rusz
*Department of Telecommunications*
*Brno University of Technology*
Brno, Czech Republic
xruszm00@vut.cz

4th Petr Musil
*Department of Telecommunications*
*Brno University of Technology*
Brno, Czech Republic
musilp@vut.cz

5th Petr Mlýnek
*Department of Telecommunications*
*Brno University of Technology*
Brno, Czech Republic
mlynek@vut.cz

***Abstract*—This article introduces design of affordable BPL modem for measuring noise levels on power lines and ensuring reliable communication. Utilizing the MOD957 module from Xingtera, the modem exhibits resilience to high noise levels, crucial for maintaining communication integrity. Throughput tests demonstrate improved performance even under noisy conditions. The article underscores the importance of noise measurement, especially in ensuring dependable communication and controlling the balance of the power grid, as well as detecting potential faults in the power grid.**

***Index Terms*—BPL modem, noise measurement, PLC, power grid, power line communication, throughput testing**

## I. Introduction

Transmission of data over power lines finds wide application across various industries. Broadband transmission high–speed internet over existing electrical grids is being utilized within household settings [1]. In the energy sector, it is employed for narrowband data transmission to monitor, control, and detect faults in the distribution network, ensuring reliable electricity distribution [2]. In the area of the Internet of Things (IoT), it serves as alternative means of data transmission from sensors [3]. The automotive industry employs data transmission through power lines for control systems within the vehicle. The broad utilization of PLC (Power Line Communication) technology is expected in the coming years, particularly in communication between electric vehicles and charging stations [4].

### A. PLC technology

Communication utilizing modulation of high–frequency signals onto the 50–60 Hz mains voltage network has been known since the early 20th century when it used to be for transmitting signaling data [5]. Power Line Communication offers numerous advantages, such as cost–effectiveness of the technology and the capability for reliable transmission over greater distances without the need to build new communication infrastructure. In addition to wired and optical lines, PLC strongly competes with wireless technologies, which are limited by interference and obstacles.

### B. Division of PLC Communication

PLC technology is divided according to the utilized bandwidth into:

- *Narrowband - NB PLC* utilizes a narrowband with frequencies in the hundreds of kHz range and is characterized by low transmission rates, reaching a maximum of 1 Mbit/s. The primary purpose of NB PLC is the transmission of signaling data. In the European region, it is standardized by EN 50065-1 issued by the European Committee for Standardization (CENELEC), defining the frequency range of NB PLC between 3 and 148,5 kHz [18], [19]. Depending on the quality of the transmission path, NB PLC enables transmissions ranging from hundreds of meters to several kilometers. Each region has its own definitions of operational frequency ranges. In the United States, the US FCC band is defined in the range of 10–490 kHz, in the Japanese region, it is the AIRB band in the range of 10–450 kHz, and in China, it is the range of 3–500 kHz [20]. Standards in the field of narrowband powerline communication include PRIME, G3-PLC, ITU-T G.hnem, IEEE 1901.3.
- *Broadband - BB PLC* utilizes significantly wider bandwidth compared to NB PLC, which can extend up to 500 MHz depending on regional standards. The achievable throughput of this technology reaches up to 1 Gbit/s. However, the communication distance is considerably shorter than that of narrowband technology, limited to a maximum of 300 m [21]. ITU-T recommendations in the G.hn standard specify the bandwidth over power lines in the range of 2 to 100 MHz [22]. Examples of BB PLC standards include HomePlug AV2, HomePlug Green Phy, HDPLC, ITU-G.hn and IEEE 1901.

### C. Applications of PLC

- The use of *broadband communication* in domestic scenarios enables user–friendly distribution of data connectivity throughout the entire premises. In addition to the advantage of utilization without the need for building new infrastructure, we recognize further benefits in the easier integration of smart appliances, incorporation of renewable energy sources, and their control elements [6]. Modern home BPL (Broadband over Powerline) modems implementing MIMO (Multiple Input Multiple Output) techniques provide real data throughput of around 1 Gbit/s. They offer a reliable way to build high–speed home networks and allow users to connect end devices using LAN and WiFi connectivity [7], [8].

- *SmartGrids* utilize PLC technology in multiple aspects to ensure reliable distribution of electrical energy and optimize the utilization of energy resources. The current trend of connecting a large number of renewable sources to the electrification system requires increased demands for controlling and regulating the distribution network [9]. The current trend of integrating a large number of renewable sources into the electrification system requires increased demands for controlling and regulating the distribution network. A prerequisite for a modern distribution system is the implementation of an advanced metering structure, AMI (Advanced Metering Infrastructure), providing online access to measurements conducted at regular time intervals to the control center. The advantage of data collection is the ability to balance energy demand in specific network segments and identify instances of energy theft [14]. The investigation of using broadband (BB) and narrowband (NB) PLC technology for automated meter reading of AMI meters is addressed in [10]. Communication between the concentrator and end smart meters is primarily affected by noise introduced into the network from household loads [11]. Enhancing the performance of PLC technology in AMI infrastructure is discussed in [12], [13].
- The *automotive industry* utilizes PLC communication for data exchange within the vehicle or for communication between the electric vehicle and the charging station. In–vehicle data transmission systems offer the advantage of significantly reducing the amount of wiring harnesses routed inside the electric vehicle. PLC systems in this usage scenario are influenced by strong impulse noise, which may affect the reliability of data unit reception [15]. The authors of the article [16] propose an optimization scheme providing effective protection against impulse interference. PLC communication finds extensive application in data exchange between the electric vehicle and the charging station via the power cable. The increasing number of electric vehicles (EVs) necessitates the construction of unmanned EV charging stations (EVSE) capable of real-time communication with EVs. The advantage of implementing a PLC system lies in its low cost and reliability compared to CAN buses and wireless technologies [17].

## II. Hardware design

The motivation for constructing our own BPL modem is its integration into the AMI infrastructure. Primarily, it serves as a noise measurement device in the electrical grid, generated by active devices in the electrical grid, and secondarily, as a reliable communication component in channels affected by high levels of noise.

When selecting a suitable BPL module capable of implementing BPL communication and applicable to PCB (Printed Circuit Board) design, emphasis was placed on the availability of development kits, software support, and accessible documentation for integration.

### A. Selecting a suitable BPL module

Selection of a suitable BPL module involved examining and testing the options of three development kits. Each of the development kits supports a different BPL communication standard.

The PLC Stamp 1200 Ev. Kit operates on the HomePlug AV2 standard and supports connection to L, N, PE conductors, enabling communication in both SISO and MIMO modes. The main drawback of this solution is the insufficient documentation and weak software support for reading parameters of the electrical network noise level. The PLCtool tool only allows reading the transmission tone mask.

Two solutions from XINGTERA were tested. All tested development kits provide the same security using symmetric block cipher AES-128. The XTV3031 development kit allows connection of all three phase conductors. Both development kits from this company allow plug-and-play solutions. BPL modules have implemented communication parts via UART and Ethernet directly on the PCB module. The module also includes a signal binding part to the power distribution. The manufacturer provides good documentation for the XTV3031 kit. Software support is at a good level and would meet the requirements for reading the noise level, but the option to export these data is missing.

For the design of the BPL modem, the MOD957 module was selected, which is part of the XTV6000 development kit. The module operates in the range of 2 to 80 MHz with the G.hn standard. The module is designed with a focus on low consumption, maximum 2,3 W. The software solution provided by the manufacturer, program XCT3.1_Rel, allows reading data such as SNR_PROBE, SNR_DATA, PSD_RX, and most importantly NOISE for our application. All measured values can be stored in .csv files with very good resolution of 3202 samples for the entire operating spectrum with a width of 78 MHz.

### B. Block diagram of BPL modem

Design of the PCB modem is based on the reference design of the BPL module MOD957 provided by the manufacturer. The block diagram of the circuitry is illustrated in Figure 1.

The power block includes the connection of the PCB to the 230 VAC power supply. The design incorporates an EMI filter located before the power supply unit converting

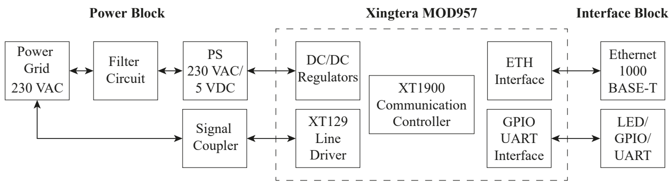



Fig. 1. Block diagram of the proposed BPL modem.

230 VAC to 5 VDC. The AC/DC power supply was chosen with consideration for the quality of its own filtering circuit to minimize the transmission of interference to the coupled high-frequency signal through coupling capacitors into the power supply grid.

Behind the MOD957 module, there is a signal interface block. This block primarily ensures smooth signal propagation between the Ethernet chip on the module and the RJ45 connector. GPIO outputs are used to indicate states using LED diodes and to operate the module in pairing/reset states. The module allows connection of the UART interface, which is routed on the PCB.

### *C. Power Block*

The power circuit consists of two parts. The first functions as a filtering circuit and is composed of a combination of power inductors and resistors. It filters signals around 20 MHz. It includes a reverse protective fuse for currents greater than 2 A. The second part consists of a protective capacitor and a choke coil. This part filters signals around 50 MHz. Surge protection is provided by a varistor. To prevent the generation of excess noise that could disrupt BPL communication, the PSI1 RAC05-05SK power supply is chosen, which meets noise production limits. The power supply features small dimensions and can deliver a current of up to 1 A, providing sufficient current reserve for potential implementation of other peripherals on the PCB. Since the MOD957 module itself includes a coupling component for high-frequency signals, this signal is differentially adapted when connected to the terminals of the mains voltage using coupling capacitors.

### *D. Xingtera MOD957*

BPL module MOD957 includes two connectors. The first one is used to connect the high-frequency signal to the terminals of the mains voltage. The second one outputs pins for the Ethernet chip, UART communication, LED power status indication, PLC communication, ETH communication, and pairing/reset functionality. The module is equipped with a DC/DC converter from 5 VDC to 3,3 VDC to power the LED diodes.

### *E. Interface Block*

Interface block consists of two parts. The first part connects the Ethernet chip to the RJ45 connector, and the second part is for connecting GPIO indicator pins and UART communication. The RTL8221F Ethernet chip mounted on the module is connected differentially and length-matched through TVS diodes to the RJ45 connector with filtering transformers. When selecting TVS diodes, emphasis was placed on their capacitance value to ensure very fast transmission between the ETH interface and the RJ45 connector. The TVS diodes also contain additional resistors to absorb any unwanted impulses. In the RJ45 connector layout, additional parallel combinations of capacitors were also considered to smooth the signal waveform.

## III. PCB MANUFACTURING AND CASE DESIGN

In the previous chapter, the PCB design was described. This chapter details the manufacturing process of the PCB and the enclosure design. The final version of the PCB design was preceded by the construction of a prototype v0.1, on which the filtering section of the power circuit was fine-tuned, and the protective part of the ETH connection was modified. During the production of the final prototype v1.1, emphasis was placed on minimizing size. The final version of the BPL modem PCB design v1.1 is shown in Figure 2. Most of the components in the power supply section consist of through-hole technology (THT) components. In this section, it was not possible to save much space due to the size of the THT components themselves and the need to maintain safe distances from low-voltage circuits. This is achieved by an air-insulated gap between conductive parts, known as clearance, and a drilled separation hole, known as creepage, which determines the minimum distance between two conductive parts across the surface of the board. The sizes of low-voltage components were chosen in the SMD (Surface Mount Device) size, with a size of 603 for resistors and 1206 for capacitors.

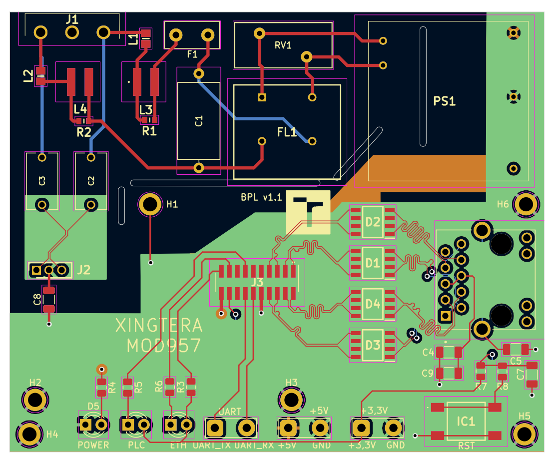



Fig. 2. Final PCB design of BPL modem v1.1.

The size of the final prototype v1.1 was reduced to 85 × 70 mm. The design is divided into two logical parts: a black high-voltage section and a green low-voltage section. These parts are appropriately separated from each other. The PCB includes a UART interface for debugging purposes and additional power connectors at the bottom for +5 V and +3, 3 V, allowing the connection of additional peripherals, such as a microcontroller to processing noise level data. Four mounting holes are placed on the PCB for attachment to a 3D enclosure.

The PCB is made using 4-layer technology. The first layer contains the routing of most high-voltage, low-voltage, and signal paths. The second layer consists of a grounded plane (GND), and the third layer is dedicated to distributing the +5 V power. The thickness of the board was chosen to be 1,6 mm. Ordering a better layer assembly, JLC04161H-3313, from JLCPCB allowed for narrower differential pair traces and reduced spacing between conductors in pairs. The manufacturing cost of the 4 boards were 7 $, with shipping from China costing 20 $. The total cost for manufacturing and assembling all components of one piece of prototype v1.1 is 100 $.

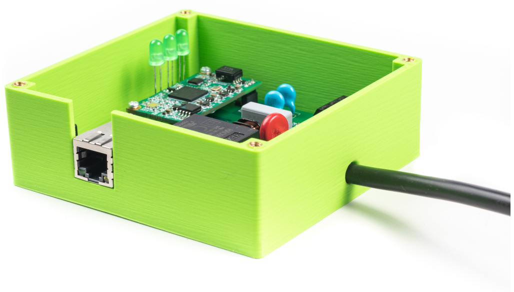


Fig. 3. BPL modem mounted in 3D case.

Assembled BPL modem is housed in a case made by a 3D printer, as depicted in Figure 3. The case dimensions are 100 × 90 mm with a height of 37 mm. The case includes a safety feature to secure the input power cable against accidental disconnection.

## IV. Comparison of BPL modem v1.1 throughput with commercially available solutions

Prototype v1.1 was tested for throughput in one topology alongside commercially available modems. Throughput measurements were conducted using the iPerf3 program with TCP transmission settings, a data window size of 2 MB, a measurement interval of 1 second, and a measurement duration of 5 minutes. The resulting throughput value is the arithmetic average of five measurements. Computer 1, acting as the server, was connected to the location marked in the BPL 01 topology images, while computer 2, operating as the client, was connected to the location marked BPL 02.

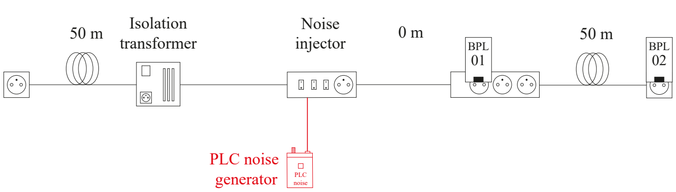



Fig. 4. Measurement topology of the influence of nearby noise.

The topology depicted in Figure 4 compares the constructed prototype, its development kit, and the commercially available devolo Magic 2 (G.hn standard modems) modem for resilience to four various levels of noise from a PLC noise generator. The topology consists of a 50-meter extension cord to which an isolation transformer is connected, to ensure noise isolation from the electrical network. A noise injector is connected to the isolation transformer, into which the PLC noise generator is plugged. The first of a pair of BPL modems is connected to the extension cord immediately after the noise injector. The second BPL modem is connected via a 50-meter extension cord. The results of the throughput measurements are presented in Table I.

TABLE I

Throughput of G.hn modems as a function of the level of nearby noise.

| Noise level | 0 | MIN | MID | MAX |
|---|---|---|---|---|
| **Modem** | **Throughput [Mbit/s]** | | | |
| devolo Magic 2 | 351 | 129 | 88 | 57 |
| Xingtera XTV6000 | 360 | 207 | 146 | 95 |
| Prototype v1.1 | 302 | 197 | 154 | 121 |

## V. Device noise level measurement using BPL modem v1.1

Knowledge about noise level in the spectrum is an important parameter for reliable communication of BPL modems. In addition to the impact of noise level on the resulting throughput of modems, it is important to monitor its level in terms of the balance of the power grid, affecting the operation of other devices connected to the grid, cable faults, and malfunctions of devices connected to the power grid.

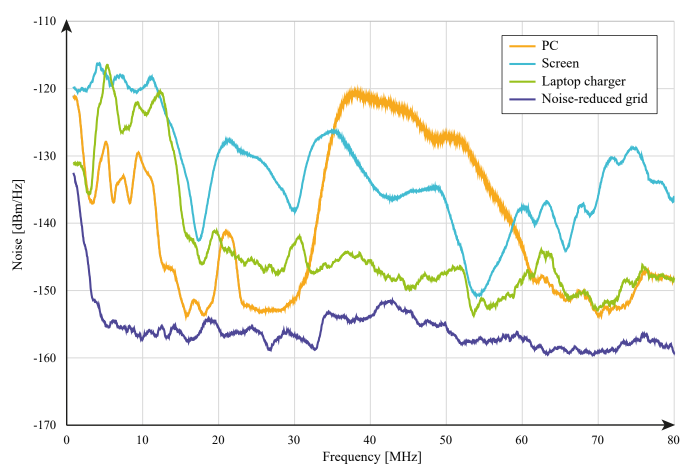



Fig. 5. Spectrum of noise levels commonly connected in the vicinity of BPL modems.

The constructed modem v1.1 is capable of periodic reading noise levels in the spectrum from 2 to 80 MHz and exporting them to .csv files for further processing. Noise levels from commonly connected devices to the electrical network, along with modems, are displayed in Figure 5. A special measurement topology was created for measuring the noise of individual devices, which included 50 m of attenuation for signals from the electrical network and an isolation transformer to which an additional 50 m of wiring was connected to minimize interference. In this assembled topology, individual devices were tested via a 2 m extension cord between BPL 01 and BPL 02. The graph reveals an increased noise production of tested devices in the range from 2 MHz to 12 MHz. The laptop charging adapter subsequently exhibits a decreasing trend, whereas other devices show an increasing trend. The PC screen reaches additional peaks at 35 MHz and the PC at 38 MHz.

## VI. Discussion of the measurements

The throughput measurements of the modems at four levels of injected noise showed that the constructed modem v1.1 has lower throughput in the scenario without noise compared to others. However, as the noise level increases, its throughput, as shown in Table I, increases compared to other modems. Under the highest noise level, the constructed prototype exhibits 21% higher throughput than the Xingtera XTV6000 development kit and 53% higher than the devolo Magic2.

The measurement of noise levels using BPL modems captured in the graph in Figure 5 shows following: if we connect BPL modems near the measured devices, such as a computer and a screen, there may be a significant decrease in modem throughput. This decrease occurs due to the high level of noise produced in a large part of the modem's communication spectrum. To verify the correctness of the noise level measurements, correlation was performed with a spectrum analyzer. Signal consistency was demonstrated in the frequency domain, with discrepancies appearing in the signal level offset.

## VII. Conclusion

The article describes the development of an affordable BPL modem capable of reading noise levels on the electrical lines within its communication bandwidth. Noise level data correlates with spectrum measurements using an expensive spectrum analyzer. The hardware part is based on the MOD957 module from Xingtera. Throughput measurements have shown that the constructed modem is more resistant to high noise levels compared to other modems. These factors predispose it to reliable communication in a noisy environment that it can spectrally measure. Automated reading of noise levels on the electrical lines will be used for further scientific research.

## References


[1] Huangqiang Li, Yunlian Sun, and Fangxian Jia, "Application of power line communication to the home network", in *2008 11th IEEE International Conference on Communication Technology*, 2008, pp. 521-524.

[2] L. T. Berger, A. Schwager, and J. J. Escudero-Garzás, "Power Line Communications for Smart Grid Applications", *Journal of Electrical and Computer Engineering*, vol. 2013, pp. 1-16, 2013.

[3] A. A. Zhilenkov, D. D. Gilyazov, I. I. Matveev, and Y. V. Krishtal, "Power line communication in IoT-systems", in *2017 IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering (EIConRus)*, 2017, pp. 242-245.

[4] "Powerline Communication for E-Mobility and Other Applications", *Codico*. .

[5] M. Schwartz, "Carrier-wave telephony over power lines: Early history [History of Communications]", *IEEE Communications Magazine*, vol. 47, no. 1, pp. 14-18, 2009.

[6] Z. Hasirci, I. H. Cavdar, N. Suljanovic, and A. Mujcic, "An application of the broadband PLC for smart homes in Turkey", in *2013 36th International Conference on Telecommunications and Signal Processing (TSP)*, 2013, pp. 222-226.

[7] M. Orgon, M. Stefanicka, I. Schmidt, I. Zolotova, and D. Cupkova, "Testing home PLC network in multi-storey house", in *2019 11th International Congress on Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT)*, 2019, pp. 1-6.

[8] R. Fujdiak, J. Slacik, M. Orgon, P. Mlynek, J. Misurec, J. Hallon, and J. Halgos, "Investigation of Power Line Communication and Wi-Fi Coexistence in Smart Home", in *2018 10th International Congress on Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT)*, 2018, pp. 1-4.

[9] Z. Hasirci, I. H. Cavdar, N. Suljanovic, and A. Mujcic, "An application of the broadband PLC for smart homes in Turkey", in *2013 36th International Conference on Telecommunications and Signal Processing (TSP)*, 2013, pp. 222-226.

[10] T. Sangsuwan, S. Thepphaeng, and C. Pirak, "Experimental performance analysis of powerline communication technologies in AMI systems", in *The 20th Asia-Pacific Conference on Communication (APCC2014)*, 2014, pp. 382-386.

[11] K. G. H. Mangunkusumo, A. S. Surya, D. R. Jintaka, and H. B. Tambunan, "NB-PLC Performance Evaluation for Advanced Metering Infrastructure", in *2021 3rd International Conference on High Voltage Engineering and Power Systems (ICHVEPS)*, 2021, pp. 414-418.

[12] A. Omri, J. Hernandez Fernandez, A. Sanz, and M. R. Fliss, "PLC Channel Selection Schemes for OFDM-based NB-PLC Systems", in *2020 IEEE International Symposium on Power Line Communications and its Applications (ISPLC)*, 2020, pp. 1-6.

[13] A. M. Liton, Z. Ren, R. Dong, and S. Xin, "An Efficient Networking Approach for Broadband PLC Networks", in *2021 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB)*, 2021, pp. 1-7.

[14] M. Lisowski, R. Masnicki, and J. Mindykowski, "PLC-Enabled Low Voltage Distribution Network Topology Monitoring", *IEEE Transactions on Smart Grid*, vol. 10, no. 6, pp. 6436-6448, 2019.

[15] M. Mohammadi, L. Lampe, M. Lok, S. Mirabbasi, M. Mirvakili, R. Rosales, and P. van Veen, "Measurement study and transmission for in-vehicle power line communication", in *2009 IEEE International Symposium on Power Line Communications and Its Applications*, 2009, pp. 73-78.

[16] Z. Xu, C. Yang, Z. Tan, and Z. Sheng, "Raptor Code-Enabled Reliable Data Transmission for In-Vehicle Power Line Communication Systems With Impulsive Noise", *IEEE Communications Letters*, vol. 21, no. 10, pp. 2154-2157, 2017.

[17] Z. Liang, C. Shengbin, C. Kang, L. Ling, and M. Huanhuan, "Oriented Electric Vehicles and DC Charging Piles Feasibility Analysis of Power Line Communication Based on OFDM", in *2019 IEEE Innovative Smart Grid Technologies - Asia (ISGT Asia)*, 2019, pp. 2595-2600.

[18] "EN 50065-1, Signaling on Low-Voltage Electrical Installations in the Frequency Range 3 kHz to 148.5 kHz-Part 1: General Requirements, Frequency Bands and Electromagnetic Disturbances." .

[19] G. Lopez, J. Matanza, D. De La Vega, M. Castro, A. Arrinda, J. I. Moreno, and A. Sendin, "The Role of Power Line Communications in the Smart Grid Revisited: Applications, Challenges, and Research Initiatives", *IEEE Access*, vol. 7, pp. 117346-117368, 2019.

[20] G. Hallak, C. Niess, and G. Bumiller, "Measurement Setup for Notch Evaluation of Narrowband PLC Devices", in *GLOBECOM 2017 - 2017 IEEE Global Communications Conference*, 2017, pp. 1-6.

[21] P. Mlýnek, M. Rusz, L. Benešl, J. Sláčik, and P. Musil, "Possibilities of Broadband Power Line Communications for Smart Home and Smart Building Applications", *Sensors*, vol. 21, no. 1, 2021.

[22] "Converging Technologies - Moving from HomePNA to G.hn: A Guide for Service Providers", pp. 1-13, 2013.

# Android Tracking Application Based on LTE Timing Advance

Jakub Michalek
*Department of Telecommunication*
*Brno University of Technology*
Brno, Czech Republic
186140@vut.cz

***Abstract*—This article introduces a novel Android application developed in the Kotlin programming language, which enables users to tracking the location of their devices through the implementation of a custom algorithm. The focal point of this innovation lies in leveraging the Timing Advance parameter derived from the Long-Term Evolution (LTE) network to ascertain the most optimal track for location tracking. The technical overview describes the technologies used, while the tracking application chapter offers a detailed look at the architecture and implementation details. Special emphasis is placed on the use of Timing Advance and its role in the position tracking algorithm. The testing results confirm the efficiency and accuracy of the proposed solution. The article further deals with possible problems and proposed future extensions of the application. This work provides a comprehensive view of an innovative approach to location tracking in mobile devices.**

***Index Terms*—Android, LTE, timing advance, tracking**

## I. INTRODUCTION

Localization has been identified as one of the most popular applications in modern society, especially when a huge number of devices are interconnected. Various types of services, including autonomous driving and indoor navigation, require high-resolution location information in both outdoor and indoor environments, which motivates continuous research in recent years. Although global navigation satellite systems (GNSS) such as the Global Positioning System (GPS) can provide continuous location information in outdoor environments, performance degradation typically occurs when satellite signals are blocked, such as in some urban canyons and indoor environments [1], which launches an effort to support cell-based localization techniques.

In this paper, to deal with the above problems, we propose a new localization technique based on the Timing Advance (TA) parameter detected in Android OS mobile device in commercial LTE systems. Specifically, we are developing a mobile application in the Kotlin language that is able to track the track along which the user is moving in real time. Furthermore, after completing the tracing, optimize the resulting track for greater accuracy in determining the location. Results show that the proposed approach achieves sufficient throughput for both urban and non-urban routing. The primary contributions of this article are succinctly outlined below.

The remainder of this paper is structured as follows: Section II delineates the characteristics of the TA parameter in LTE networks, while Section III elucidates the proposed algorithm for location determination. The functionality of the device tracing application is explicated in Section IV, followed by the presentation of measurement results in Section V. Finally, Section VI encapsulates our conclusions.

## II. LTE NETWORKS

LTE is the 4th generation of wireless communication systems. It was designed to provide much higher data rates and lower latency compared to legacy Universal Mobile Telecommunication System (UMTS) and GSM systems. This was obtained by using flat architecture and enabling higher bandwidth, higher order of modulation and better support of multiple-input multiple-output (MIMO). [2]

### A. Timing advance in LTE

In LTE, Time-Division Multiple Access (TDMA) is utilized as part of the strategy for allocating radio frequency bandwidth among multiple users. Each active user in LTE is allocated a specific time slot during which their transmission should reach the base station (BS), ensuring minimal interference with transmissions from other users. Due to the finite speed of electromagnetic propagation, the message must be sent from the user equipment (UE) earlier than the designated time slot. Additionally, UE clocks are not perfectly accurate nor synchronized, necessitating timing based on the arrival times of transmissions from the BS.

Consequently, the UE sends its message at a predetermined negative offset known as the TA, relative to the start of a message slot delineated by signals received from the BS. Furthermore, signals from the BS require time to traverse the distance to the UE and vice versa. Hence, the TA represents the round-trip time from the BS to the UE and back to the BS, effectively twice the one-way propagation time. [3]

Initially, the BS determines a suitable value for TA through brief communication with the UE over a separate random access channel (RACH). The UE must establish a connection with the BS before knowing the TA. Subsequently, the BS transmits this initial value of TA to the UE for use in data transmission.

As TA is contingent on the distance between the UE and the BS, it undergoes changes if the UE relocates. During data communication, the BS continually monitors the arrival time of messages from the UE. It then sends commands back to the UE to effect minor adjustments to the TA, ensuring continued avoidance of overlap with messages from other UEs.

Due to the potential mobility of the UE, the current TA value remains valid for only a brief duration, as determined by a timer set by the BS. If there is no further data communication within this timeframe, a new TA must be negotiated using the RACH. Consequently, the UE's radio maintains a valid TA only under two conditions: during an active data connection, and for a short period thereafter, as constrained by the TA Timer.

The TA parameter can be converted into the distance between the BS and the UE. Each TA step corresponds to a distance of 78 meters, which significantly surpasses the accuracy of 3G propagation delay (234 meters). This TA value serves as the basis for geolocation estimation.

In LTE, there exists an upper limit on the possible values of TA. The initial TA sent from the BS to the UE is an 11-bit quantity, capped at a maximum value of 1282. Incremental updates sent from the BS to the UE are signed quantities packed into six bits. Consequently, this maximum value allows for a maximum one-way distance of just over 100 kilometers. [4]

### B. Timing advance on Android

On Android, it is possible to obtain the TAparameter from the LTE connection information provided by the Telephony Manager Class. It is important to note that on many devices from different manufacturers the TA is reported incorrectly and on them one TA step is 144 meters.

In addition, it is also necessary to grant permission access to the phone to applications that want to obtain this parameter. This permission also allows the application to read additional information about the BS to which the device is currently connected. [5]

## III. Algorithm Explanation for LTE Tracking

A proprietary algorithm has been devised to enable real-time tracking of the user and to compute tracks based on the acquired data. This algorithm hinges on calculating track points utilizing the known positions of the base stations to which the device is connected, alongside the TA parameters negotiated by the device with the base station at any given moment. The calculation of new points along the current track occurs whenever there is a transition in the connected base station, as well as when the device approaches the newly connected station. Figure 1 illustrates both scenarios upon which the algorithm operates.

When the device transitions its connection from one base station, say B1, to another, B2, due to a change in position, a modified linear interpolation method is employed to ascertain the device's position. To accomplish this, the first step involves calculating the direction vector, denoted as **u**, originating from B2 and terminating at B1:

$$\mathbf{u} = (lat_{B1} - lat_{B2}; lon_{B1} - lon_{B2}) \tag{1}$$

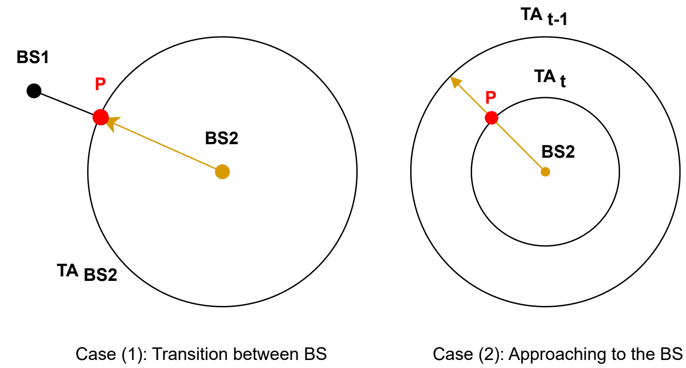



Fig. 1. Cases for new position calculation.

On this vector, at a distance TA from station B2, a calculated point P is established, which denotes the position of the device following the base station change. The coordinates of point P are determined as follows:

$$\phi = atan(\mathbf{u}) \tag{2}$$

$$\begin{aligned} lat_P = asin(sin(lat_{B2})cos(\frac{d}{R}) \\ +cos(lat_{B2})sin(\frac{d}{R})cos(\phi)) \end{aligned} \tag{3}$$

$$\begin{aligned} lon_P = lon_{B2} + atan(sin(\phi)sin(\frac{d}{R})cos(lat_{B2}) \\ , cos(\frac{d}{R}) - sin(lat_{B2})sin(lat_P)) \end{aligned} \tag{4}$$

,where:

- $\phi$ is azimutal angle of direction vector **u**,
- $lat_{B2}$ and $lon_{B2}$ are the coordinates of station B2,
- D is the distance determined by TA in meters,
- R is the radius of the earth.

The formula employed is a modified version of the Haversine formula, renowned for its accuracy in calculating distances between two points on a sphere, such as the Earth. This formula operates with coordinates in radians; hence, it necessitates conversion to radians prior to computation and subsequent conversion back to degrees to obtain coordinates. Similarly, when the device approaches the newly connected base station, the same formula is utilized to compute positions. The only adjustment made is to the distance parameter, D, as the value of the TA parameters decreases in this scenario.

For the tracing process itself, this algorithm is systematically applied to the entire dataset, comprising information about the parameters of the base stations retrieved from the device. The intricacies of data acquisition and the application of the algorithm to derive the user's track will be elucidated in the subsequent chapter.

## IV. Tracking application

An Android application has been developed to facilitate the tracking of user location, leveraging the Kotlin programming language and structured upon the Model-View-ViewModel (MVVM) architectural pattern. This application employs a SQLite database embedded within the device for data storage, with management facilitated through utilization of the Room library.

### A. Data collection and processing

In order to obtain device location data, initiating the application and activating location tracking functionality is imperative. Concurrently, granting access permissions to both phone and location services is essential. Authorization to access the phone is requisite for extracting pertinent information from the network, notably pertaining to the base station to which the device is currently connected. Equally critical is the authorization for location access, which facilitates retrieval of device location data from GPS. This data serves as a reference point for comparison with the calculated device location within the application.

Following the initiation of tracing, the application undertakes periodic data retrieval from both the base station and the GPS sensor. This data collection occurs seamlessly in the background, even when the application is not actively in use, facilitated through the implementation of a Foreground service. A crucial criterion for storing this acquired data in the database is the device's connection to the LTE network, alongside the identification of the current base station connection. Subsequently, the geographic coordinates corresponding to the identified base station are assigned from the local base station database. It is pertinent to note that the local database encompasses an extensive repository of over 50,000 LTE base stations attributed to the Czech mobile operator O2, covering the geographic scope within which the application was subjected to testing. Upon fulfillment of both prerequisites, the record of the device's current location is persisted in the database, adhering to the format illustrated in Figure 2.

```
@Entity(tableName = "locations")
data class Location (
    @PrimaryKey(autoGenerate = true)
    val id: Long =0,
    val time: String, // Date and time of record
    var trackID: Int, // Track identification
    val btsCi:Int, // BTS identification
    val btsTac:Int, // BTS local area code
    val btsType:Int, // Type of mobile network
    var btsX:Double, // BTS Coordination's
    var btsY:Double,
    val btsTA:Int, // Timing advance value
    val btsRSSI:Int, // Receive signal strength indicator
    var gpsX:Double, // GPS Coordination's
    var gpsY:Double
)
```

Fig. 2. Structure of location record in Room database.

Before the actual tracing, the data from the location database is pre-processed to a specific path. Records lacking the TA parameter are omitted from consideration. Moreover, redundancy within the dataset is addressed through a methodical approach whereby consecutive records sharing identical TA parameters and originating from the same base station are consolidated into a singular record. This process effectively curtails redundant data, enhancing the efficiency and relevance of the dataset for subsequent analysis.

### B. Live tracking

The real-time tracing process involves the continuous acquisition of position data, followed by the utilization of an algorithm to ascertain the current device position from the TA parameter. As elaborated in Chapter III, the calculation of the new position occurs selectively, triggered either by a change in the connected base station or by the device nearing the presently connected base station. It is crucial to acknowledge that this condition introduces a delay in obtaining the current position, necessitating due consideration. Figure 3 delineates the stepwise process involved in acquiring the current position utilizing the proposed algorithm.

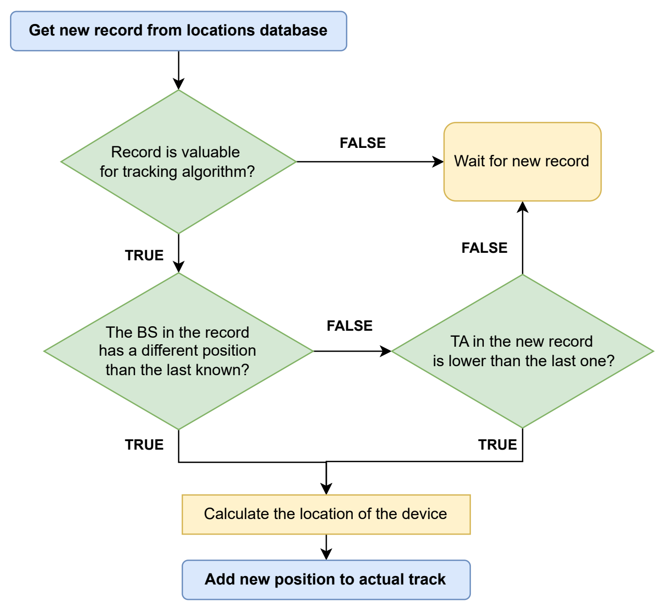



Fig. 3. Live tracking position calculation flow chart.

The application utilizes the osmdroid library to visualize device location on map widgets, enabling seamless integration with maps sourced from OpenStreetMap and facilitating interaction with them. During tracing sessions, the default display on the map showcases the current calculated location of the device. Additionally, users have the option to enable the visualization of the calculated track, GPS position, and the location of the presently connected base station. Notably, an annulus delineates the area within which the device can potentially be situated, based on the knowledge derived from the TA parameter. Figure 4 illustrates a screenshot capturing the live trace of the device, showcasing these features in action.

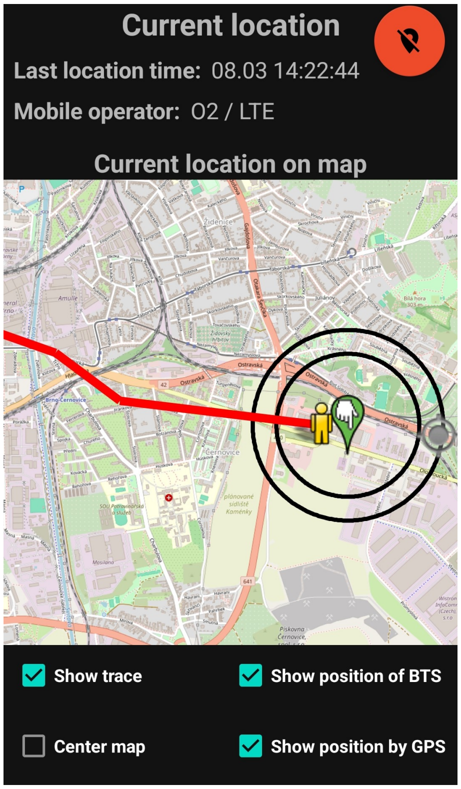



Fig. 4. Live tracking fragment - person icon is calculated location.

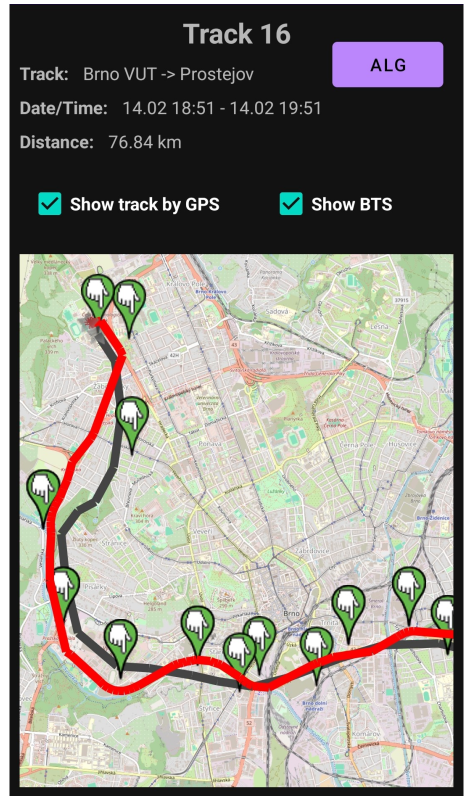



Fig. 5. Track history fragment - black line is calculated and red is determined from GPS, green marks are BS positions.

### C. Track history

Upon completion of live tracing, users have the option to preserve the track within the track history. Within the track history list, each recorded track is characterized by its identifier, name, date, start point, track duration, and distance. Furthermore, users can customize track names, view detailed track information, and delete tracks as needed directly from this list interface.

Upon selecting the track detail button to display a specific track, a dedicated fragment is launched, showcasing the calculated track on the map. Users are empowered to toggle the visibility of additional track details, including the track obtained via GPS and the location of base stations connected during the track, through checkbox controls. Figure 5 provides a visual representation of the fragment displaying the calculated track alongside pertinent track data.

The displayed calculated track undergoes additional optimization utilizing a Simple Moving Average (SMA) technique, employing a window width of 3. Traditionally employed in time series and financial analysis, the SMA serves to mitigate short-term fluctuations within the calculated track, thereby elucidating the overarching trend of device movement. This optimization step enhances the interpretability and coherence of the displayed track trajectory, facilitating a more insightful analysis of the device's movement patterns.

## V. Results

This chapter aims to evaluate the accuracy of the proposed algorithm for track determination by comparing it against the track calculated from GPS-derived positions, which serves as the reference data. The study encompasses several records obtained through the application, all conducted on the same mobile device during driving sessions. Through this comparative analysis, insights into the efficacy and reliability of the proposed algorithm in accurately determining tracks will be gleaned.

In the first case, a composite track from Kladno to Brno is examined, covering urban, suburban, and highway driving segments, totaling approximately 270 km. The measurement reveals an average Root Mean Square Error (RMSE) of approximately 1260 m, with a standard deviation of 800 m. The maximum positional error remains within 4000 m, as depicted in the upper portion of Figure 6.

The Cumulative Distribution Function (CDF) presented in the lower panel of Figure 6 illustrates the probability distribution of exceeding certain error thresholds. Approximately 50 percent of errors surpass the 950 m threshold (median), with around 10 percent of errors extending beyond the 2000 m. Such error occurrences are primarily attributed to the relatively sparse deployment of base stations in less populated areas, compounded by the inherent delay in retrieving the TA parameter from these stations.

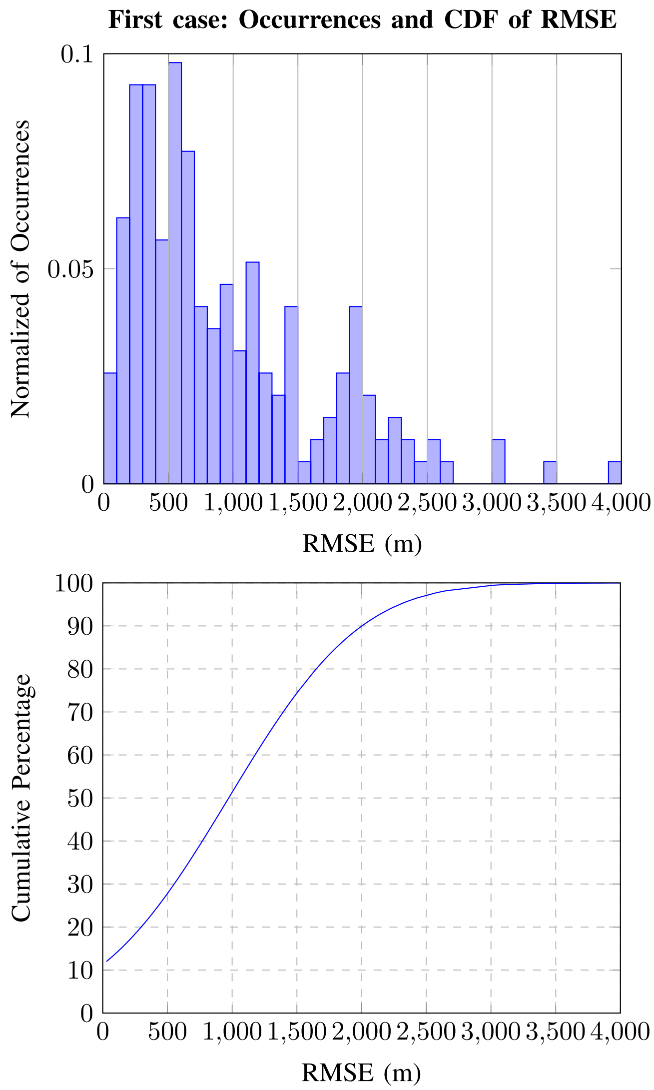



Fig. 6. RMSE statistics for first case.

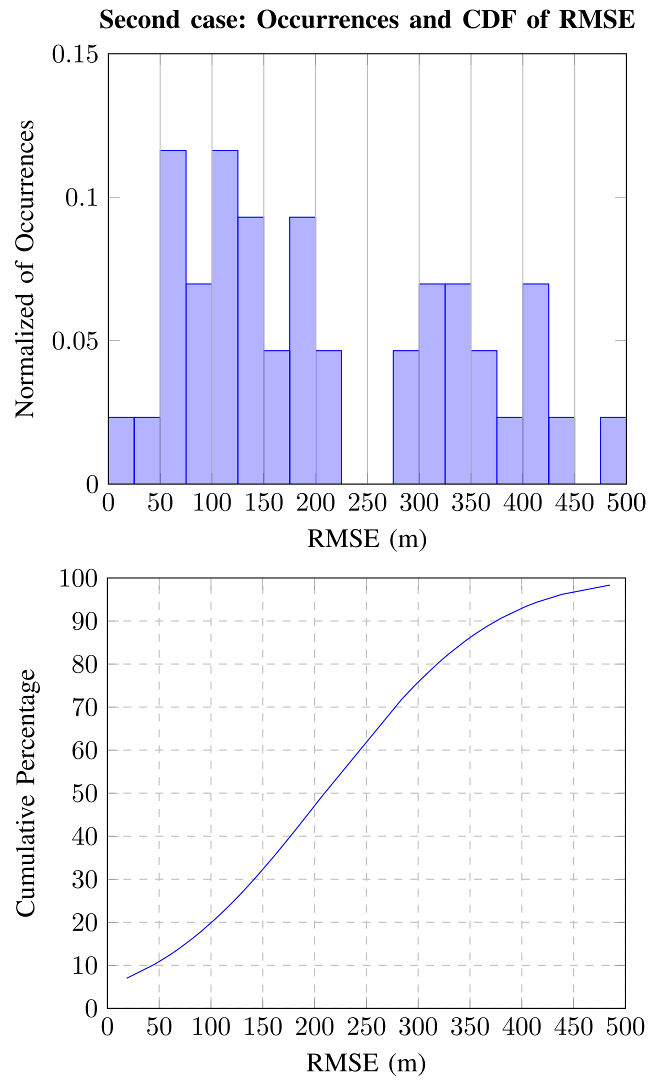



Fig. 7. RMSE statistics for second case.

In the second case, the focus is on a track traversing the densely populated city of Prague, covering approximately 17 km. This urban environment poses distinct challenges and opportunities for track determination. The evaluation reveals an average RMSE of approximately 245 m, with a standard deviation of 129 m. Importantly, the maximum positional error remains below 500 m, as illustrated in the plot of Figure 7.

The CDF, depicted in the lower panel of Figure 7, provides insights into the probability distribution of exceeding certain error thresholds. Approximately 50 percent errors surpass the 260 m threshold, with around 10 percent errors extending beyond the 360 m. Notably, in contrast to the previous case, the algorithm exhibits significantly improved accuracy within the urban confines of Prague. This enhanced performance can be attributed to the denser deployment of base stations and their consequent overlap within the urban landscape.

In the final case, a comparison is made between track calculations derived from 10 measurements, focusing on the potential for reducing positional determination errors through the application of statistical mechanisms and the Kalman filter. This assessment pertains to a combined track from Prostějov to Brno, covering a smaller city, a larger urban center, and a highway. The calculated RMSE values and their respective means for all measurements, alongside the application of statistical methods, are summarized in Table I.

From the table, it's evident that both the algorithm alone and its application with SMA yield comparable results. However, in practical terms, utilizing SMA offers distinct advantages, particularly in tracks featuring significant local fluctuations. This method effectively smooths out such fluctuations, resulting in enhanced accuracy and visually clearer tracks. Conversely, the use of the Kalman filter yielded notably inferior results compared to the previous methodologies. This discrepancy in error rates likely stems from suboptimal parameter selection for the filter, suggesting potential for optimization in future implementations.

TABLE I
RMSE VALUES FOR INDIVIDUAL RECORDS OF THE SAME TRACK.

| | RMSE [m] | | |
|---|---|---|---|
| **Track** | **Only algorithm** | **SMA calculation** | **Kalman filter** |
| 1 | 743 | 825 | 1312 |
| 2 | 825 | 883 | 1386 |
| 3 | 1126 | 995 | 1583 |
| 4 | 1090 | 1180 | 1535 |
| 5 | 1018 | 979 | 1503 |
| 6 | 1552 | 1364 | 1866 |
| 7 | 1456 | 1274 | 1962 |
| 8 | 600 | 964 | 1374 |
| 9 | 1233 | 1338 | 2067 |
| 10 | 925 | 938 | 1417 |
| **Average** | **1057** | **1074** | **1600** |

## VI. CONCLUSION

This article introduced a mobile application designed for tracking a user's device and the algorithm employed for determining the device's location utilizing information sourced from the LTE network. The algorithm primarily relies on the Timing Advance parameter acquired from the connected base stations, particularly emphasizing its variation during base station transitions. It integrates linear interpolation, the Haversine formula for calculations, and incorporates SMA for result refinement. The method's advantage lies in achieving acceptable device routing accuracy solely through the LTE mobile network connection.

The mobile application offers comprehensive functionalities for displaying both the current track and the history of previously recorded tracks. Notably, its clear presentation of the user's location and accompanying data aids users in comprehending the LTE mobile network structure and facilitates comparison of the proposed algorithm's accuracy with GPS-based location determination. Moreover, a key advantage of the application is its minimal data consumption, requiring only a mobile network connection for location determination without utilizing mobile data, thus further enhancing its utility.

## REFERENCES

[1] R. Bajaj, S. Ranaweera, and D. Agrawal, "Gps: Location-tracking technology." *IEEE Computer*, vol. 35, pp. 92–94, 04 2002.

[2] N. Ismail, Y. Fahmy, and M. Khairy, "Connected users localization in lte networks," in *2019 15th International Computer Engineering Conference (ICENCO)*, 2019. doi: 10.1109/ICENCO48310.2019.9027466 pp. 7–12.

[3] L. Jarvis, J. McEachen, and H. Loomis, "Geolocation of lte subscriber stations based on the timing advance ranging parameter," in *2011 - MILCOM 2011 Military Communications Conference*, 2011. doi: 10.1109/MILCOM.2011.6127575 pp. 180–187.

[4] B. Horn, "Timing advance in lte," 2022. [Online]. Available: http://people.csail.mit.edu/bkph/cellular_repeater_TA.shtml

[5] Developers, "Permissions on android," Google LLC, 2022. [Online]. Available: https://developer.android.com/guide/topics/permissions/overview

# Using 5G-IoT Networks for Portable Electric Vehicle Charging Stations


1st Michal Mikulášek
*Department of Telecommunications*
*Brno University of Technology*
Brno, Czech Republic
xmikul58@vutbr.cz

2nd Lukáš Jablončík
*Department of Telecommunications*
*Brno University of Technology*
Brno, Czech Republic
lukas.jabloncik@vut.cz

3rd Pavel Mašek
*Department of Telecommunications*
*Brno University of Technology*
Brno, Czech Republic
masekpavel@vut.cz



***Abstract*—The transition to electric vehicles (EV) as part of corporate fleets poses challenges for monitoring and reimbursing home charging activities. This paper explores the application of Internet of Things (IoT) networks to facilitate the use of portable EV charging stations by employees for home charging of corporate vehicles. The core objective is to enable precise reporting of the energy consumed during home charging sessions directly to employers, ensuring accurate compensation and energy usage tracking. This article looks at the theoretical and practical assessment of whether 5G-IoT networks are suitable for this particular case. An experimental setup was established, deploying a portable EV charging station integrated with 5G-IoT connectivity to test its effectiveness in real-world residential settings under varying radio conditions. The experiment focuses on assessing the reliability of data transmission, usability in home conditions, and user experience in interfacing with the charging infrastructure. Results demonstrate the feasibility of using 5G-IoT networks for this application, highlighting the potential for seamless, efficient, and secure communication between portable charging stations and corporate monitoring systems. The paper also identifies potential challenges, such as network coverage in residential areas and the challenges in device design.**



***Index Terms*—LTE Cat-M1, NB-IoT, electric vehicle, portable charging station, charging**


## I. Introduction

The European Union has set a target to reduce emissions by at least 55% by 2030. The European Union's commitment to reducing emissions through the "Fit for 55" package highlights the continent's proactive approach to combat climate change. This ambitious target is part of a comprehensive strategy that includes the electrification of the transportation sector, which is one of the largest contributors to greenhouse gas emissions [1]. The uptake of EVs in Europe has seen a remarkable increase, driven by both policy initiatives and advances in EV technology. Governments across Europe are introducing various incentives, such as subsidies for the purchase of EVs, reduced taxes and exemptions from certain charges, to encourage consumers to switch to battery electric vehicle (BEV) from Internal Combustion Engine (ICE). The European Green Deal aims to support the installation of 1 million charging points by 2025. This expansion is crucial to support the expected surge in EV uptake, as EVs are forecast to account for a significant proportion of all vehicles on European roads by 2030 [2].

Adequate charging infrastructure will be a key factor in the transition to electromobility. Charging stations can be divided into levels 1, 2 and 3. This provides a framework for understanding the range of charging speeds available to consumers. Level 1 and 2 Alternating Current (AC) charging stations are typically used for charging in homes or workplaces and provide slower charging for a few hours. The typical output of these stations is up to 22 kW. Level 3 Direct current (DC) fast charging stations offer the convenience of fast charging, making them ideal for public places and highways. The development of ultra-fast charging stations, capable of delivering 350 kW or more, will revolutionize long-distance electric vehicle travel by reducing charging times almost to the level of traditional internal combustion engine vehicle refueling times. The development and construction of slow charging stations will play an important role in normal daily charging [3].

In addition to fixed charging stations called wallboxes, there are also portable charging stations. These should solve the problems of charging electric vehicles, especially in urban areas with limited access to private parking spaces. Portable charging stations can be placed in a variety of environments, offering flexibility for both EV owners and businesses. For businesses with a fleet of EVs, portable charging stations provide a practical solution for employees who need to charge their company vehicles at home, and solve the problems associated with installing permanent charging points in rented or shared accommodation. The requirements for a portable stations are: secure and reliable data transmission from the charging station, the ability to transmit from underground garages, the ability to use Open Charge Point Protocol (OCPP) 1.6 and higher, low hardware cost, the lowest possible data volume, and guaranteed availability for energy management. The objective of this paper is to determine whether 5G-IoT networks, such as LTE Cat-M1 or NB-IoT, are suitable for facilitating the communication and operational requirements of portable charging stations (EVs). This research is motivated by the market need for portable charging solutions in urban environments and beyond presents a set of unique challenges, including the need for robust and reliable data communications, the ability to operate in areas with poor signal penetration such as underground parking garages, and the need for low operating costs due to minimal data consumption [4]–[6].


Fig. 1. Cubish portabox [7].

The key outputs of our work can be summarized as follows:

- the suitability of 5G IoT for communication of the portable chargers,
- sufficient coverage for portable chargers network,
- future development and mass deployment.

The rest of the paper is organized as follows. The description of the LTE Cat-M1 and NB-IoT technology and IoT network requirements is given in Section II. In Section III, the individual parts of a portable charging station are listed below, including the selection of suitable components. Next, the measurement results are presented in Section IV. Finally, we conclude the paper in the Section V and future works are mentioned.

## II. 5G-IoT Networks

Narrowband IoT (NB-IoT) and LTE Cat M1, two low-power wide area (LPWA) technologies, have become the essential part of the fifth generation (5G) of cellular systems. Initially defined by the 3rd generation partnership project (3GPP) in release 13, both technologies were introduced as alternatives to the already available technologies for long-range communication, i.e., LoRa(WAN) and Sigfox. These technologies were builds upon the 3GPP long term evolution (LTE) standard with additional features designed for the needs of the industry sector. Compared with the technologies working in the license-exempt frequency spectrum (LoRaWAN, Sigfox, Wireless M-BUS, etc.), the 3GPP cellular technologies for the internet of things (CIoT) and industry 4.0 utilize the dedicated (licensed) frequency spectrum exactly as the fourth generation of the cellular systems do [8], [9].

The importance of massive machine-type communication (mMTC) has been identified worldwide, and since 3GPP release 15, the CIoT technologies have been integral part of the 5G networks. The list of features increases in every new 3GPP release, as the new requirements for the simple, performance-limited, often battery-powered devices are driven by the foreseen communication scenarios. Remote sensing and metering, assets tracking, online monitoring, and data gathering are the use-cases the LPWA technologies were designed for [10], [11].

Compared side by side based on their specifications, LTE Cat M1 should outperform the NB-IoT in case of the data throughput in both uplink and downlink data transmissions. The improved speeds were achieved by increasing the bandwidth to 1.4 MHz (LTE Cat M) vs. 200 KHz (NB-IoT), resulting in the uplink data rates up to 1 Mbps (LTE Cat M1) vs. up to 62.5 Kbps in case of the NB1. The list of the communication parameters is shown in Table I [12]–[16].

TABLE I
Comparison of the CIoT technologies defined in 3GPP Releases 13 and 14 [15], [16].

| Technology | NB-IoT (NB1) | NB-IoT (NB2) | LTE Cat M1 | LTE Cat M2 |
|---|---|---|---|---|
| 3GPP | Release 13 | Release 14 | Release 13 | Release 14 |
| Spectrum | Licensed | | | |
| Frequency | 700-2100 MHz | | | |
| Bandwidth | 200 kHz | | 1.4 MHz | 5 MHz |
| Link budget | 164 dB | | 155.7 dB | |
| Max. EIRP | 23 dBm | | | |
| Max. payload | 1600 B | | 8188 B | |
| UL data rate | 0.3-62.5 Kbps | 0.3-159 Kbps | HD: 375 Kbps;<br>FD: 1 Mbps | HD: 2.625 Kbps;<br>FD: 7 Mbps |
| DL data rate | 0.5-27.2 Kbps | 0.5-127 Kbps | HD: 300 Kbps;<br>FD: 0.8 Mbps | HD: 2.35 Mbps;<br>FD: 4 Mbps |
| Consumption | Tx: 240 mA<br>Rx: 12 mA<br>PSM: <1uA | Tx: 240 mA<br>Rx: 46 mA<br>PSM: <3uA | Tx: 360 mA<br>Rx: 46 mA<br>PSM: <3uA | Tx: 360 mA<br>Rx: 70 mA<br>PSM: <8uA |
| Security | LTE security | | | |

### A. NB-IoT and LTE Cat M1

Both of the selected LPWA technologies represent 3GPP ratified representatives operating in a licensed band. However, they target different mMTC scenarios as NB-IoT represents a strictly power-saving oriented technology with a highly prolonged communication range with relaxed communication delay requirements. On the other hand, LTE Cat M1 is targeted toward more demanding scenarios which require the transmission of larger data volumes, higher data throughputs, and lower latency. Particularly, LTE Cat M1 still preserves significantly reduced complexity compared to full-fledged LTE systems.

The first difference between NB-IoT (LTE Cat NB1) and LTE Cat M1 is evident from the comparison of spectrum utilization. NB-IoT relies on a significantly reduced 180 kHz (200 kHz including guard bands) channel with 15 kHz spacing (additional 3.75 kHz variant for single tone transmission) when the multi-tone transmission is used, enabling the communication speed of up to 62 kbps. This reduced bandwidth allows NB-IoT to be deployed in three different ways. Firstly, it can be deployed in-band as a single physical resource block (PRB) of a traditional LTE system. NB-IoT can also be deployed in a standalone mode occupying a single channel of a global system for mobile communications (GSM) or in a guard band of the conventional LTE system [13].

On the contrary, LTE Cat M1 does not use 1 but 6 PRB, resulting in a bandwidth of 1.08 MHz (1.4 MHz when guard bands are included) with 15 kHz subcarrier spacing. As a result, it allows LTE Cat M1 to achieve a 1 Mbps data rate in full-duplex mode and 300 kbps in half-duplex. The ability to operate in full-duplex mode is another advantage of LTE Cat M1 over NB-IoT, which operates only in a half-duplex mode [13], [15], [16].

As representatives of LPWA standards, both NB-IoT and LTE Cat M1 provide extended coverage compared to traditional LTE. The increased link budget is primarily achieved through repetitions. In terms of NB-IoT, the maximum coupling loss (MCL) is as high as 164 dB, which is a 20 dB increase. LTE Cat M1, on the other hand, provides approximately a 10 dB gain over traditional LTE, resulting in a total MCL of 155.7 dB. These differences are mainly coming from the multi-state modulation scheme of LTE CaT M1 combined with the lower number of repetitions. From the perspective of maximum transmission powers, both systems provide two power classes with maximum values of 20 dBm (Power Class 5) and 23 dBm (Power Class 3) [13].

### B. LTE Cat NB2 and LTE Cat M2

The previous sections described the initial releases of NB-IoT and LTE Cat M1.

NB-IoT (LTE Cat NB2), release 14 most importantly added handovers to the RRC-connected state. Also, the transmission speeds were increased to 159 kbps and 127 kbps in the uplink and downlink directions, respectively. Further, multicast transmission, a new 14 dBm power class, and observed time difference of arrival (OTDOA) positioning were implemented [15], [16].

LTE Cat M1 was also significantly improved in release 14. Bandwidth was extended to 5 MHz, which also improved throughput of 7 and 4 Mbps for uplink and downlink, respectively. Thanks to the increased transport block size (TBS), the data rates of LTE Cat M1 devices are increased. As for NB-IoT, OTDOA localization and multicast transmissions are supported. Release 15 then introduces a 14 dBm power class [15], [16].

## III. Portable Electric Vehicle Charging Station

Several portable charging stations were investigated to test 5G-IoT network technologies. Namely, Sparkline and Cubish Portabox from Inchanet shown in figure 1, Porty portable charging station from Ejoin, Feyree EV Portable Charger Type 2 and TAYSLA Electric Car Charger Type 2. Three of these portable charging stations are from European manufacturers, two portable charging stations are from Chinese manufacturers. Detailed specifications are listed in the following table II.

From the above mentioned portable charging stations, the Portabox charging station was selected. The main reason was that the Portabox has enough space inside for the control unit. Moreover, this portable charging station has a simple design and it was possible to connect the experimental controller supporting the 5G-IoT network to it.

TABLE II
Comparison of portable charging stations

| Charger name | Charging power | Number of phases | Connectivity |
|---|---|---|---|
| Portabox | 22 kW | 3 | - |
| Sparkline | 22 kW | 3 | - |
| Feyree charger | 11 kW | 3 | - |
| Taysla charger | 3,7 kW | 1 | - |
| Ejoin Porty | 22 kW | 3 | WiFi, 4G |

### A. Portabox

The Portabox is equipped with overvoltage, undervoltage and leakage current protection of our original development. The switching elements are rated for more than 150% of the rated continuous load. The control electronics performs self-diagnosis and displays error codes. If you use the charging station in your company fleet, it can be connected via OCPP to the remote management system and your charging sessions will be synced automatically.

Portabox continuously measures available electrical power capacity to prevent circuit breaker overload and to monitor building power reserve. This wireless solution ensures maximum available charging power at all times [7].

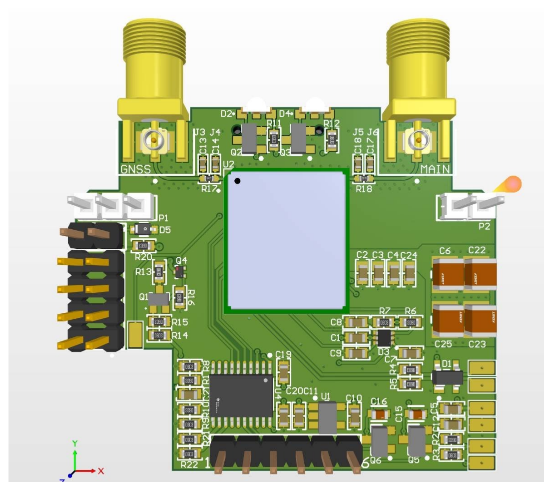


Fig. 2. Communication board prototype.

### B. Communication Module Selection

During preparation, several communication modules compatible with NB-IoT and LTE Cat-M were analyzed. Main focus was on receiver sensitivity, maximal transmission power, supported communication protocols and easy integration in custom control unit.

Resulting comparison can be seen in Table III. Based on this comparison, we have decided to use Quectel's BG77 modules as it provided all necessary communication protocols (PPPoS mainly) and it was also extensively used in other projects in the past. All compared modules supported NB-IoT2 and LTE Cat-M in 3GPP Release 14 versions. There are starting to appear modules, which support 3GPP Release 15, but they are not as common as 3GPP Release 14.

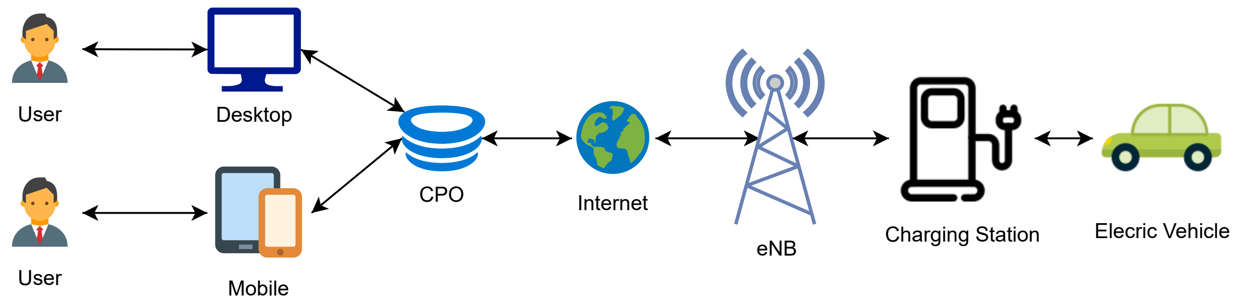



Fig. 3. Comunication schematic.

TABLE III
COMPARISON OF COMMUNICATION MODULES

| Module | Technologies | Receiver sensitivity | Transmit power |
|---|---|---|---|
| TX62-W (Telis) | NB2+LTE-M | unknown | 23 dBm |
| Sara-R5 (u-Box) | NB2+LTE-M | -116 dBm | 23 dBm |
| BG77 (Quectel) | NB2+LTE-M | unknown | 20 dBm |
| SIM7080G (SIMCom) | NB2+LTE-M | unknown | 20 dBm |
| nRF9160 (Nordic Semi) | NB2+LTE-M | -108 dBm | 23 dBm |

### C. Open Charge Point Protocol

The Open Charge Point Protocol (OCPP) 1.6 is one of the most important protocols in the development of electric vehicle charging infrastructure. It provides a standardised communication framework between charging points and control systems, called a Charge Point Operator (CPO). OCPP 1.6 was developed by the Open Charge Alliance and supports a wide range of features important for efficient and flexible operation of charging stations. These features include remote starting and stopping of charging sessions, status notification, firmware management and error reporting. Version 1.6 also introduces smart charging features that allow charging points to adjust charging speeds based on network demand. This promotes network stability and optimises power consumption [6], [16].

## IV. MEASUREMENT CAMPAIGN

The aim of the measurement was to verify whether LTE-M and NB-IoT technologies are suitable for portable charging stations. For the measurements, a scenario with a typical use of a portable charging station was created. It is assumed that a typical user will use two to three typical locations for charging an electric vehicle. A location at work, a location at home and a location at the cottage house. Three specific locations have been identified. The first location was the BUT university campus. This simulated the environment of a typical user at work. The second location was an apartment building in the center of Brno. This location simulates a place where the user will charge on weekdays. The third location was identified in a village in the mountains where the user may have a house cottage.

The BG77LA-64-SGNS module from Quectel was chosen for the measurement. The module was controlled by an ESP-32 microprocessor. It communicated with the Portabox charging station using the RS-485 bus. The Smart Fuel Pass portal from Ultima Payments was chosen as the CPO and the Thingsboard environment was chosen for sending telemetry data. The unit communicated with the CPO using OCPP 1.6J and telemetry was sent using MQTT [17].

Employers will manage the charging stations primarily using CPO and data will be sent using OCPP. Therefore, the portable charging station shall be able to send initiation of communication with the kWh meter status. During charging, the charging station shall send charging status every five minutes. When charging is complete, the final status in the meter is sent. The charging station can be switched off/on during charging, or it can be set according to the OCPP parameters.

TABLE IV
MEASUREMENT RESULTS

| Place | Network | Connection Time (s) | RSSI (dBm) | RSRP (dBm) | SINR (dB) |
|---|---|---|---|---|---|
| University Campus | LTE-M | 16,21 | -87 | -101 | 11 |
| Center of Brno - ground floor | LTE-M | 16,40 | -70 | -80 | 30 |
| Center of Brno - 1st und. floor | LTE-M | 17,67 | -94 | -106 | 11 |
| Center of Brno - 2nd und. floor | LTE-M | 130,38 | -95 | -108 | 9 |
| Bohutín village | NB-IoT | 83,86 | -81 | -91 | 20 |

During the measurement, the value of the time required from switching on the charging station to sending the first data was recorded. Then the RSSI, RSRP and SINR values were recorded when the first data was sent. The results are shown in Table 3. These parameters were then stored every minute. The objective was to determine the change in the parameters over time.

A new printed circuit board (PCB) was designed for the measurement. The basis of the PCB is the BG77 module. The BG77 communicates with the ESP-32 control processor using a Universal Synchronous/Asynchronous Receiver and Transmitter (USART). The power supply voltage of the module is 3.3 V, but the logic pins need 1.8 V. Therefore, a logic shifter was used. There are two LEDs to indicate the connection to the network. An antenna can be connected to the PCB using the U.FL (IPEX) connector or using the SMA connector. The PCB design is shown in the figure 2.

## V. Conclusion and Future Works

The measurement results show that the BG77 module with the selected technologies has no problem to connect in the typical locations where the user will be located. Even in worse radio conditions, the module was always able to connect to the network.

Depending on the technology, results differed quite a bit. In case of NB-IoT, it was very dependent on resulting network conditions and ECL levels. For ECL 0 and 1 module was able to communicate with OCPP server and transfer additional debug information over MQTT. When network conditions decreased and ECL changed to 2, modules started to experience reduced quality of service. MQTT was unable to transfer data in sufficient time (less than 15 seconds) and OCPP commands were substationally delayed. On the other hand, LTE Cat-M produced more promising results. Even at borderline radio conditions it was able to maintain connection with OCPP with usable delay between commands and MQTT debug informations were transmitted.

Fig. 4. Testing setup.

In future work it is planned to limit the amount of data sent and during charging. Especially the telemetry data. It is also planned to measure the exact response time to commands. Different antennas will be tried to improve the transmission parameters. Subsequently, more detailed measurements will be taken and further modifications will be proposed to optimize the operation.

## References


[1] European Council, "Fit for 55: The EU plan for a green transition," https://www.consilium.europa.eu/en/policies/green-deal/fit-for-55-the-eu-plan-for-a-green-transition/, 2021, 1.3.2024.

[2] M. Metais, O. Jouini, Y. Perez, J. Berrada, and E. Suomalainen, "Too much or not enough? planning electric vehicle charging infrastructure: A review of modeling options," *Renewable and Sustainable Energy Reviews*, vol. 153, p. 111719, 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S136403212100993X

[3] J. H. Lee, D. Chakraborty, S. J. Hardman, and G. Tal, "Exploring electric vehicle charging patterns: Mixed usage of charging infrastructure," *Transportation Research Part D: Transport and Environment*, vol. 79, p. 102249, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S136192091831099X

[4] I. Mahmud, M. B. Medha, and M. Hasanuzzaman, "Global challenges of electric vehicle charging systems and its future prospects: A review," *Research in Transportation Business Management*, vol. 49, p. 101011, 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S221053952300069X

[5] V. Moghaddam, I. Ahmad, D. Habibi, and M. A. Masoum, "Dispatch management of portable charging stations in electric vehicle networks," *eTransportation*, vol. 8, p. 100112, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2590116821000102

[6] C. Zhao and X. You, "Research and implementation of ocpp 1.6 protocol," in *Proceedings of the 2017 2nd International Conference on Machinery, Electronics and Control Simulation (MECS 2017)*. Atlantis Press, 2016/06, pp. 283–291. [Online]. Available: https://doi.org/10.2991/mecs-17.2017.149

[7] AutoNaBijeni.cz, "Cubish portabox - přenosné nabíječky," https://www.autonabijeni.cz/prenosne-nabijecky/cubish-portabox/, 2023, 28.2.2024.

[8] M. Vaezi, A. Azari, S. R. Khosravirad, M. Shirvanimoghaddam, M. M. Azari, D. Chasaki, and P. Popovski, "Cellular, Wide-Area, and Non-Terrestrial IoT: A Survey on 5G Advances and the Road Toward 6G," *IEEE Communications Surveys & Tutorials*, vol. 24, no. 2, pp. 1117–1174, 2022.

[9] S. Painuly, S. Sharma, and P. Matta, "Future Trends and Challenges in Next Generation Smart Application of 5G-IoT," in *2021 5th International Conference on Computing Methodologies and Communication (ICCMC)*, 2021, pp. 354–357.

[10] S. Ugwuanyi, G. Paul, and J. Irvine, "Survey of IoT for Developing Countries: Performance Analysis of LoRaWAN and Cellular NB-IoT Networks," *Electronics*, vol. 10, no. 18, p. 2224, 2021.

[11] S. W. H. Shah, A. N. Mian, A. Aijaz, J. Qadir, and J. Crowcroft, "Energy-Efficient MAC for Cellular IoT: State-of-the-Art, Challenges, and Standardization," *IEEE Transactions on Green Communications and Networking*, vol. 5, no. 2, pp. 587–599, 2021.

[12] R. Mozny, P. Masek, M. Stusek, K. Zeman, A. Ometov, and J. Hosek, "On the Performance of Narrow-band Internet of Things (NB-IoT) for Delay-tolerant Services," in *2019 42nd International Conference on Telecommunications and Signal Processing (TSP)*. IEEE, 2019, pp. 637–642.

[13] O. Liberg, M. Sundberg, Y.-P. E. Wang, J. Bergman, J. Sachs, and G. Wikström, *Cellular Internet of Things: From Massive Deployments to Critical 5G Applications*, second edition ed. Academic Press, 2020.

[14] M. Stusek, D. Moltchanov, P. Masek, K. Mikhaylov, O. Zeman, M. Roubicek, Y. Koucheryavy, and J. Hosek, "Accuracy Assessment and Cross-validation of LPWAN Propagation Models in Urban Scenarios," *IEEE Access*, vol. 8, pp. 154 625–154 636, 2020.

[15] GSMA, "NB-IoT Deployment Guide to Basic Feature set Requirements," https://www.gsma.com/, GSMA, White Paper, June 2019.

[16] GSMA, "LTE-M Deployment Guide to Basic Feature Set Requirements," https://www.gsma.com/, GSMA, White Paper, June 2019.

[17] Ultima Payments a.s., "Smart fuel pass - your digital fueling solution," https://smartfuelpass.com/, 2023, 25.2.2024.

# Reliability of WBG Transistors

Jan Horký
*Department of Microelectronics*
*Brno University of Technology*
Brno, Czech Republic
146830@vut.cz
0009-0000-3319-8410

***Abstract*—Silicon Carbide (SiC) and Gallium Nitride (GaN) transistors have the potential to revolutionize power electronics industry. They have better resistance, size, and efficiency than current Silicon (Si) transistors. On the other hand, as Silicon FET has been used for several decades, its reliability and robustness has been well known and documented. The aim of this paper is to present a review of reliability and radiation issues of GaN and SiC transistors in order to mitigate potential issues in future designs.**

***Keywords—Bandgap, Si, GaN, SiC, Transistor, HEMT, Radiation, Reliability, Robustness.***

## I. Introduction

There is always a demand for electronic devices to be smaller, cheaper, and more efficient. As the traditional Silicon transistors are being pushed to their theoretical limits, every new advancement brings only a small improvement, but with unbearable cost. Therefore, new types of semiconductors and transistors had to be found. New Silicon Carbide (SiC) and Gallium Nitride (GaN) transistors seem to be the most promising. They deliver higher efficiency and switching speeds, lower resistance, and better radiation immunity [1]. As new types of devices, they also bring uncertainty and risks in terms of their reliability and radiation hardness [1].

## II. Wide Bandgap

The band gap can be described as the energy required to excite electrons from the valence band to the conduction band. The greater the band gap is, the greater the energy is needed, hence the devices with greater bandgap can withstand higher voltages, temperatures, and radiation.

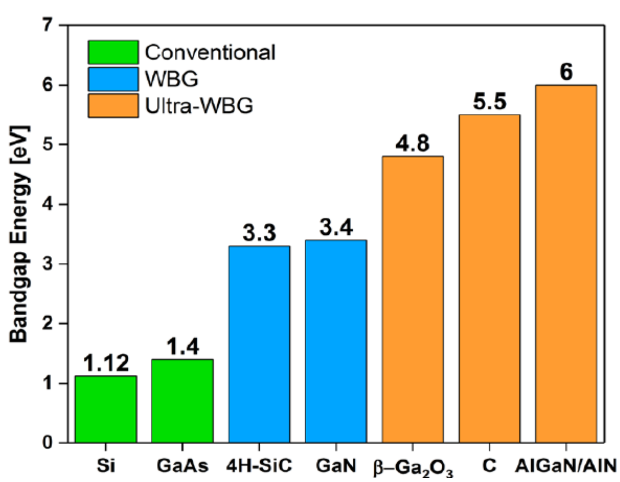



Figure 1: Bandgap values of various materials [2].

Figure 1 shows the bandgap values of various semiconductors. Silicon Carbide and Gallium Nitride semiconductors have three times higher energy than the Silicon ones while other semiconductors like the Gallium Oxide ones have even higher values.

## III. Silicon Carbide Transistor

A Silicon Carbide transistor has a similar structure as a Silicon one, as shown in Figure 2. As a SiC FET has a greater bandgap, it can be smaller for the same voltage rating, which significantly reduces its series resistance. Its high voltage rating, current capability, and switching speeds make a SiC FET ideal for power applications.

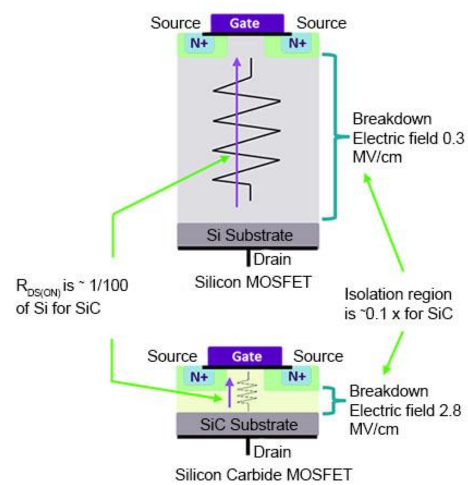



Figure 2: Structure of Silicon and Silicon Carbide MOSFET [3].

As the structure is similar, so are their reliability issues and radiation effects. Also, reliability assessment tests can be the same for both Si and SiC transistors [4].

## IV. Gallium Nitride Transistor

A Gallium Nitride transistor has a different structure than Si and SiC transistors. GaN transistor is formed by a junction between two materials with different bands. This junction forms a quantum well and two-dimensional electron gas (2DEG) inside the well. 2DEG have minimal resistance and high mobility. Therefore, GaN transistor is often described as a high electron mobility transistor (HEMT) and it can achieve greater switching speeds than it is possible with Si FET.

To have a normally-off device, E-GaN or Cascode D-GaN structure is used. Cascode uses normally-on GaN HEMT in series with traditional low voltage Si MOSFET. The MOSFET is used for switching the GaN HEMT and the GaN HEMT is used to block high voltages. The benefit is a high gate voltage rating, making the device easy to use with any ordinary gate driver. On the other hand, this solution increases the series resistance and lowers the efficiency, as the Si MOSFET has its internal body diode.

The E-GaN HEMT is created by placing the p-GaN or p-AlGaN layer under the gate to create a depletion region. Then the channel can be created by applying a positive voltage to the gate. The voltage needed to turn on the device is low, with a significant conduction above 1.7 V [1]. This makes the device susceptible to accidental turn-on. Also, as the gate breakdown voltage is low, specialized gate drivers must be used.

## V. Comparison Of Si, GaN and SiC Transistors

Different materials and bandgap values affect the parameters of all transistors. Figure 3 shows maximal theoretical values in five different categories, with GaN and SiC semiconductors superior to Si one in every category [2]. In theory, GaN semiconductor can withstand a higher electric field than SiC one, while in reality, SiC FETs have a higher voltage rating than GaN HEMTs. To minimize the price of GaN HEMTs, manufacturers grow transistors on silicon substrates. However, as there is a large mismatch in the thermal expansion and in the lattice constant, defects are produced and limit the breakdown voltage [5]. Different types of substrates can be used to improve the voltage rating, but with significant cost increase.

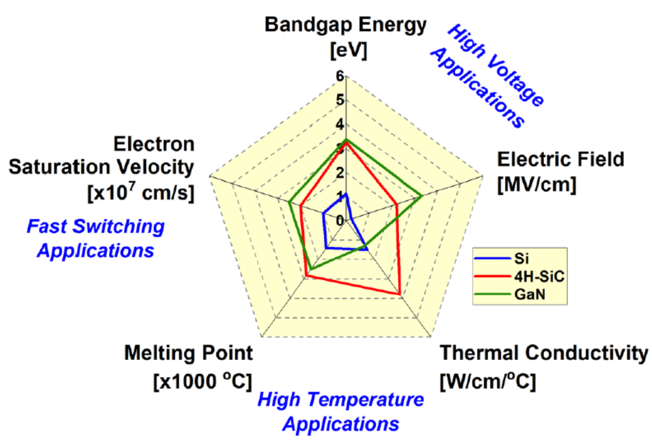



Figure 3: Theoretical limits of Si, SiC and GaN semiconductors [2].

Figure 3 also shows superior thermal conductivity of SiC semiconductor. SiC FET's $R_{DSON}$ has the lowest temperature dependency. With temperature increased from 25 °C to 150 °C, the Si FET has an $R_{DSON}$ increase of 160 %, GaN about 100 % increase while SiC FET has increase only about 50 % [6].

Lateral structure of GaN HEMT also limits the current capability. Most of the commercially available GaN HEMTs are rated up to tens of amperes, whereas SiC FETs up to hundreds of amperes. Also, Si and SiC FETs are more reliable and rugged in linear mode, which is present during switching transitions [7]. Furthermore, they can withstand the same current and voltage for longer period, making their Safe Operating Area superior to GaN HEMT's [7].

## VI. SC and OV Robustness of WBG Transistors

As the structure of Si and SiC FET is similar, so is their reliability. However, as GaN HEMT's structure is different, other types of issues can occur. The main difference is GaN HEMT's low short-circuit (SC) and overvoltage (OV) robustness and dynamic breakdown voltage. Typically required SC robustness is 10 µs withstanding time at maximum application-specific voltage and current [8], while the reported SC capability is under 1 µs at a voltage higher than 350 V [4]. For lower voltages, the robustness can increase to more than 10 µs, but fail again if SC pulses are repetitive due to dynamic $R_{DSON}$ [4]. GaN HEMTs mainly fail thermally in long-duration, low-voltage SC tests [9]. In high voltage tests, failure is caused by the high electric field induced by the hole accumulation underneath the gate in which the holes are generated from impact ionization. Failure can also be caused by high carrier density and electric field crowding at the drain-side gate edge. If SC fault is expected, a protection circuit with a 100-200 ns reaction time shall be used. [4]

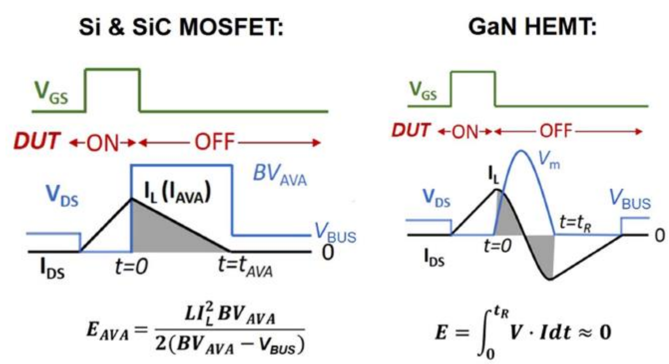



Figure 4: Surge energy process in Si, SiC and GaN transistors [4].

Overvoltage and surge energy robustness are also important for power MOSFETs. Si and SiC transistors are inherently robust due to their avalanche capability. If they are subject to surge energy their $V_{DS}$ raises until avalanche breakdown occurs and $V_{DS}$ becomes constant. Then the energy is dissipated as heat and current decreases to zero. The avalanche energy is described as the maximum energy that a MOSFET can dissipate without destruction. This process can be repeated without changing the device's parameters. However, GaN transistors do not have these properties [10].

During the surge energy burst, the resonance between GaN HEMT's output capacity and parasitic inductance creates a rapid increase in $V_{DS}$ [4]. As there is no avalanche breakdown, energy is not dissipated until the resonance voltage turns negative. Negative resonance voltage turns on the transistor and energy is depleted by reverse conduction. If the increase in $V_{DS}$ is too high, the voltage rating of the device can be exceeded, leading to the transistor's failure [11]. The comparison of both processes is shown in Figure 4.

## VII. DYNAMIC EFFECTS IN GAN TRANSISTORS

The main usage of GaN transistors is in fast-switching circuits, thus dynamic effects have great importance. Shift in the transistor's parameters may have a significant effect on overall efficiency and reliability.

One of the issues is the dynamic $R_{DSON}$. In the blocking state, leakage current electrons are generated in the transistor. During turn-on, both high voltage and high currents are present and they can induce hot electrons [12]. This process can cause charge trapping and decrease channel conductivity. Then the charge begins to be removed and $R_{DSON}$ settles back to its idle value. $R_{DSON}$ can rise multiple times higher after blocking high voltage. It impairs efficiency and increases junction temperature, particularly in high-frequency applications. The comparison of $R_{DSON}$ 2 µs after the blocking state is shown in Figure 5. The figure shows significant $R_{DSON}$ change, compared to Si MOSFET.

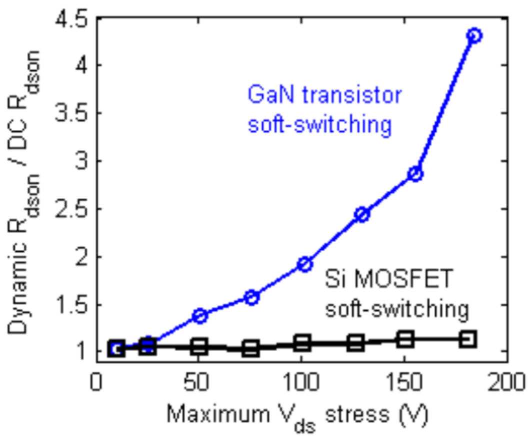



Figure 5: Dynamic $R_{DSON}$ in GaN and Si FET [13].

The size of the $R_{DSON}$'s increase is mainly caused by blocking voltage $V_{DS}$ and load current, as both affect electron trapping [4]. The increase reaches its maximum at 100-300 V for a 600 V rated transistor. The length of the de-trapping process can last up to several seconds and is affected by switching speed, frequency, and duty cycle. Therefore, the transistor shall be switched on as fast as possible.

Efficiency can also be decreased by output capacitance $C_{OSS}$ losses. The loss is produced when output capacitance is periodically charged and discharged [4]. The Figure 6 shows $C_{OSS}$ losses in GaN, SiC, and Si transistors. It can be seen that losses in GaN transistors are worse than in SiC ones in low-frequency applications. Losses rise significantly with frequencies above 10 MHz, up to 20 %. Apart from impairing efficiency they can cause significant temperature increases.

Another serious issue was reported in cascode D-GaN HEMTs. As the cascode is formed by Si and GaN transistors, a capacitance mismatch is formed. Also, in some devices, transistors are connected by bond wires with inherent parasitic inductance. During turn-off, this LC circuit can cause oscillations and switching losses. In some cases a false turn-on can occur, possibly destroying the device [14].

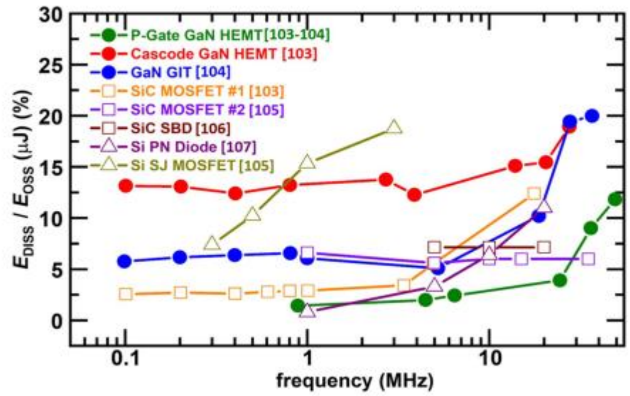



Figure 6: Ratios of the $C_{OSS}$ loss over the total $C_{OSS}$ energy [4].

## VIII. RADIATION

The greater band gap makes WBG transistors more immune to space radiation although this radiation hardness refers to Total Ionizing Dose (TID) only. Cumulative TID can change the device's parameters, shift thresholds, increase leakage currents, and decrease functionality. Standard commercial Si MOSFETs are radiation hardened up to 30 krad while a greater band gap makes SiC transistors robust up to 100 krad without any additional shielding or processes. SiC FETs can withstand higher radiation but with significant $C_{GS}$ changes above 300 krad. Radiation hardening in GaN transistors is dependent on gate design with normally-on transistors hardened up to 6 Mrad. E-GaN HEMT can withstand radiation up to 500 krad while cascode D-GaN HEMT is expected to have much lower radiation hardening as it is limited by Si MOSFET in the cascode. Future Ultra-WBG devices like ß-GaO transistors should have even higher radiation hardening with expected values up to 50 Mrad. [2]

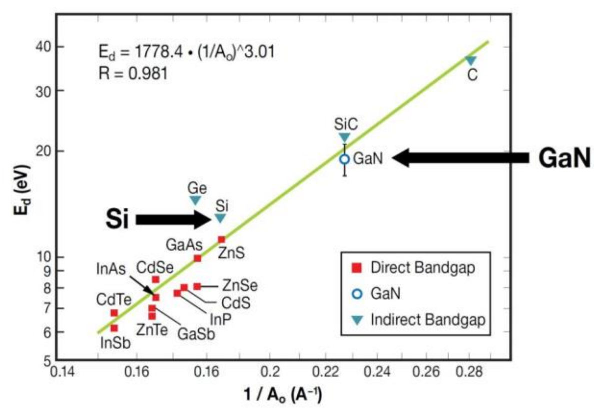



Figure 7: Displacement threshold energy for various materials [15].

When charged particles move through a semiconductor they can displace atoms and create defects. The Displacement Damage Dose (DDD) has long-term degradation characteristics as TID although has a different physical mechanism. In long-term exposure, device parameters can change if sufficient displacement occurs. It usually leads to an increase in $R_{DSON}$, threshold voltage, reduction in transconductance and saturation current, and other parameters [15]. GaN and SiC transistors have higher displacement threshold energy and therefore better reliability than Si ones, as shown in Figure 7. For GaN devices, the root cause is mainly the change in 2DEG mobility by carrier scattering from radiation-induced defects, and 2DEG density by charged traps that decrease carrier concentration [16]. Radiation can also increase gate lag, which limits the maximum operating frequency and reliability of GaN transistors. Gate lag can be created by donor or acceptor traps. It mainly depends on trap density, energy, and distribution [17]. Their effect on the transistor's drain current during turn-on is shown in Figure 8.

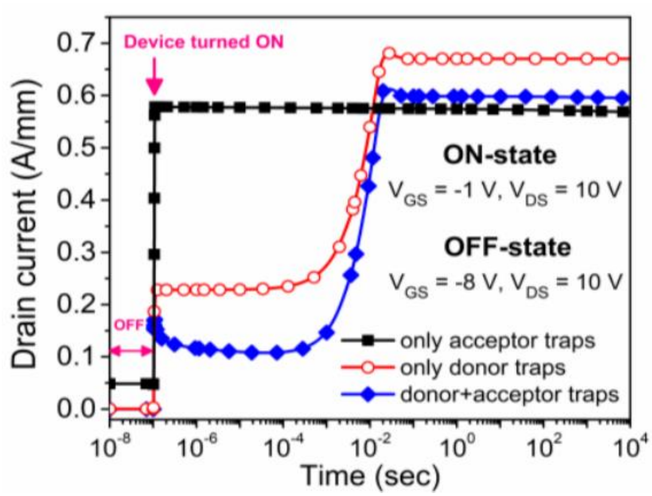



Figure 8: Effects of traps on the gate lag in GaN transistors [17].

The main issues in MOSFETs are Single Event Effects (SEE). When the device is struck by high-energy particle, ionization is caused, and a large number of electrons and holes is generated. It can degrade parameters and create glitches but it can also lead to device's destruction.

The most important SEEs in MOSFETs are Single Event Gate Rupture (SEGR), Single Event Burnout (SEB), and Single Event Dielectric Rupture (SEDR).

When the high energy particle strikes the MOSFET through the Gate and Body, electrons and holes are generated. If the transistor is in off-state, the positive voltage on the Drain removes most of the electrons and holes remain. The opposite principle applies to the Gate. A group of electrons on one side and a group of holes on the other side create an electric field. If the field is too strong, the isolation between the Gate and the Body is broken, which leads to SEGR and the destruction of the device. This issue applies to Si and SiC transistors, as there is a thin layer of oxide beneath the Gate. As the oxide is not present in e-GaN transistors, they do not suffer from SEGR. On the other hand, they suffer from SEDR where Source and Drain become shorted. The process is similar to SEGR but with a much smaller cross-section [2].

In Si and SiC transistors, the charged particle can strike the device and turn on the parasitic bipolar NPN transistor and create SEB. The NPN transistor can carry a large current. If the current is not limited or stopped, the device can be destroyed. The destruction can also occur in GaN transistors. They also suffer from SEB, but instead of opening parasitic NPN between Source and Drain, a high current is created by shorting the Drain and Substrate [2]. It happens for higher voltages than in Si and SiC devices but with higher cross-section [2]. High-energy particles can also permanently change leakage currents in Si, SiC, and GaN devices [2]. However, this change usually does not lead to device malfunction.

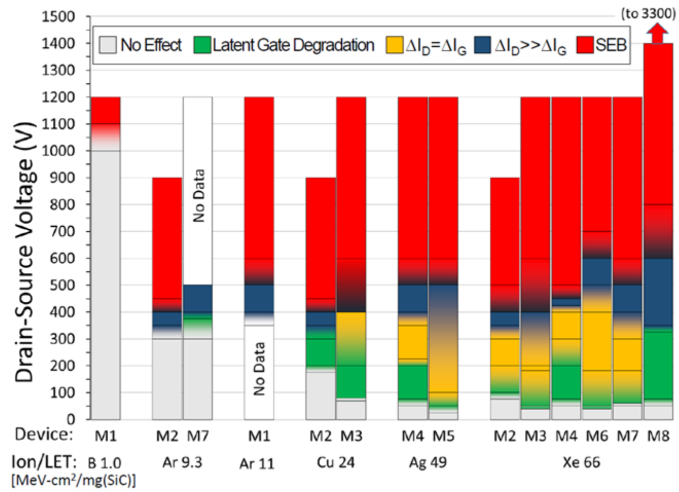



Figure 9: Radiation effects in SiC transistors [2].

Figure 9 shows the radiation effects on SiC transistors. Transistors are marked as Mx, column height denotes the transistor's voltage rating. Columns are put into groups, based on the particle's type and energy. The grey part of any column shows that if the corresponding Drain-Source voltage and particle energy was used, no radiation effect was observed. In the green part, latent gate degradation was shown, which led to an increase in gate current by at least 1 nA. In the yellow part, some devices show the formation of a Gate-Drain leakage current. For the higher voltages, the leakage current increased and changed its direction to Drain-Source, as shown in blue colour. If the Drain-Source voltage increased further the device experienced SEB and was destroyed (red colour). In summary, SiC transistors rated for $V_{DS}$ 1200 V should not be used at voltages above 500 V in any radiation environment with LETs > 10 MeV-cm2/mg [2]. The radiation effects depend on the device structure and therefore the radiation hardness can be improved. The SEB susceptibility can be reduced by making the epilayer thicker, which decreases the electric field [18]. That also leads to the degradation of $R_{DSON}$, therefore some compromise must be made. The Gate-Drain leakage current degradation (yellow colour) can be mitigated by common hardening techniques, for example, drain neck width reduction [18].

Figure 10 shows the SEB susceptibility comparison of SiC and Si transistors based on Drain-Source voltages and LETs. SiC transistors are generally less susceptible to SEB than unhardened silicon MOSFETs but SiCs are usually more susceptible to permanent non-destructive effects [2].

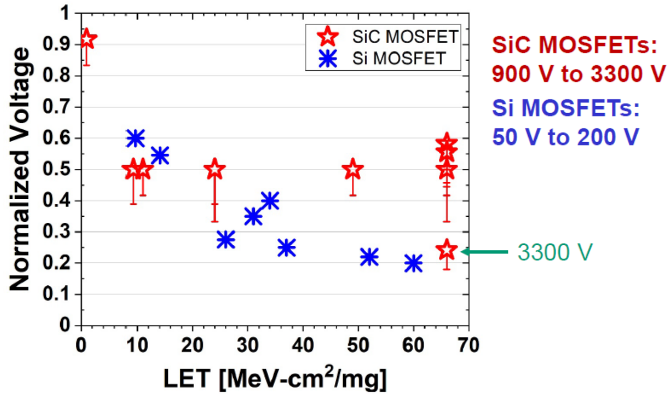



Figure 10: Single Event Burnout in Si and SiC MOSFETs [2].

For GaN devices, great variability exists. Their susceptibility depends heavily on their design and voltage rating. Most low-voltage transistors do not show any significant radiation effects. Some 100 V and 200 V rated devices show only some minor degradation while others show significant SEB susceptibility. In 600 V and higher-rated devices, SEB usually occurs at voltages above 50 % of their rating at LETs above 40 MeV-cm2/mg [2].

## IX. Conclusions

In this paper, Gallium Nitride and Silicon Carbide transistors were compared. Their parameters can make devices smaller, lighter, and more efficient. But their properties can also form new design challenges and reliability issues. Therefore, their reliability was studied.

In general, as SiC transistors have a similar structure to Si transistors, their reliability and robustness are also similar. SiC FETs benefit from a bigger band gap, which make them more radiation robust although robustness applies to the Total Ionizing Dose and Displacement Damage Dose. SiC FETs still suffer from Single Event Effects. To use them in the radiation environment, significant voltage derating is necessary.

GaN HEMT's overvoltage, overcurrent, and surge energy robustness is generally lower than in Si and SiC transistors, as GaN HEMT does not have avalanche capability. In some cases, high-energy bursts can lead to the transistor's destruction.

High electron mobility makes GaN transistor ideal for high-frequency applications. However, at some point, increasing the frequency starts lowering the overall efficiency, as dynamic effects are present. GaN transistors suffer from dynamic $R_{DSON}$, which can be significant at high frequencies or voltages. Output capacitance losses also contribute to lower efficiency.

GaN transistor also has better radiation immunity than Si one. Different structure also leads to different radiation effects. As e-GaN does not have a gate oxide, SEGR is not possible. GaN transistor also suffers from SEB like SiC one and Si one but the mechanism is different and happens at higher voltages.

Both SiC and GaN transistors seem to be promising for future space and commercial applications but mentioned reliability issues must be treated carefully.

## References

[1] G. Longobardi, "GaN for power devices: Benefits, applications, and normally-off technologies," in *2017 International Semiconductor Conference (CAS)*, Sinaia, Romania, Oct. 2017, pp. 11–18. doi: 10.1109/SMICND.2017.8101144.

[2] J. Lauenstein, "Wide Bandgap Power – SiC, GaN – Radiation Reliability," presented at the *Nuclear and Space Radiation Effects Conference*, Santa Fe, USA, Nov. 2020.

[3] Editorial Team, "What are Silicon Carbide (SiC) MOSFETs?," everythingPE. https://www.everythingpe.com/community/what-are-silicon-carbide-sic-mosfets (accessed Jan. 10, 2024).

[4] J. P. Kozak *et al.*, "Stability, Reliability, and Robustness of GaN Power Devices: A Review," *IEEE Trans Power Electron*, vol. 38, no. 7, pp. 8442–8471, Jul. 2023, doi: 10.1109/TPEL.2023.3266365.

[5] A. C. Liu *et al.*, "The Evolution of Manufacturing Technology for GaN Electronic Devices," *Micromachines*, vol. 12, no. 7. MDPI AG, Jul. 01, 2021. doi: 10.3390/mi12070737.

[6] E. O. Prado, P. C. Bolsi, H. C. Sartori, and J. R. Pinheiro, "An Overview about Si, Superjunction, SiC and GaN Power MOSFET Technologies in Power Electronics Applications," *Energies (Basel)*, vol. 15, no. 14, Jul. 2022, doi: 10.3390/en15145244.

[7] Infineon, "Understanding the trade-offs between Si & GaN in high-reliability applications," Appl. Note 1225.

[8] R. S. Chokhawala, J. Catt, and L. Kiraly, "A discussion on IGBT short-circuit behavior and fault protection schemes," *IEEE Trans Ind Appl*, vol. 31, no. 2, pp. 256–263, 1995, doi: 10.1109/28.370271.

[9] A. Castellazzi, A. Fayyaz, S. Zhu, T. Oeder, and M. Pfost, "Single pulse short-circuit robustness and repetitive stress aging of GaN GITs," in *2018 IEEE International Reliability Physics Symposium (IRPS)*, Mar. 2018, pp. 4E.1-1-4E.1-10. doi: 10.1109/IRPS.2018.8353593.

[10] R. Zhang and Y. Zhang, "Power device breakdown mechanism and characterization: review and perspective," *Jpn. J. Appl. Phys*, vol. 62, Feb.2023, Art. no. SC0806, doi: 10.35848/1347-4065/acb365.

[11] R. Zhang, J. P. Kozak, Q. Song, M. Xiao, J. Liu, and Y. Zhang, "Dynamic Breakdown Voltage of GaN Power HEMTs," in *2020 IEEE International Electron Devices Meeting (IEDM)*, Dec. 2020, pp. 23.3.1-23.3.4. doi: 10.1109/IEDM13553.2020.9371904.

[12] M. Meneghini *et al.*, "Trapping in GaN-based metal-insulator-semiconductor transistors: Role of high drain bias and hot electrons," *Appl. Phys. Lett.*, vol. 104, no. 14, Apr. 2014, Art. no. 143505, doi: 10.1063/1.4869680.

[13] B. Lu, T. Palacios, D. Risbud, S. Bahl, and D. I. Anderson, "Extraction of Dynamic On-Resistance in GaN Transistors: Under Soft- and Hard-Switching Conditions," in *2011 IEEE Compound Semiconductor Integrated Circuit Symposium (CSICS)*, Oct. 2011, pp. 1–4. doi: 10.1109/CSICS.2011.6062461.

[14] P. Xue, L. Maresca, M. Riccio, G. Breglio, and A. Irace, "Investigation on the Short-Circuit Oscillation of Cascode GaN HEMTs," *IEEE Trans Power Electron*, vol. 35, no. 6, pp. 6292–6300, Jun. 2020, doi: 10.1109/TPEL.2019.2947274.

[15] J. Shepard, "GaN in orbit and beyond." Power Electronic Tips. https://www.powerelectronictips.com/gan-in-orbit-and-beyond-faq/ (accessed Jan. 20, 2024).

[16] S. J. Pearton, F. Ren, E. Patrick, M. E. Law, and A. Y. Polyakov, "Review—Ionizing Radiation Damage Effects on GaN Devices," *ECS Journal of Solid State Science and Technology*, vol. 5, no. 2, pp. Q35–Q60, Nov. 2016, doi: 10.1149/2.0251602jss.

[17] K. De and G. Dutta, "Investigation of Trap Induced Gate Lag Phenomenon in AlGaN/GaN High Electron Mobility Transistors," in *2018 4th IEEE International Conference on Emerging Electronics (ICEE)*, Dec. 2018, pp. 1–5. doi: 10.1109/ICEE44586.2018.8938017.

[18] Xingguang Zhu, "Radiation Hardness Study on SiC Power MOSFETs," in *International Conference on Silicon Carbide and Related Materials*, Washington, DC: ICSCRM, Sep. 2017.

# Investigating alternative carbonaceous materials for cost-effective anodes for sodium-ion batteries

Antonín Šimek
*Department of electrical and electronic Technology*
*Faculty of Electrical Engineering and Communication*
Brno, Czech Republic
Antonin.Simek1@vut.cz

Ondřej Čech
*Department of electrical and electronic Technology*
*Faculty of Electrical Engineering and Communication*
Brno, Czech Republic
cechondrej@vut.cz

Tomáš Kazda
*Department of electrical and electronic Technology*
*Faculty of Electrical Engineering and Communication*
Brno, Czech Republic
kazda@vutbr.cz

***Abstract*—This study explores the potential of alternative carbonaceous materials as cost-effective candidates for sodium-ion anodes. As the demand for energy storage solutions continues to grow, the search for affordable and sustainable battery components is becoming increasingly important. Carbon-based materials appear promising due to their abundance, low cost and favourable electrochemical properties. In this investigation, various carbon-based materials sourced from Czech Republic, including pitch coke, petroleum coke and dry quenched coke, are evaluated for their suitability as anode materials. The results of this study contribute to ongoing efforts to develop cost-effective and environmentally friendly energy storage solutions based on sodium-ion battery technology.**

***Keywords*—sodium-ion batteries, anodes, carbonaceous materials, electrochemistry, sustainability**

## I. Introduction

With the growing demand for energy, gradual depletion of fossil fuel reserves, and the imperative to reduce carbon emissions, the significance of clean energy generation and storage has surged. Renewable energy sources offer cost-effective, sustainable, and emission-free alternatives to traditional fuels [1]. Lithium-ion batteries (Li-ion) have rapidly diversified their applications, powering everything from portable electronics to stationary storage and electric vehicles. Yet, concerns about renewable energy integration, cost escalations, and geopolitical dependencies on Li-ion materials have spurred interest in alternative technologies like sodium-ion batteries (Na-ion) [2].

Na-ion batteries, leveraging abundant and ethically sourced materials, emerge as promising complementary technology to Li-ion. Despite their lower theoretical capacity, larger ion radius and higher standard electrode potential of sodium, Na-ion batteries offer viable solutions for stationary storage and affordable electromobility [3]. A comparison of the basic properties of Li-ion and Na-ion technology is shown in Table I. While Europe lags in Li-ion battery manufacturing, there's an opportunity to catch in Na-ion technology with concerted research and development efforts. Thus, this paper holds immense significance in keeping pace with global advancements in energy storage.

TABLE I. Comparison of the basic properties of Li-ion and Na-ion technology and Li and Na as an element [1-4]

| Property | Li-ion | Na-ion | Unit |
|---|---|---|---|
| Energy density | 250 | 130 | $Wh.kg^{-1}$ |
| Cycle life | ~1000 | ~500 | Cycles |
| Voltage | 3.6-4.2 | 2.0-3.5 | V |
| Price | 100 | 50-70 | % |
| Ionic radius of Li/Na | 0.76 | 1.02 | Å |
| Electrochemical potential of Li/Na | –3.04 | –2.71 | V |
| Abbundance of Li/Na | 0.0017 | 2.4 | % |
| Melting point of Li/Na | 180.5 | 97.8 | °C |
| Density of Li/Na | 0.53 | 0.97 | $g.cm^{-3}$ |

This work focuses on the development of anode materials, as graphite cannot be effectively used as an anode material in sodium batteries due to the larger ionic radius of sodium ions compared to lithium ions [1, 4]. Sodium ions are larger in size (with an ionic radius of approximately 1.02 Å) compared to lithium ions (with an ionic radius of approximately 0.76 Å) [1]. This difference in size prevents efficient intercalation of sodium ions into the graphite structure, leading to poor performance and battery instability. As a result, alternative anode materials compatible with larger sodium ion sizes, such as hard carbon and other carbonaceous materials that have disordered structures, are typically used in sodium batteries [3, 4]. These materials typically achieve capacities in the range of 200-250 $mAh.g^{-1}$, which is significantly less than the theoretical 372 $mAh.g^{-1}$ for graphite in Li-ion batteries [3, 4].

## II. Active materials used

The following three samples obtained from a local supplier in the Czech Republic were chosen for this research into alternative, cost-effective active anode materials: pitch coke (PC), calcined petroleum coke (CPC) and dry quenched coke (DQC). These materials are classified as coke materials, which are solid carbonaceous residues obtained from the thermal decomposition of organic materials such as coal, oil or biomass

[5]. These materials typically range from $100 to $500 per tonne, considerably cheaper than hard carbon, priced around $500 to $2000 per tonne. Graphite's cost varies: natural graphite is approximately $1000 to $3000 per tonne, while synthetic graphite ranges from $5000 to $20000 per tonne. The general composition of these materials is given in Table II.

TABLE II. GENERAL COMPOSITION OF MATERIALS

| **Material** | **Material composition (atomic %)** | | | |
|---|---|---|---|---|
| | ***Carbon*** | ***Suplhur*** | ***Ahs*** | ***Volatile matter*** |
| PC | 90-95 | 0.5-6 | 0.5-5 | 3-12 |
| CPC | 85-95 | 1-6 | 0.1-5 | 8-15 |
| DQC | 85-95 | 0.5-3 | 0.5-3 | 2-8 |

### A. *Pitch coke*

Pitch coke, a derivative of coke production, arises during the carbonization process. It results from the treatment of coke oven gases, which are extensively utilized in various industrial processes. It is characterized by its high carbon content and typically contains impurities such as sulphur and ash. Pitch coke is often used in the production of carbon electrodes for the steel and aluminium industries due to its high carbon content and good electrical conductivity. [6]

### B. *Calcined petroleum coke*

Calcined petroleum coke, a high-quality carbonaceous material, finds extensive usage in various industrial applications. Primarily utilized in the production of graphite electrodes, it serves as a key ingredient in carbon paste products, utilized in electrode and anode manufacturing processes. Additionally, it acts as a decarburize in foundries and steel mills, aiding in the control of carbon content in steel production. With a fixed carbon content ranging between 97 to 99.5%, calcined petroleum coke stands as a valuable resource in the metallurgical and manufacturing sectors. [7]

### C. *Dry quenched coke*

Dry quenched coke, also known as breeze coke, is a by-product of coke production that undergoes a quenching process without water. This type of coke is typically produced by heating coal in the absence of air, followed by rapid cooling using an inert gas or by simply allowing it to cool naturally. Dry quenched coke is characterized by its porous structure and is commonly used as a fuel in blast furnaces, foundries, and other industrial processes due to its high calorific value and good reactivity. [8]

## III. CHARACTERISATION OF ACTIVE MATERIALS

Grinding was carried out prior to the analyses as the materials were supplied in grain sizes of 1-5 mm, which is not suitable for use in anode materials. A two-step milling process was used, with the first step carried out in a Fritsch Analysette 3 SPARTAN vibrating ball mill using a steel grinding bowl and a steel ball with a diameter of 5 cm. Grinding was carried out in three cycles of 20 minutes each. The second step was carried out in a Fritsch Pulverisette 7 planetary mill using zirconium-oxide grinding bowl and zirconium-oxide balls with a diameter of 5 mm (balls to sample ratio 10:1). Absolute ethanol was added to the grinding bowl to prevent tamping of the material. The grinding procedure was set to 5×30 min, 500 rpm, 15 min pause.

After grinding the samples, scanning electron microscope (SEM) images were taken and energy dispersive spectroscopy (EDX) was performed. For this, a Tescan Lyra 3 electron microscope equipped with a Brooker XFlash 5010 EDX detector was used. Results of EDX analysis can be seen in Table III, SEM images of materials used can be seen in Fig. 1. Raman spectra were also measured using a Witec Alpha 300R instrument. The spectra were measured using a green laser with a wavelength of 532 nm and a power of 5 mW. Measured Raman spectra can be seen in Fig. 2. X-ray powder diffraction (XRD) was performed on a Rigaku MiniFlex 600 with a D/teX Ultra detector. The results of this analysis are shown in Fig. 3.

TABLE III. RESULTS OF EDX ANALYSIS

| **Material** | **Content of element (atomic %)** | | | | |
|---|---|---|---|---|---|
| | ***C*** | ***S*** | ***Al*** | ***Si*** | ***Ca*** |
| PC | 99.06 | 0,94 | - | - | - |
| CPC | 98.59 | 1.41 | - | - | - |
| DQC | 95.70 | 0.35 | 1.44 | 2.18 | 0.32 |

As can be seen from the results of the EDS analysis, the cleanest sample is PC, which contains the least amount of sulphur (S) and highest amount of carbon. On the other hand, DQC contains the most impurities, which in addition to sulphur also contains calcium (Ca), aluminium (Al) and silicon (Si).

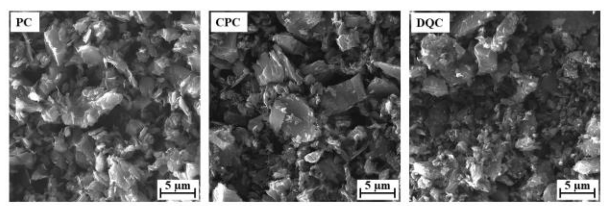



Fig. 1. SEM images of different coke materials after milling

SEM images show consistent particle shape across all samples, with grain sizes measured in micrometers. Images were captured at 10000x magnification, with a field of view of 27.7 µm.

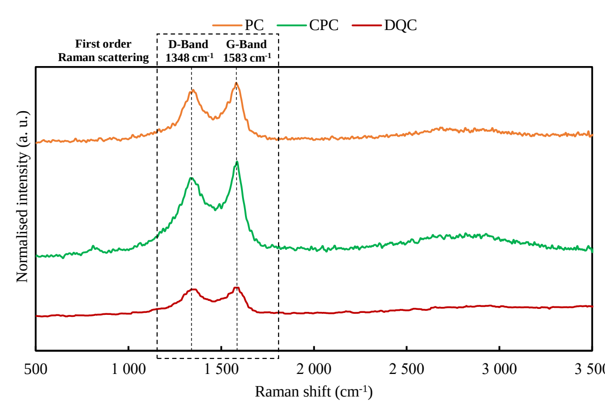



Fig. 2. Normalised Raman spectra of different coke materials

In the measured Raman spectra, two main peaks can be observed from the first order Raman scattering, where incident photons interact with the sample by inducing vibrational modes without any change in the electronic energy level. The peak located at approximately Raman shift 1348 $cm^{-1}$ is referred to as the D-band and according to [9] is associated with structural defects and disorder. It arises due to the presence of $sp^3$ hybridized carbon atoms, which result from structural defects, such as vacancies, grain boundaries, or edge defects. The intensity of the D-band provides information about the degree of disorder or defects present in the carbon material. The second main peak, referred to as the G-band, is located at a Raman shift of approximately 1583 $cm^{-1}$ and is associated with the graphitic structure and symmetry of the material. It originates from the in-plane vibration of $sp^2$ hybridized carbon atoms in a hexagonal lattice. The G-band is a signature of the ordered graphitic structure and represents the stretching of the C-C bond in the graphene layers. The intensity and position of the G-band are sensitive to the number of graphene layers and the degree of graphitization.

When comparing the Raman spectra, it is possible to observe the differences of the individual carbonaceous materials. CPC achieves the highest intensity of the observed peaks and also has the largest difference between the D-band and G-band peaks. This material has a greater degree of graphitization compared to the other two. On the other hand, DQC achieves the lowest intensity, which also does not have much difference between D-band and G-band. This material is therefore the most disordered.

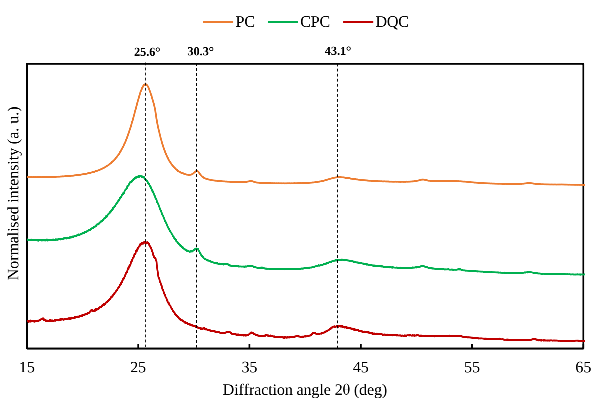



Fig. 3. Normalised XRD spectra of different coke materials

From the measured XRD spectra it is possible to observe the main peak located at 25.6° which according to [9] is typically associated with the (002) reflection of graphitic carbon materials. This peak corresponds to the interlayer spacing between the graphene layers in the crystalline structure. PC and DQC than have similar crystalline structures or arrangements of graphene layers, resulting in a consistent interlayer spacing. CPC has this peak shifted to around 24°, which indicates a slight difference in the interlayer spacing compared to the other samples. A peak at around 43° corresponds to the (100) reflection of graphitic carbon materials. This peak arises from the stacking of graphene layers within the crystalline structure.

## IV. Electrochemical Characterization

### A. *Preparation of negative electrodes*

To verify the electrochemical properties of the investigated materials, it was necessary to prepare negative electrodes. These electrodes were prepared using a standard procedure, i.e., using a magnetic stirrer, Polyvinylidenefluoride (PVDF) binder was dissolved together with N-Methyl-2-pyrrolidone (NMP) solvent. Subsequently, SuperP was added as a conductive additive. Finally, the investigated active material (PC, CPC and DQC) was added. Resulting mass ratio of PVDF:SuperP:Active material was 10:10:80. After mixing for approximately 24 hours, a paste was formed, which was then coated onto a 25 µm thick aluminium foil using K-hand coater rods, resulting in the formation of a 200 µm thick layer. Subsequently, the layer was dried at 50°C in the presence of air. After drying, electrodes with a diameter of 18 mm were die-cut and then pressed with a pressure of 30 kN. After pressing, the electrodes were dried in a vacuum dryer at 110°C for 24 hours. Subsequently, the electrodes were transferred to a Jacomex glove box in the presence of an argon inert atmosphere (both $O_2$ and $H_2O$ bellow 1 ppm).

### B. *Composition of electrochemical half-cells*

The electrochemical properties were subsequently measured using electrochemical half-cells that were assembled in EL-CELL® electrochemical test cells. The half-cells were assembled in a glove box under argon inert atmosphere. This half-cell consisted of metallic sodium from supplier Sigma Aldrich (counter electrode), a Whatman GF/C® glass separator, 130 µl of ionically conductive electrolyte (custom preparation), and a prepared negative electrode (working electrode). The electrolyte used was unipolar sodium hexafluorophosphate dissolved in ethylene carbonate and propylene carbonate in a 1:1 volume ratio.

### C. *Electrochemical meassurements*

Cyclic voltammetry (CV) was measured to determine the electrochemical activity of the electrodes. CV was measured at four different scan rates (5; 1; 0.5 and 0.1 $mV.s^{-1}$) in a potential window of 0.01-2.5 V. The CV results are shown in Fig. 4.

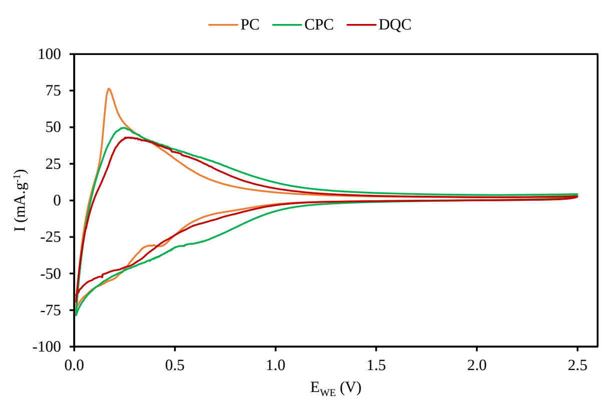



Fig. 4. Results of CV performed at 0.1 $mV.s^{-1}$ scan rate

From the CV results it can be seen that the best results are achieved by PC, which has a sharp anodic peak at a potential of 0.17 V. Moreover, this waveform is very similar to that of hard carbon measured in by X. Zhang et al. [10]. For the other two samples, this anodic peak is not very sharp and is shifted towards a higher potential. The position of the anodic peaks varied for each sample by both the working electrode potential ($E_{WE}$) and the current response (I). The position of the anodic peaks is given in Table IV.

TABLE IV. THE POSITION OF THE ANODIC PEAKS

| Material | Potential (V) | Current response ($mAh.g^{-1}$) |
|---|---|---|
| PC | 0.17 | 76.43 |
| CPC | 0.22 | 49.62 |
| DQC | 0.25 | 43.00 |

Subsequently, galvanostatic cycling with potential limitation (GCPL) was performed, during which cyclic charging and discharging of the electrodes with different C-rates was performed. Initially, 10 cycles were performed at 0.1C load. Subsequently, the load was increased to 0.2C; 0.5C; 0.8C and 1C. At each of these loads, 5 cycles were performed. The load was then stepped down to 0.1C in the same increments. GCPL was also measured in the range of 0.01-2.5 V. The GCPL results are shown in Fig. 5 (discharge capacity vs cycle number) and in Fig. 6 (Working electrode potential-$E_{WE}$ vs discharge capacity).

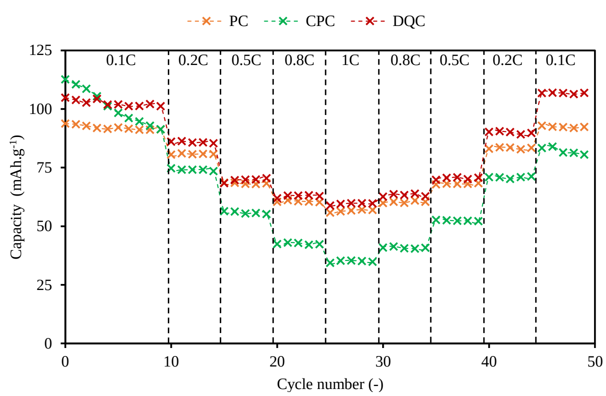



Fig. 5. Comparison of discharge capacities during cycling at different C-rates

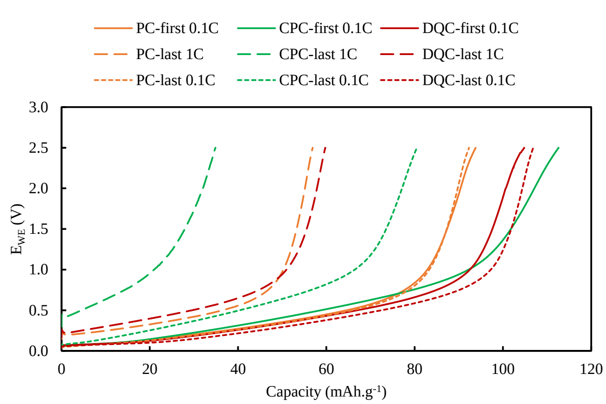



Fig. 6. Comparison of discharge curves obtained from the first and last 0.1C cycle and from the last 1C cycle

From the measured GCPL results it is possible to observe, that CPC material achieves the highest discharge capacity of 112.60 $mAh.g^{-1}$ in the first cycle at 0.1C load. However, this discharge capacity decreased very sharply during the first 10 cycles at 0.1C (18.85% decrease). For PC this decrease was only 2.65% and for DQC the decrease was 3.46%. As the load was gradually increased up to 1C, the DQC sample reached the highest capacity of 59.73 $mAh.g^{-1}$. At the last 1C cycle, there was a 43.01% decrease in discharge capacity compared to the first 0.1C cycle. For sample PC, this decrease was 39.34% and for sample CPC, the decrease was 69.03%.

As the load was gradually decreased up to 0.1C, the highest capacity of 106.88 $mAh.g^{-1}$ was achieved by DQC, which showed an increase in capacity of 1.97% compared to the first cycle. This phenomenon is most likely due to the gradual increase in charging and discharging efficiency. This increase in efficiency may be due to the gradual formation of the solid electrolyte interphase (SEI) passivation layer which forms when the surface of the electrode comes into contact with the liquid electrolyte. This layer can initially impede ion transport and reduce capacity, but over time, it may stabilize and become more beneficial for battery performance, leading to improved capacity retention and possibly increased capacity at the end of cycling [11]. It is also possible that during the initial cycles, the carbon material at the negative electrode undergoes a conditioning or activation process that improves its electrochemical properties over time. This may include removal of surface impurities, better infiltration of the electrolyte into the electrode structure, or a change in the atomic structure of the carbon to provide more active sites for electrochemical reactions [11]. For other materials tested, this decrease between the first and last cycle was 1.59% for PC and 28.49% for CPC.

From the discharge curves in Fig. 6, it can be further observed that there is an increase in the potential of the $Na^+$ extraction plateaus at 1C loads. Samples PC and DQC then have almost identical shape of discharge curves differing only in discharge capacity. In contrast, sample CPC is shifted towards higher potentials. For clarity, the discharge capacity in the first 0.1C cycle, the last 1C cycle and the last 0.1C cycle is plotted in the bar chart in Figure 7 bellow.

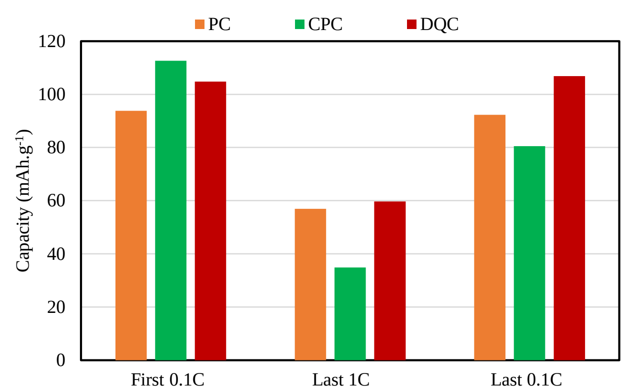



Fig. 7. Comparison of discharge capacities of individual samples

## V. CONCLUSION

Na-ion battery technology is emerging as a promising complement to Li-ion batteries and offers a cost-effective and sustainable solution for energy storage applications. Advances in the use of carbonaceous materials, in particular pitch coke, petroleum coke and dry quenched, as low-cost alternatives for Na-ion anodes present a significant opportunity to address the growing demand for affordable energy storage solutions. With ongoing research and development, these advancements hold the potential to reshape the future of renewable energy storage.

The most successful sample was DQC. Despite the most amount of impurities, not-so-good Raman spectrum and the absence of a sharp anodic peak in the cyclic voltammetry, this sample achieved the highest discharge capacity of 106.88 $mAh.g^{-1}$. Moreover, this discharge capacity increased by 1.97% after 50 cycles at different loads due to electrode formatting and increasing efficiency caused by SEI layer optimisation. If long-term cycling was performed, this discharge capacity is expected to decrease, but only in units of percent. The highest capacity at initial cycling was achieved by the CPC sample, but the decrease in this capacity was very steep.

In conclusion, it can be said that the investigated anode materials originating from the Czech Republic have low price and high availability compared to hard carbon or graphite. Their electrochemical properties are slightly worse than hard carbon, but this could be solved by a specific form of chemical cleaning or further calcination to remove unwanted impurities. These materials therefore appear to be a very suitable candidate for the future development of low-cost and sustainable anode materials for Na-ion batteries.

## Acknowledgment

This work was supported by the specific graduate research of the Brno University of Technology No. FEKT-S-23-8286. CzechNanoLab project LM2023051 funded by MEYS CR is gratefully acknowledged for the financial support of the measurements at CEITEC Nano Research Infrastructure.

## References

[1] D. Choi, N. Shamim, A. Crawford, Q. Huang, C. K. Vartanian, V. V. Viswanathan, M. D. Paiss, M. J. E. Alam, D. M. Reed, and V. L. Sprenkle, “Li-ion battery technology for grid application”, *Journal of Power Sources*, vol. 511, 2021.

[2] T. Perveen, M. Siddiq, N. Shahzad, R. Ihsan, A. Ahmad, and M. I. Shahzad, “Prospects in anode materials for sodium ion batteries - A review”, *Renewable and Sustainable Energy Reviews*, vol. 119, 2020, doi: 10.1016/j.rser.2019.109549.

[3] N. Tapia-Ruiz, A. R. Armstrong, H. Alptekin, M. A. Amores, H. Au, J. Barker *et al.*, “2021 roadmap for sodium-ion batteries”, *Journal of Physics: Energy*, vol. 3, no. 3, Jul. 2021.

[4] K. Kubota and S. Komaba, “Review—Practical Issues and Future Perspective for Na-Ion Batteries”, *Journal of The Electrochemical Society*, vol. 162, no. 14, Oct. 2015, doi: 10.1149/2.0151514jes.

[5] The editors of Britanica, “Coke”, *Encyclopedia Britannica*, 2008.

[6] A. Sarkar, D. Kocaefe, Y. Kocaefe, D. Sarkar, D. Bhattacharyay, B. Morais, and J. Chabot, “Coke–pitch interactions during anode preparation”, *Fuel*, vol. 117, pp. 598-607, 2014.

[7] L. Edwards, “The History and Future Challenges of Calcined Petroleum Coke Production and Use in Aluminum Smelting”, *JOM*, vol. 67, no. 2, pp. 308-321, 2015.

[8] S. Gangadharan, A. Bhushan, S. Kumar, R. Singh, K. M. Mohan, A. Shankar, and A. S. Reddy, “Coke dry quenching and implications on coke quality and blast furnace performance”, *Ironmaking & Steelmaking*, vol. 50, no. 1, pp. 75-83, Jan. 2023.

[9] K. Li, Q. Liu, H. Cheng, M. Hu, and S. Zhang, “Classification and carbon structural transformation from anthracite to natural coaly graphite by XRD, Raman spectroscopy, and HRTEM”, *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, vol. 249, 2021.

[10] X. Zhang, X. Dong, X. Qiu, Y. Cao, C. Wang, Y. Wang, and Y. Xia, “Extended low-voltage plateau capacity of hard carbon spheres anode for sodium ion batteries”, *Journal of Power Sources*, vol. 476, 2020.

[11] J. Meng, G. Jia, H. Yang, and M. Wang, “Recent advances for SEI of hard carbon anode in sodium-ion batteries: A mini review”, *Frontiers in Chemistry*, vol. 10, Sep. 2022.

# Understanding the Impact of Temperature on Li-ion Batteries

1st Martin Šedina
*Department of Electrotechnology*
*Faculty of electrical engineering and communication*
*Brno university of technology*
Brno, Czech Republic
xsedin00@vutbr.cz

2nd Tomáš Kazda
*Department of Electrotechnology*
*Faculty of electrical engineering and communication*
*Brno university of technology*
Brno, Czech Republic
kazda@vutbr.cz

***Abstract*—With the demand for reliable energy storage solutions continuing to grow, understanding the complex relationship between temperature and Li-ion batteries is crucial. This article focuses on the effect of temperature on key parameters of Li-ion batteries. It explores how temperature changes affect the lifetime, load capacity and overall performance of Li-ion batteries. Understanding these critical aspects should help optimise Li-ion batteries in different operating environments to ensure their longevity and efficiency. It also shows experiments with the testing of automotive cells at different temperatures and the effect of the DoD on the battery performance and life span at these temperatures.**

***Index Terms*—Li-ion batteries, Low temperatures, High temperatures**

## I. Introduction

Lithium-ion batteries have become the cornerstone of advanced energy storage, are everywhere in smartphones, wearables, laptops and powering a rapidly growing number of electric vehicles (EVs). The integration of Li-ion batteries into EVs has led to an increased focus on optimizing battery life, which is affected by factors such as current load and operating temperature, etc.

Operating temperature is a key topic of discussion due to its direct influence on the condition of Li-ion batteries. High temperatures can irreversibly reduce battery capacity and cause battery damage, demanding a balance between performance and temperature management. Conversely, low temperatures pose a risk, particularly in terms of electrolyte freezing, which can compromise battery functionality and lead to battery damage. [1]

Addressing these issues becomes more important in the global context of EV deployment, where vehicles are expected to operate smoothly in a variety of climatic conditions. Ongoing research and innovation are focused on finding solutions to ensure that Li-ion batteries are sufficiently resilient to operating conditions.

## II. Li-ion batteries

Lithium-ion batteries are the newest and most advanced battery technology on the market, which provide high energy density with high voltage of the cell and a good ratio between power and capacity. Li-ion batteries represent a wide range of used materials, especially for the cathode, see 1. However, only some types are mainly used on the market. Almost all batteries use graphite as an anode material. Graphite has a theoretical capacity of 372 $mAhg^{-1}$ and voltage vs. $Li/Li^{+}$ around 0.2 V. Comparison of all commercially used battery chemistry is shown in Figure 1. [2] [3]

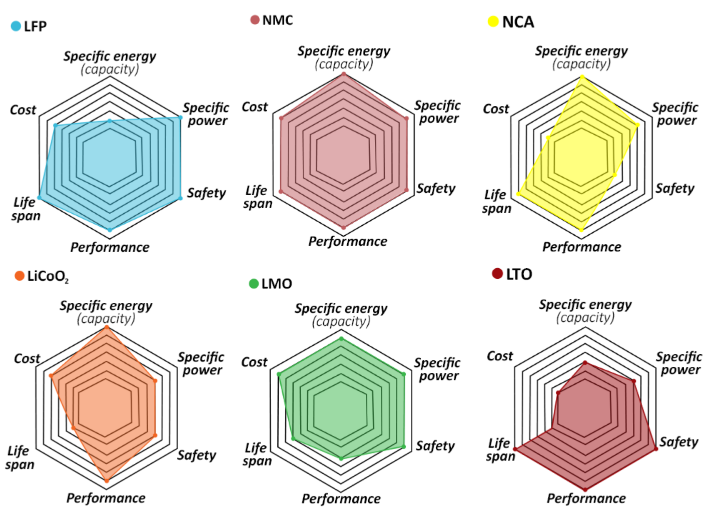



Fig. 1. Comparison of properties of individual battery chemistry [2]

### A. Current technologies used in EVs and storages

The LFP battery is primarily used in electric vehicles used in China and for energy storage. They have the advantage of being free of heavy metals, long life and high reliability. [2] [4] [5]

- **Potential vs. $Li/Li^{+}$**- 3.45 V
- **Teoretical capacity**- 170 $mAhg^{-1}$

The NMC battery is the most widely used battery type for electric vehicles. The current generation uses the NMC622

($LiNi_{0.6}Mn_{0.2}Co_{0.2}O_2$) configuration and in 2022 has transitioned to the NMC811 ($LiNi_{0.8}Mn_{0.1}Co_{0.1}O_2$) configuration. The battery has a high energy density, very high theoretical capacity and high stability. [2] [4] [5]

- **Potential vs. Li/Li$^+$**- 3.7 V
- **Teoretical capacity**- 280 mAhg$^{-1}$

The NCA battery is used in the NCA8155 ($LiNi_{0.8}Co_{0.15}Al_{0.05}O_2$) configuration and is currently used by Tesla in their vehicles. The battery has the highest energy density and very high theoretical capacity. [2] [4] [5]

- **Potential vs. Li/Li$^+$**- 3.7 V
- **Teoretical capacity**- 279 mAhg$^{-1}$

### B. *Other technologies*

The LCO ($LiCoO_2$) battery is the first developed commercial cathode material, in nowadays is mostly used in smartphones and wearables. The battery has a high energy density and with the additives, nowadays operate with a maximum voltage of 4.4+ V. [2] [6]

- **Potential vs. Li/Li$^+$**- 3.85 V
- **Teoretical capacity**- 274 mAhg$^{-1}$

LMO ($LiMn_2O_4$) is a material that was developed as a cheaper replacement for the first more powerful LCO batteries. These batteries were used by e.g. Citroen Zero. These batteries have a low environmental impact and good stability. [2] [4] [7]

- **Potential vs. Li/Li$^+$**- 4.1 V
- **Teoretical capacity**- 148 mAhg$^{-1}$

LTO ($Li_4Ti_5O_{12}$) is a special type of battery because, in the case of all of the above, its name is made from the cathode material used in the battery. LTO is on the other hand anode material. It is typically combined with NMC or LFP cathode. The cathode used then influences the voltage of the battery and its capacity. LTO has a theoretical capacity of 175 mAhg$^{-1}$ and voltage vs. Li/Li$^+$ 1.5 V. Due to high voltage vs Li has lower voltage than batteries with graphite, on the other hand, LTO has a much longer cycle life than graphite. [2] [8]

## III. Battery degradation

Lithium-ion battery degradation is a natural process that occurs over time, reducing the battery's ability to hold a charge. This process is a significant concern in many applications, particularly electric vehicles, where battery performance and lifespan are critical. Therefore, understanding and mitigating lithium-ion battery degradation is a key focus in battery research. All degradation processes are shown in Figure 2.

### A. *SEI layer growth*

The SEI layer, or Solid Electrolyte Interphase, is a protective coating that forms on the negative electrode of lithium-ion batteries during their initial cycle. Its growth is initiated by the breakdown of electrolytes, causing a loss of electrolyte material. As the battery ages, the SEI layer continues to expand primarily due to factors such as particle cracking and changes in electrolyte composition. This layer protects the electrode but also prevents the movement of lithium ions and restricts the flow of electrolytes. This increases the impedance of the cell, leading to a decrease in battery performance over time. Unfortunately, the formation of this SEI layer results in irreversible losses of the battery capacity. [9] [10] [11]

### B. *Lithium plating*

Lithium plating is an unwanted process where lithium metal is deposited on the surface of the negative electrode instead of being incorporated into it. This phenomenon can be caused by various factors such as high-speed charging, low and high operating temperatures, overcharge state, high current, exceeding the maximum cell voltage and insufficient electrode area. The consequences of lithium plating are significant and lead to irreversible damage such as electrode plating, SEI layer growth and dendrite growth. These dendrites can damage the separator and cause internal short circuits. [9] [10] [11]

### C. *Positive electrode structural change and decomposition*

The decomposition and degradation of a positive electrode are greatly influenced by its chemical composition. The degradation process is caused by phase transformation, oxidation, electrolyte degradation and surface passivation, and acid action. Changes in transition metal composition, high state of charge and high temperatures can cause accelerated degradation, impedance rise and sharp drop in cell capacitance. [9] [10] [11]

### D. *Particle fracture*

Particle fracture is a common problem in lithium-ion batteries, especially for materials with a high theoretical specific capacity. The significant volume changes that occur in electrode materials during cycling create stress and mechanical stress, leading to particle breakage and decreased cycling life. Various factors contribute to the likelihood of particle cracking, including operating temperature, current density, particle size, and manufacturing process. Consequences include the breakdown of electrical contacts, reduced capacity, growth of solid electrolyte interphase, as well as the formation of passivation layers on positive electrodes and potential electrode delamination. [9] [10] [11]

### E. *Electrolyte decomposition*

The electrolyte in lithium-ion batteries is crucial for ion transport, but it undergoes complex reactions that affect battery performance. Some ageing processes in lithium-ion batteries are mainly caused by electrolyte degradation. The complex composition of the electrolyte introduces problems with complex side reactions during ageing. Understanding electrolyte degradation is essential for optimising battery design and improving battery performance. The decomposition of electrolytes in Li-ion batteries can be caused by the formation of the SEI layer, high temperature, overcharging and over-discharging, chemical reactions inside the battery etc. [12] [13]

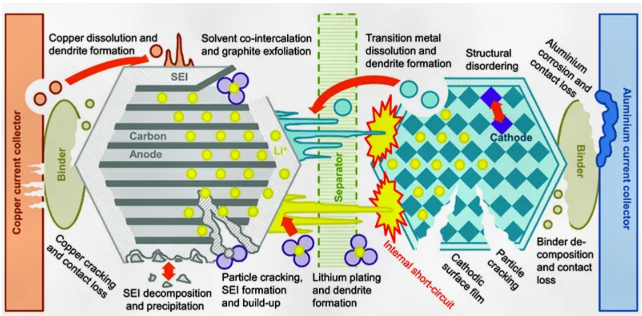



Fig. 2. Comparison of properties of individual battery chemistry [14]

## IV. Effect of the operation temperature on Li-ion batteries

The operating temperature of Li-ion batteries significantly affects their performance and safety, with an ideal range of 15 °C to 35 °C and a standard operating range for most batteries of -20 °C to 60 °C. Keeping within this temperature range is critical to avoid negative impacts on chemical reactions, ionic conductivity and overall battery performance. Operating Li-ion batteries below the lower limit in cold environments results in slow reactions and reduced energy efficiency. On the other hand, exceeding the upper limit can lead to performance degradation and, in extreme cases, thermal runaway. Accurate temperature control is essential to ensure Li-ion batteries operate within the specified range, improving both performance and safety. [15] [16]

### A. Low temperatures

The cycling performance of Li-ion batteries in low temperatures is notably impacted by internal resistances, particularly charge-transfer resistance (Rct). Rct reflects the slow kinetics of Faradic reactions at electrode–electrolyte interfaces. The resistance is influenced by the state of charge (SOC), with discharged electrodes exhibiting higher Rct. This disparity in Rct levels between charging and discharging processes contributes to the greater difficulty in charging at low temperatures. Other things could be freezing of the electrolyte and stopping the reaction in the battery or lithium plating which reduce battery capacity. [17] [18]

### B. High temperatures

The effect of high temperatures on Li-ion batteries is complex and multifaceted, with heat generation being a critical factor in operation. Heat is generated due to reversible and irreversible processes, including entropy heat from reversible electrochemical reactions and various irreversible processes such as active polarisation, Ohmic heating, stirring and enthalpy changing. The temperature rise inside the Li-ion caused by these processes contributes to cyclic ageing and calendar ageing, which eventually affects the performance and lifetime of the battery. Studies of individual components such as electrodes and electrolytes reveal thermally induced crystallographic transformations and decomposition. At the system level, ageing of the entire Li-ion occurs, with binder migration and changes at the SEI leading to capacity loss. In addition, thermal runaway can be a critical issue at high temperatures, leading to uncontrolled heat generation, structural damage and potential explosion. Understanding and controlling high-temperature effects is essential to optimise Li-ion performance, prevent ageing due to thermal effects and ensure safety during operation. [19] [20]

## V. Previous experiment

In the previous experiment, we tested 18650 cells with LTO chemistry in extreme temperatures up to below -30 °C. This battery performed very well at low temperatures, during cycling at -30 °C capacity of the battery dropped around 64 % and stayed there during low-temperature cycling at 0.2C. The capacity drift is possibly made by self-heating during current flow during the battery cycling. After heating up to room temperature and continuing in cycling battery did not show any irreversible capacity drop. The result from the cycling could be seen in 3

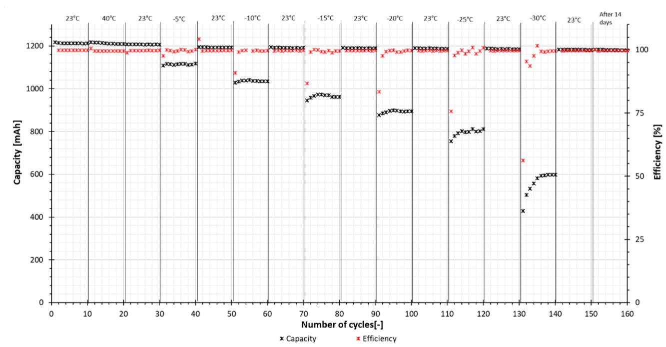



Fig. 3. Capacity fade and coulombic efficiency during low-temperature cycling of LTO cell [21]

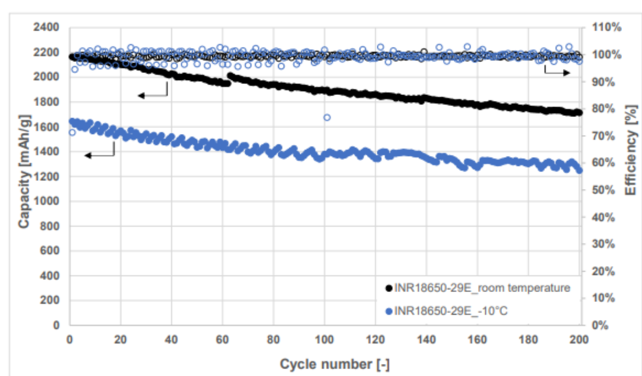



Fig. 4. Capacity fade and coulombic efficiency during low-temperature cycling NMC cells [22]

In another experiment, my colleague tested 18650 NMC cells. Batteries were cycled at -10 °C and at room temperature to the comparison of the battery working at low temperatures and in ideal operating conditions. After 200 cycles battery cycled at low temperature showed a little bit higher degradation (1.1 %) than the battery cycled at room temperature, a measurement of the degradation was made at room temperature, for the result see Figure 4. On the other hand, the battery was able to work outside of the operation windows defined by the manufacturer. [22]

## VI. Experiment

As was mentioned at the beginning, understanding the high and low thermal effects on the batteries is key for the prediction of operation large battery applications. For this reason, we set the experiment where we tested large automotive pouch cells with the capacity of 78 Ah with NMC721 chemistry to test how combination of 70 % depth of discharge and different operating temperatures affect battery life spam. Results should help to set operation conditions and prediction of operation in second-life energy storage.

For the testing were set testing temperatures 5, 22 and 45 °C and 70 % depth of discharge with the currents 0.33C-0.33C and charging protocol constant current-constant voltage with the cut-off current 0.05C . Befor the start of the cycling and after each 100 cycles is made load test with 0.1C-0.1C, 0.2C-0.2C and 0.2C-0.5C and electrochemical impedance spectroscopy at room temperature to measure battery parameters without temperature influence.

### A. Results

All batteries achieved 500 cycles and from that point, it is clearly seen an effect of temperature on battery capacity drop. On Figure 5 are shown batteries cycled at 70 % DoD and Table I showing capacity decrease after cycling.

At 70 % DoD is temperature effect not that significant, due to not fully utilization of the battery. At 45 °C there is no significant influence of the high temperature and capacity loss has mostly linear trend of capacity lost, on the other hand the capacity loss is double then than at room temeprature. At 5 °C batteries suffer more for reversible capacity loss due to not having enough energy to effectively heat itself and long-term working at these conditions can lead to faster capacity fade. For the precise setting of the real cyclable capacity of the batteries at 5 °C the real capacity lost is counted from cycle 501, the first cold cycle after reheating, to reduce the effect of low temperature on battery capacity measurements and to have more accurate degradation data. The highest degradation shown battery at 45 °C.

TABLE I
Comparison of the capacity fist vs 500 cycle with the capacity drop

| No001 (5°C) | No002 (5°C) | No003 (22°C) | No004 (22°C) | No005 (45°C) | No006 (45°C) |
|---|---|---|---|---|---|
| 0.802 | 0.845 | 0.971 | 0.969 | 1.005 | 1.006 |
| 0.783 | 0.783 | 0.926 | 0.921 | 0.904 | 0.901 |
| 7.19 % | 7.38 % | 4.64 % | 5.00 % | 10.08 % | 10.41 % |

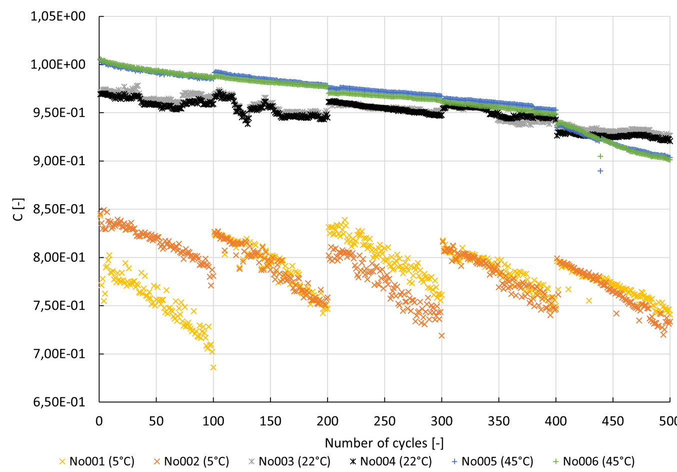



Fig. 5. Development of the capacity for 70 % DoD

## VII. Conclusion

The temperature effect on lithium-ion batteries is a significant factor affecting their performance and lifetime. High temperatures accelerate degradation, while low temperatures reduce performance. Advances in thermal management systems offer significant potential to reduce these problems. As can be seen from the measurements, the temperature effect can be reduced by setting a reduced DoD.

## Acknowledgment

This work was supported by the specific graduate research of the Brno University of Technology No. FEKT-S-23-8286. This work was supported by the EIT Urban Mobility project SEVES.

## References

[1] G. Leoncini, R. Mothier, B. Michel, and M. Clausse, “A review on challenges concerning thermal management system design for medium duty electric vehicles”, Applied Thermal Engineering, vol. 236, 2024.
[2] R. M. Salgado, F. Danzi, J. E. Oliveira, A. El-Azab, P. P. Camanho, and M. H. Braga, “The Latest Trends in Electric Vehicles Batteries”, Molecules, vol. 26, no. 11, 2021.
[3] J. Asenbauer, T. Eisenmann, M. Kuenzel, A. Kazzazi, Z. Chen, and D. Bresser, “The success story of graphite as a lithium-ion anode material – fundamentals, remaining challenges, and recent developments including silicon (oxide) composites”, Sustainable Energy  Fuels, vol. 4, no. 11, pp. 5387-5416, Oct. 2020.
[4] N. Nitta, F. Wu, J. T. Lee, and G. Yushin, “Li-ion battery materials: present and future”, *Materials Today*, vol. 18, no. 5, pp. 252-264, 2015.
[5] B. E. Murdock, K. E. Toghill, and N. Tapia-Ruiz, “A Perspective on the Sustainability of Cathode Materials used in Lithium-Ion Batteries”, Advanced Energy Materials, vol. 11, no. 39, 2021.
[6] K. Wang, J. Wan, Y. Xiang, J. Zhu, Q. Leng, M. Wang, L. Xu, and Y. Yang, “Recent advances and historical developments of high voltage lithium cobalt oxide materials for rechargeable Li-ion batteries”, Journal of Power Sources, vol. 460, 2020.
[7] J. Ma, T. Liu, J. Ma, C. Zhang, and J. Yang, “Progress, Challenge, and Prospect of LiMnO 2: An Adventure toward High-Energy and Low-Cost Li-Ion Batteries”, Advanced Science, vol. 11, no. 2, 2024.
[8] P. U. Nzereogu, A. D. Omah, F. I. Ezema, E. I. Iwuoha, and A. C. Nwanya, “Anode materials for lithium-ion batteries: A review”, Applied Surface Science Advances, vol. 9, 2022.

[9] J. S. Edge, S. O'Kane, R. Prosser, N. D. Kirkaldy, A. N. Patel, A. Hales, A. Ghosh, W. Ai, J. Chen, J. Yang, S. Li, M. -C. Pang, L. Bravo Diaz, A. Tomaszewska, M. W. Marzook, K. N. Radhakrishnan, H. Wang, Y. Patel, B. Wu, and G. J. Offer, "Lithium ion battery degradation: what you need to know", Physical Chemistry Chemical Physics, vol. 23, no. 14, pp. 8200-8221, Apr. 2021.

[10] S. E. J. O'Kane, W. Ai, G. Madabattula, D. Alonso-Alvarez, R. Timms, V. Sulzer, J. S. Edge, B. Wu, G. J. Offer, and M. Marinescu, "Lithium-ion battery degradation: how to model it", Physical Chemistry Chemical Physics, vol. 24, no. 13, pp. 7909-7922, Mar. 2022.

[11] M. M. Kabir and D. E. Demirocak, "Degradation mechanisms in Li-ion batteries: a state-of-the-art review", International Journal of Energy Research, vol. 41, no. 14, pp. 1963-1986, 2017.

[12] Y. Liao, H. Zhang, Y. Peng, Y. Hu, J. Liang, Z. Gong, Y. Wei, and Y. Yang, "Electrolyte Degradation During Aging Process of Lithium-Ion Batteries: Mechanisms, Characterization, and Quantitative Analysis", Advanced Energy Materials.

[13] C. L. Campion, W. Li, and B. L. Lucht, "Thermal Decomposition of LiPF[sub 6]-Based Electrolytes for Lithium-Ion Batteries", Journal of The Electrochemical Society, vol. 152, no. 12, 2005.

[14] "SPECIFIC EXPLANATION OF THE LITHIUM ION BATTERY DEGRADATION", TYCORUN, 02023.

[15] S. Ma, M. Jiang, P. Tao, C. Song, J. Wu, J. Wang, T. Deng, and W. Shang, "Temperature effect and thermal impact in lithium-ion batteries: A review", Progress in Natural Science: Materials International, vol. 28, no. 6, pp. 653-666, 2018.

[16] T. M. Bandhauer, S. Garimella, and T. F. Fuller, "A Critical Review of Thermal Issues in Lithium-Ion Batteries", Journal of The Electrochemical Society, vol. 158, no. 3, 2011.

[17] H. Luo, Y. Wang, Y. -H. Feng, X. -Y. Fan, X. Han, and P. -F. Wang, "Lithium-Ion Batteries under Low-Temperature Environment: Challenges and Prospects", Materials, vol. 15, no. 22, 2022.

[18] Y. Ji, Y. Zhang, and C. -Y. Wang, "Li-Ion Cell Operation at Low Temperatures", Journal of The Electrochemical Society, vol. 160, no. 4, pp. A636-A649, Jan. 2013.

[19] S. Ma, M. Jiang, P. Tao, C. Song, J. Wu, J. Wang, T. Deng, and W. Shang, "Temperature effect and thermal impact in lithium-ion batteries: A review", Progress in Natural Science: Materials International, vol. 28, no. 6, pp. 653-666, 2018.

[20] D. R. Wright, N. Garcia-Araez, and J. R. Owen, "Review on high temperature secondary Li-ion batteries", Energy Procedia, vol. 151, pp. 174-181, 2018.

[21] The 5 th International Conference on Nanomaterials: Fundamentals and Applications. Košice: Pavol Jozef Šafárik University in Košice Publishing ŠafárikPress, 2021.

[22] Advanced Batteries Accumulators and Fuel Cells – 22 nd ABAF. Brno: Brno University of Technology Faculty of Electrical Engineering and Communication Department of Electrical and Electronic Technology, 2021.

# Lithium-ion battery SOH analysis


Marek Sedlařík
*Department of Electrical and Electronic Technology*
*Faculty of Electrical Engineering and Communication, Brno University of Technology*
Brno, Czech Republic
203338@vut.cz

Tomáš Kazda
*Department of Electrical and Electronic Technology*
*Faculty of Electrical Engineering and Communication, Brno University of Technology*
Brno, Czech Republic
kazda@vut.cz

Petr Vyroubal
*Department of Electrical and Electronic Technology*
*Faculty of Electrical Engineering and Communication, Brno University of Technology*
Brno, Czech Republic
vyroubal@vut.cz



***Abstract*— Battery State-of-Health modelling can significantly reduce the amount of costly laboratory tests in the application being analyzed. This paper discusses the prediction of the State-of-Health, an indicator of battery life, using support vector regression. This experiment is performed on a sample cell of a lithium-ion battery, which is subjected to a method known as the Constant Current Constant Voltage method, where the battery is charged and discharged at a constant current of 0.5 C. Although this method in this paper is applied in laboratory conditions and it is a controlled method, or it deviates from the battery cycling of real applications, it can be used in these applications, thus the scope of this research can predict the State-of-Health also in the areas of batteries used in mobile devices or electromobility. The State-of-Health indicator then determines whether the battery is still suitable for that primary application. Assuming that we can predict this parameter with some accuracy, it is then also possible to tell after what length of time a battery will need to be replaced and when it will be suitable for secondary applications such as stationary storage. Once it reaches that state, this calculation can be further applied to those applications as well.**



**Keywords — Li-ion battery, SOH, SVR, Estimation, CCCV**


## I. Introduction

In recent decades, lithium-ion (Li-ion) batteries have evolved into a key technology for energy storage in a variety of applications from consumer electronics to electric vehicles and stationary power systems. Due to their high energy density, long lifetime, and ability to recharge without significant reduction in capacity, Li-ion batteries have become the preferred solution for many primary and secondary applications.

Primary applications are notably crucial for devices including electric vehicles and portable electronics, providing essential power for their core functionalities. On the other hand, in secondary applications, such as stationary energy storage, Li-ion batteries serve a supportive role, either as backup power solutions or as key elements in enhancing the reliability and stability of the electrical grid.

The life of Li-ion batteries is affected by many factors, including the temperature of the operating environment, the depth of discharge, the magnitude of the charge and discharge current and the overall number of battery cycles. Together, these factors determine how long a battery maintains its capacity and therefore how efficiently it stores electrical energy.

The State-of-Health (SOH) serves as a pivotal metric for gauging the aging process of batteries, offering insight into the extent of capacity or power reduction over time. A fresh battery boasts an SOH of 100%, indicating optimal performance and capacity. As the SOH wanes to 80%, it signals that the battery no longer possesses the requisite energy to fulfil the demands of specific e-mobility applications, marking its transition to the end-of-life (EOL). The intricacy of SOH lies in its non-measurable nature, necessitating the development of estimation techniques to ascertain this critical parameter accurately. Ensuring precise estimation of SOH is paramount for the maintenance of safety and reliability in e-mobility operations, highlighting the significance of ongoing research in this area. [1]

SOH estimation techniques are categorized into physics-based and data-driven methods. The physics-based category includes empirical, equivalent circuit, and electrochemical models, with the crucial aspect being the updating of model parameters. On the other hand, data-driven methods rely less on physical understanding and more on operational data to capture the degradation characteristics of SOH. Data-driven strategies aim to establish a relationship between SOH and various features. These features comprise both direct attributes such as voltage, current, and temperature, and indirect attributes, which typically consist of statistical analyses of direct features. In the field of estimating the SOH a variety of data-driven methodologies have been employed. Among these tools, recurrent neural network (RNN), support vector regression (SVR), relevance vector machine (RVM) and Gaussian process regression (GPR) are most significant. Specifically, the SVR approach is used in this study due to its robust performance in predictive accuracy and computational efficiency. These methods have been documented extensively for RNN [2], SVR [3,4], RVM [5], and GPR [6].

## II. Experiment

### *A. Measurement*

The estimation of SOH is investigated for a Samsung INR18650-35E battery with lithium-nickel-manganese-cobalt-oxides (NMC) as a cathode material and graphite as an anode material, 300 cycles were measured using the constant current constant voltage (CCCV) method. The battery was charged and discharged at 0.5 C which corresponds to a current of 1.7 A at a battery capacity of 3400 mAh and in a voltage range of 2.65 – 4.20 V. Voltage profiles during charging and discharging are shown in Fig. 1 and 2, respectively, where the dark blue color curve is the first cycle and the dark red color curve is the last cycle. Detailed information for the battery cell used in this paper is listed in Table I.

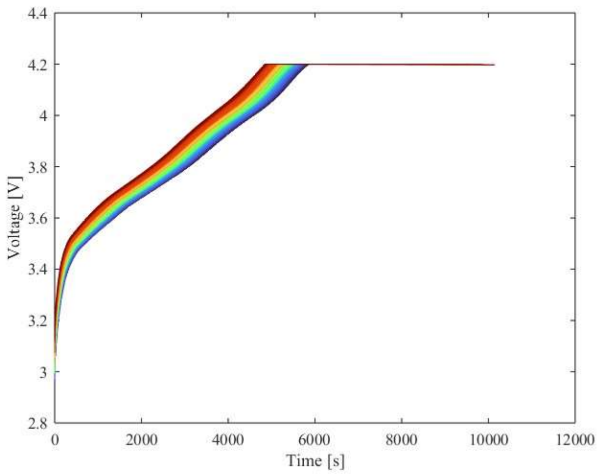



Fig. 1: Battery voltage dependence on time during charging (the dark blue curve is the first cycle and the dark red is the last cycle).

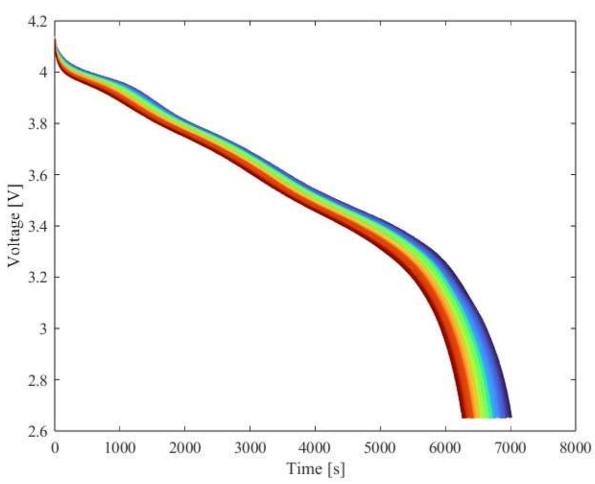



Fig. 2: Battery voltage curves during discharging (the dark blue curve is the first cycle and the dark red is the last cycle).

The CCCV approach is widely recognized as the default protocol for cycling Li-ion batteries. This method unfolds in two key stages: initially, the battery is charged at a consistent current rate until it hits a predetermined voltage threshold. Following this, the charging mode switches to a constant voltage, maintaining this voltage level while gradually reducing the current. The switch from constant current to constant voltage occurs once the cell's voltage reaches its maximum value, a level often set based on the specific application's demands or the battery's nominal capacity. It is during the constant voltage phase that the battery's SOH can be particularly impacted, as slower charging rates may enhance longevity and reduce stress on the battery.

The CCCV method is sensitive to the charge/discharge current rate (C-rate), temperature, and the battery's intrinsic chemistry, with higher C-rates potentially limiting lithium diffusion within the electrodes and thus, affecting the battery's capacity and lifespan. This limitation underscores the importance of optimizing the C-rate for rapid charging while also balancing the trade-off between charge speed and battery health. [7]

TABLE I: LI-ION BATTERY PARAMETERS

| **Samsung INR18650-35E** | |
|---|---|
| ***Parameter*** | ***Value*** |
| Type | Cylindric |
| Nominal Capacity | 3.4 Ah |
| Nominal Voltage | 3.6 V |
| Maximum Voltage | 4.2 V |
| Minimum Voltage | 2.5 V |
| Maximum Current | 8 A |

## B. Data selection

Data separation is performed in about 1/3 of the measured cycles, of which 95 are selected for training and the rest for validation. The data has been separated into individual charge and discharge cycles where a parameter is extracted from each cycle, the set of which forms the input data for the support vector regression.

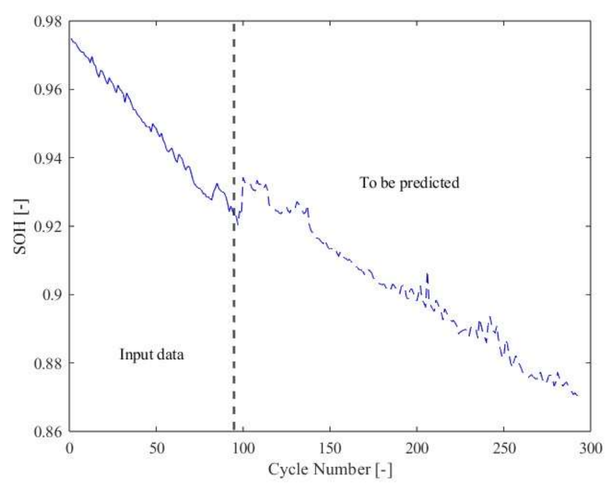



Fig. 3: Measured and separated SOH to Input data and Validation data

## C. Feature extraction

In this study, we evaluate two parameters that can be measured in real time: Time Interval of an Equal Charging Voltage Difference (TIECVD) and Time Interval of an Equal Discharging Voltage Difference (TIEDVD). These indicators tend to degrade as the battery ages, which implies that they could provide insights into the battery's capacity. TIECVD is calculated as the duration it takes for the battery voltage to increase from 3.5 V to 4.2 V during the charging process. Conversely, TIEDVD measures the time it takes for the voltage to decrease from 3.8 V to 3.6 V during discharge. It is important to mention that only assuming the battery is discharging at a constant current, so TIEDVD has a major impact. Otherwise, TIECVD becomes the sole indicator of battery health used for SOH estimation. [8]

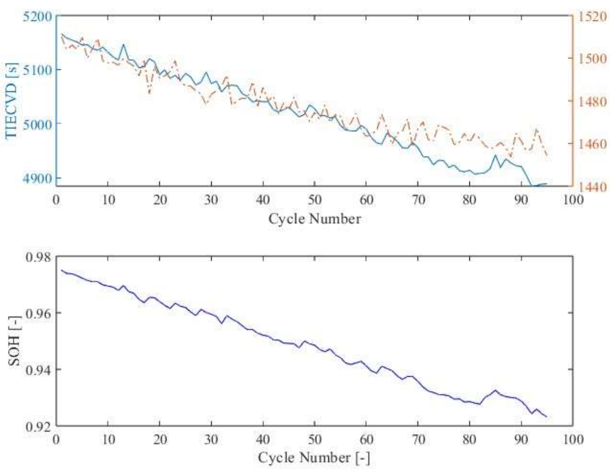



Fig. 4: Input features (TIECVD, TIEDVD) compared with SOH.

## D. SOH estimation

SOH represents the ratio between actual capacity $Q_i$ and nominal capacity $Q_0$ of the battery.

$$SOH_i = \frac{Q_i}{Q_0} \tag{1}$$

Support Vector Regression (SVR) represents an adaptation of Support Vector Machines (SVM) for regression tasks, allowing the methodology to address issues with small sample sizes and non-linear data. SVR constructs hyperplanes in a high-dimensional space, which serves as a decision surface to predict output values with high accuracy. When applied to the evaluation of a battery's SOH, feature data is non-linearly mapped and subsequently regressed within a multi-dimensional space to estimate the SOH value.

There is a sample set $S = \{(x_i, y_i), i = 1,2, \dots, N\}(x_i \in R^n, y_i \in R)$, where $x_i$ and $y_i$ are the $i$-th input and output feature values. $N$ denotes the sample size, and $n$ stands for the feature dimensionality. The mapping function employed by SVR transforms the original low-dimensional data into a higher-dimensional space, following the equation:

$$f(x) = \omega \cdot \varphi(x) + b, \tag{2}$$

where $\omega$ denotes the weight vector, $b$ signifies the bias, and $\varphi(x)$ is the nonlinear mapping function. If the relaxation variables $\zeta_t$ and $\zeta_i$ are inserted into the equation, the solution to the problem of finding $\omega$ and $b$ can be explained mathematically as below:

$$\min R(\omega, b, \bar{\zeta}) = \frac{1}{2} \parallel \omega \parallel^2 + C \sum_{i}^{n} (\bar{\zeta}_t + \bar{\zeta}_t^*), \tag{3}$$

The deviation of the predicted values from the actual values is within the specified error epsilon $\epsilon$ for each data point as specified in the equations:

$$\text{s.t} \begin{Bmatrix} y_i - \omega \cdot \phi(x) - b \le \varepsilon + \bar{\zeta}_i \\ \omega \cdot \phi(x) + b - y_i \le \varepsilon + \bar{\zeta}_i^* \\ \zeta_r \bar{\zeta}_t^* \ge 0. \end{Bmatrix}, \tag{4}$$

Where $C$ denotes the penalty term reflecting the trade-off between model complexity and the degree to which deviations larger than $\epsilon$ are tolerated. By introducing Lagrange multipliers and using the kernel function, the SVR function can be redefined as:

$$f(x) = \sum_{i=1}^{N} (\alpha_i - \alpha_i^*) K(x_i, x_j) + b, \tag{5}$$

In this equation $K(x_{\mathrm{i}}, x_j)$ is the kernel function, where $x_i$ are the input vectors for training samples, and $x_j$ are input vectors for test samples. Lagrange operators are represented by $\alpha_i$ and $\alpha_i^*$.

Various kernel functions are usually used in the SVR model: separate linear kernel function, polynomial kernel function, RBF kernel function and sigmoid kernel function. Gaussian or Radial Basis Functions (RBF) are chosen to solve the SOH estimation.

$$K_{\mathrm{RBF}}(x_i, x_j) = \exp\left(-\frac{1}{2\sigma^2} \|x_i - x_j^2\|\right), \tag{6}$$

where the width of the RBF kernel function is represented by $\sigma$. [4]

# III. RESULTS

To achieve the best results, it is chosen automatic optimization of hyperparameters: Kernel function width, penalty factor and error tolerance. The results of the optimization are shown in Table II and their progress is plotted in Fig. 5.

The 'Min observed objective' represents the lowest value of the objective function achieved during the hyperparameter tuning process, serving as a direct measure of model performance under various parameter settings. The 'Estimated min objective' reflects the predictive model's extrapolation of the minimum objective value based on current trends observed throughout the optimization iterations. The sharp peak observed initially suggests a suboptimal set of hyperparameters, which the optimization algorithm rapidly adjusts in subsequent evaluations.

TABLE II: HYPERPARAMETER OPTIMIZATION RESULTS

| ***Parameter*** | ***Value*** |
|---|---|
| Width of kernel function $\sigma$ | 129.54 |
| Penalty factor $C$ | 225.03 |
| Tolerated error $\epsilon$ | $1.28 \cdot 10^{-3}$ |

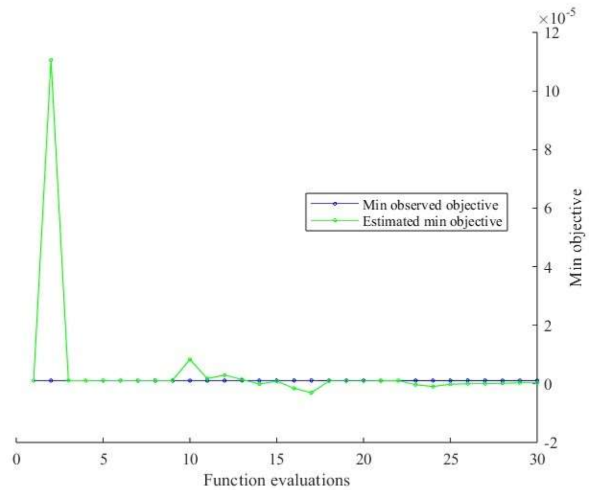



Fig. 5: Min objective vs. number of function evaluations.

The result of SOH prediction using the SVR method with hyperparameter optimization is plotted in Fig. 6. It is clear that the predicted curve maintains the same trend as the measured curve, although there is a certain deviation in the predicted data from the 100th cycle when continuous cycling is interrupted and other battery tests are performed. At this point, the battery rests before it is reconnected to continuous cycling, which is reflected by an increase in SOH.

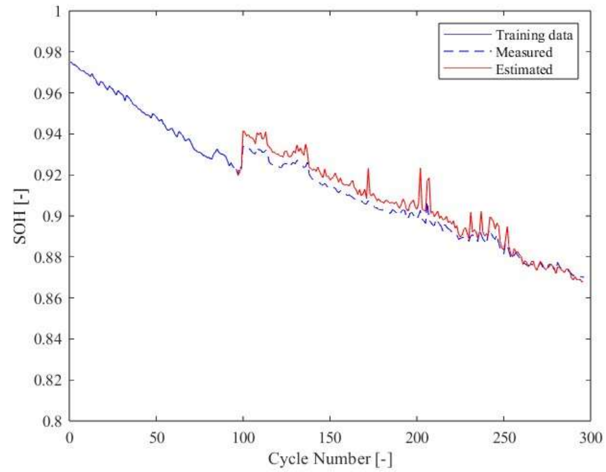



Fig. 6: Comparison between predicted and measured SOH.

### *A. Evaluation*

The root mean square error (RMSE) graph in Fig. 7 shows the deviation of the predicted SOH and measured values as a function of each cycle.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}$$

Where $n$ is cycle number, $y_i$ represents measured data and $\hat{y}_i$ is the estimated SOH value. The initial values are close to zero, where it is directly linked to the training data by the prediction. However, this is followed by a battery rest at cycle 100, where the RMSE value increased rapidly. This rest is not included in the training dataset as it would significantly change the gradient of the predicted curve. The predicted data tends to decrease the RMSE value in the later stage, which is precisely the most important part of the prediction, and the resulting trend of decrease in predicted SOH is thus almost equivalent to the measured one. The overall RMSE of the entire predicted dataset compared to the measured one is equal to $5.3 \cdot 10^{-3}$.

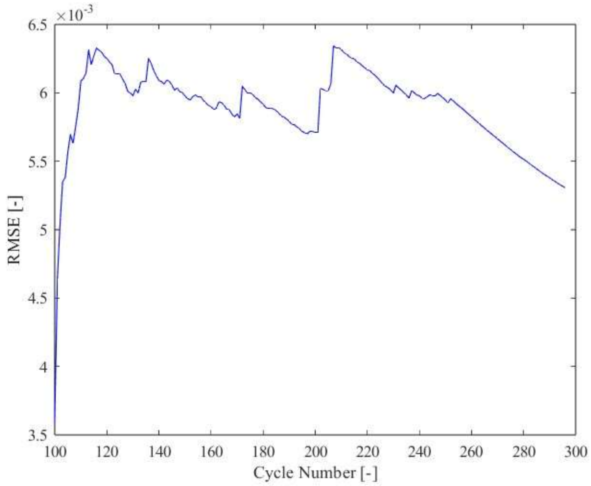



Fig. 7: Evaluation of RMSE between estimated SOH and validation data in each cycle

## IV. Conclusion

This paper focuses on the SOH estimation of Li-ion batteries using the SVR method with hyperparameter optimization. A deviation of real from measured data was observed, which may be caused by the battery resting at 100 cycles when additional tests were performed on the battery and it stopped cycling continuously for a period of time. However, this rest period was not included in the test data in response to it deviating from the resulting predicted curve vector. Although this phenomenon is considered to be natural battery behaviour, it occurred only once in the dataset. On the other hand, this situation could be resolved by making it more frequent throughout the dataset, thus further optimizing the resulting model. Nevertheless, despite this fact, the resulting RMS error of the predicted SOH parameter with respect to the measured one is equal to $5.3 \cdot 10^{-3}$, it is clear that the SOH parameter can be effectively predicted using this method. The accuracy could be further improved by using more sophisticated data-driven methods in deep learning.


## Acknowledgment

This work was supported by the BUT specific research program (project No. FEKT-S-23-8286).


## References


[1] C. Lin, A. Tang, W. Wang, and J. ‐ H. Chou, “A Review of SOH Estimation Methods in Lithium-ion Batteries for Electric Vehicle Applications”, *Energy Procedia*, vol. 75, no. 7, pp. 1920-1925, 2015.

[2] S. Ansari, A. Ayob, M. S. H. Lipu, A. Hussain and M. H. M. Saad, "Data-driven remaining useful life prediction for lithium-ion batteries using multi-charging profile framework: A recurrent neural network approach", *Sustainability*, vol. 13, no. 23, 2021.

[3] F. K. Wang, Z. E. Amogne, C. Tseng and J. H. Chou, “A hybrid method for online cycle life prediction of lithium ‐ ion batteries”, *International Journal of Energy Research*, vol. 46, no. 7, pp. 9080-9096, Jun. 2022.

[4] L. Xing, X. Liu, W. Luo, and L. Wu, “State of Health Estimation for Lithium-Ion Batteries Using IAO–SVR”, *World Electric Vehicle Journal*, vol. 14, no. 5, 2023.

[5] R. Wang, H. Feng, C. Tseng, and J. ‐ H. Chou, “Remaining useful life prediction of lithium ‐ ion battery using a novel health indicator”, *Quality and Reliability Engineering International*, vol. 37, no. 3, pp. 1232-1243, 2021.

[6] H. Pan, C. Chen, M. Gu, and J. ‐ H. Chou, “A Method for Predicting the Remaining Useful Life of Lithium Batteries Considering Capacity Regeneration and Random Fluctuations”, *Energies*, vol. 15, no. 7, pp. 9080-9096, 2022.

[7] M. Usman Tahir, A. Sangwongwanich, D. I. Stroe, and F. Blaabjerg, “Overview of multi-stage charging strategies for Li-ion batteries”, *Journal of Energy Chemistry*, vol. 84, pp. 228-241, 2023.

[8] Q. Zhao, X. Qin, H. Zhao, and W. Feng, “A novel prediction method based on the support vector regression for the remaining useful life of lithium-ion batteries”, *Microelectronics Reliability*, vol. 85, pp. 99-108, 2018.

# Recovery Of Lithium From Waste Water By Sodium Carbonate

Jiří Báňa
*Department Of Electrical and Electronic Technology*
*Brno University of Technology*
Brno, Czech Republic
0000-0002-8119-979X

Tomáš Kazda
*Department Of electrical and Electronic Technology*
*Brno University of Technology*
Brno, Czech Republic
0000-0003-1973-02Z

Antonín Šimek
*Department Of electrical and Electronic Technology*
*Brno University of Technology*
Brno, Czech Republic
0000-0002-1991-1029

Ondřej Čech
*Department Of electrical and Electronic Technology*
*Brno University of Technology*
Brno, Czech Republic
0000-0002-7878-5819

**Lithium-ion batteries are growing in production, thanks in large part to the rise of electromobility. In the future, every battery for electromobility will have to be recycled at the end of its life with a recycling efficiency of over 70 %. In the case of hydrometallurgical recycling methods and underwater crushing, some lithium is remaining in waste water. Efficient recovery of lithium from wastewater can mitigate the need for lithium mining. In this paper, we focus on the recovery of lithium carbonate from wastewater by thermal precipitation and crystallization of waste lithium using sodium carbonate.**

***Keywords— Li-ion battery, direct recycling, Lithium***

## I. INTRODUCTION

Generally, lithium-ion batteries (LIBs) comprise five primary elements: the cathode, anode, electrolyte, separator, and casing. The cathode includes an aluminum collector and an active material, typically containing transition metal oxides and lithium. The anode is typically composed of graphite coated onto a copper current collector, although there are instances where lithium titanate ($Li_4Ti_5O_{12}$ or LTO) is used instead of graphite or silicon. The casing is often constructed from aluminum or stainless steel, while the separator is typically made of a polymeric material. [1] [2]

The growing need for power sources to fuel mobile electronic devices and electric vehicles is driving a rapid increase in lithium consumption. Ensuring a cost-effective and dependable lithium supply has become paramount. To meet this demand, various potential lithium sources, including seawater, geothermal water, and wastewater, are being explored alongside conventional brine lakes and ores. This diversified approach aims to secure a stable and resilient lithium supply in the face of growing global demand. The recycling of lithium from discharged batteries is often overlooked. [3]

The hydrometallurgical process has become an important commercial technique for the recycling of metals from spent LIBs. In this approach, the batteries are crushed and dissolved in concentrated acids - primarily the cathode material. The leaching efficiency of Li, Mn, Co, and Ni from diverse cathode materials of discarded lithium-ion batteries (LIBs) is contingent on both the process temperature and duration. As the temperature rises, efficiencies typically increase, attributed to the endothermic dissolution nature of cathode materials. The impact of time on leaching efficiency is subject to variables like leaching agent type, temperature, and cathode material characteristics. Generally, prolonged leaching durations enhance efficiency, reaching a threshold beyond which additional time does not yield significant improvements. [4]

Leaching facilitating the subsequent recovery of dissolved metals by extraction, precipitation or electrodeposition. However, that in this process lithium ions are usually discharged as waste, resulting in a significant loss of valuable lithium resources. [5]

One potential disadvantage of hydrometallurgy is the generation of large amounts of wastewater containing chemical reagents and impurities, which require treatment and disposal. Additionally, the recovery of certain metals, such as lithium, can be challenging due to their high solubility and reactivity. Overall, however, hydrometallurgy offers a promising approach to the sustainable recycling of spent LIBs. [4] [5]

Research into lithium recovery is in its early stages, with proposed methods including sorption, nanofiltration, and electrolysis. Additionally, the precipitation of lithium with $CO_3^{2-}$ allows for recovery from a relatively clean aqueous solution, but due to the high solubility product constant of lithium carbonate (Ksb = $1.7 \times 10^{-3}$ at 25 °C), complete precipitation of lithium ions is not achievable. Adsorption emerges as a cost-effective and environmentally friendly means of lithium recovery from aqueous solutions. However, most adsorbents lack specificity, exhibiting low selectivity for a particular metal. Hence, the exploration of new adsorbents for the selective separation of lithium from aqueous solutions becomes imperative. [6]

Tsuruta et al. [7] suggests the potential for biological recovery of $Li^+$ ions using various microorganisms. Author discusses the findings of a study on lithium accumulation by various microorganisms, emphasizing the notable ability of certain gram-positive bacterial strains, specifically A. nicotianae IAM12342 and Brevibacterium helovolum IAM1637, to accumulate high levels of lithium. Among the tested microorganisms, A. nicotianae demonstrated the strongest performance, accumulating 126 μmol of Li/g of dry microbial cells within 1 hour.

Luo et al. [6] developed a magnetic lithium ion-imprinted polymer ($Fe_3O_4@SiO_2@IIP$) using a novel crown ether. The synthesized $Fe_3O_4@SiO_2@IIP$ exhibited optimal adsorption at pH 6, rapid kinetics (10 minutes for complete equilibrium), and adherence to an external mass transfer model, revealing a maximum adsorption capacity of 0.586 mmol/g. Impressive selectivity for Li(I) was demonstrated. The sorbent maintained high efficiency (>92%) over five cycles. A significant 89.8% lithium recovery during bed regeneration with 0.5 mol/L HCl solution was showcased. Notably, $Fe_3O_4@SiO_2@IIP$ displayed exceptional performance in real wastewater.

Lemaire et al. [8] aimed to develop a sorption/desorption method for recovering lithium from aqueous solutions, particularly leachates from used Li-ion batteries during hydrometallurgical recycling. Preliminary investigations were conducted on various solid materials, including activated carbons. This method proved inefficient for lithium uptake. The study revealed ion exchange as the mechanism for both sorption and desorption, with fast kinetics and a well-fitting model to experimental data. Concentration effects were observed, and although the lithium concentration after desorption might not be sufficient for traditional lithium salt precipitation, opportunities for reuse, such as $Li_3PO_4$ precipitation, were highlighted.

Precipitation is still the most used method for separating lithium from leached solutions. This method uses the difference in the solubility of metal compounds, which is dependent on the specific pH and temperature. Materials with low solubility, such as transition metal hydroxides or oxalates, are precipitated. Therefore, precipitants such as sodium hydroxide, trisodium phosphate, and sodium carbonate are mainly used to precipitate other dissolved metals, followed by lithium extraction in the form of lithium carbonate or trisodium phosphate by reacting lithium cation and precipitants. For example, lithium carbonate has a higher tendency to form low-solubility lithium carbonate under high pH. This fact is due to the formation of $CO_3^{2-}$ ions. Also, the solubility decreases at higher temperatures (12.9 g $L^{-1}$ at 25 °C, 10.8 g $L^{-1}$ at 40 °C), which makes precipitation easier. Compared to other compounds (LiOH – 129 g $L^{-1}$, LiCl – 815 g $L^{-1}$ at 25 °C) lithium carbonate has a substantially lower solubility, which makes it easier to precipitate. [9]

In this article, we focus on recovery of lithium from waste water using sodium carbonate and different approaches. The resulting product was subjected to scanning electron microscopy (SEM) with energy dispersive spectroscopy (EDS) and X-ray diffraction spectroscopy (XRD) to determine purity. It has been found that it is possible to achieve extraction efficiencies exceeding 80%.

## II. Recovery of spent LITHIUM

In this work, two different experiments were made. For first experiment, lithium metal from retired half-cells was used. The lithium was dissolved in demineralized water. The dissolution of lithium follows the following equation:

$$2Li\ (s) + 2H_2O\ (l) \rightarrow 2LiOH\ (l) + H_2\ (g)$$

Lithium was dissolved until a saturated solution was formed. From the saturated solution (12.8 g/100 g (20 °C)), 56.4 g was pipetted off. An equimolar amount of sodium carbonate in 100 ml of demineralized water was then added to the lithium hydroxide solution. The lithium carbonate formed was subsequently precipitated by using the reduced solubility of lithium carbonate with increasing temperature. The solution was brought to a boil using hotplate and the resulting product was filtered off. The resulting filtrate was dried, crushed and subjected to EDS and XRD analysis.

Since a high concentration of lithium in the wastewater is required for temperature precipitation, a solution with a lower concentration of lithium or lithium hydroxide was chosen for precipitation followed by crystallization. Total of 0.5 g of lithium metal from the half-cell was dissolved. Solution was recrystallized. In the crystallized mixture, sodium hydroxide is assumed to be present, which was formed by the reaction:

$$LiOH\ (l) + Na_2CO_3\ (l) \rightarrow Li_2CO_3\ (l) + NaOH\ (l)$$

The crystallized material was added to a beaker and 150 ml of ethanol was added to dissolve the sodium hydroxide.

$$C_2H_5OH\ (l) + NaOH\ (s) \rightarrow C_2H_5ONa\ (l) + H_2O\ (l)$$

$Li_2CO_3$ is insoluble in ethanol, thus the resulting suspension was filtered and washed with ethanol. The resulting filtrate was dried, rubbed and subjected to EDS and XRD analysis.

## III. Results and discussion

The yield of lithium carbonate using reduced solubility with increasing temperature was 0.9053 g. SEM image is shown in Fig. 1.

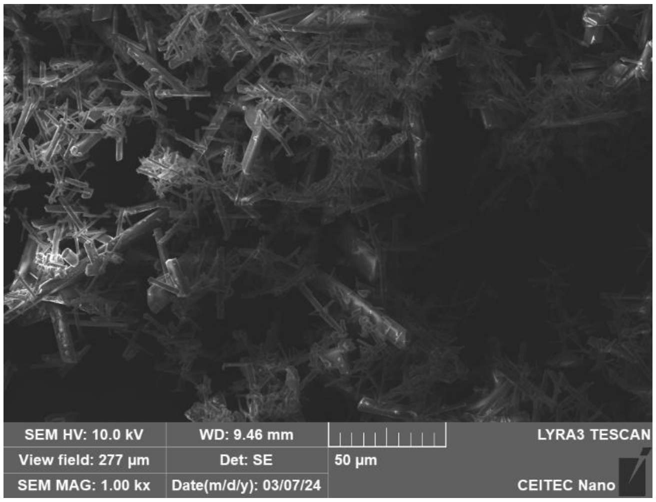



Fig. 1. SEM image of Lithium carbonate after 100°C precipitation

EDS shows presence of sodium, oxygen and carbon. EDS analysis is shown in Fig. 2. XRD analysis confirmed presence of sodium carbonate in temperature-driven precipitate. Mixture consists of lithium carbonate in 95.7 % and sodium carbonate in 4.3 %. XRD spectra is shown in Fig. 3. No other elements are present, so the mixture is free of electrolyte residues.

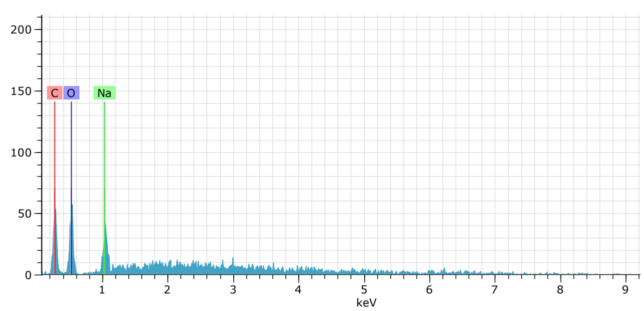



Fig. 2. EDS spectra of temperature-driven precipitation

The amount of lithium hydroxide in the solution before precipitation is 6.4 g. Therefore, the total amount of lithium corresponds to 1.84 g according to molar weight distribution. In 0.9053 g of precipitated salt of purity 95.7 % is 0.8664 g of lithium carbonate. According to molar distribution, there is only 0.16 g of pure lithium, according to molar mass distribution. The total lithium yield is therefore only 8.84 %.

Lithium carbonate powder after temperature-driven precipitation is shown in Fig. 4.

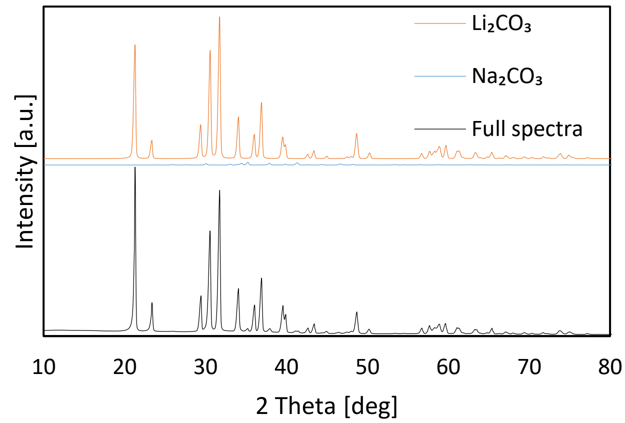



Fig. 3. XRD of Lithium carbonate after 100°C precipitation

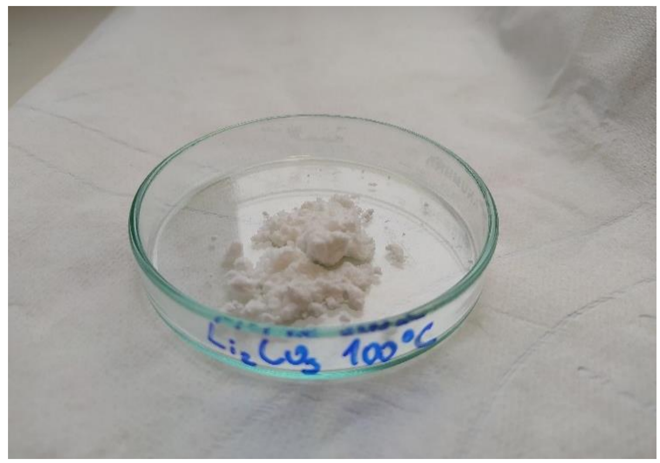


Fig. 4. Lithium carbonate after 100 °C precipitation

The second method was quantitative crystallization followed by ethanol washing. The yield was 2.826 g. SEM image of resulting substance is shown in Fig. 5.

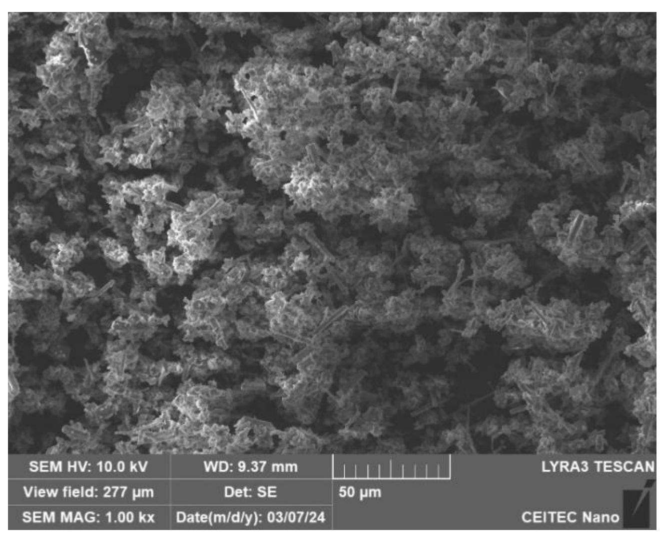



Fig. 5. SEM image of Lithium carbonate after crystallization and ethanol purification

EDS analysis shows presence of Carbon, Oxygen, Sodium and Silicone. Silicone contamination is less than 1 %. EDS measurement is shown in Fig. 6.

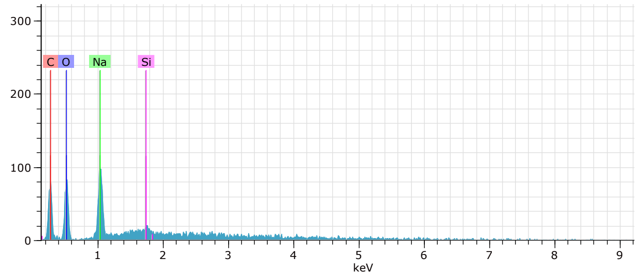



Fig. 6. EDS of Lithium carbonate after crystallization and purification and ethanol purification

XRD measurement is shown in Fig 7. As found, 84.6 % of the mixture is pure lithium carbonate. The remaining 15.4 % is lithium carbonate precipitate. No silicone was detected. From the known concentration of lithium carbonate in filtered substance and the initial concentration of lithium in the solution, we found the efficiency of the whole process to be 89.7 % in favor of lithium recovery. Total lithium recovered was calculated according to molar mass to be 0.45 g. The sample examined did not contain detectable amounts of electrolyte residue or other contamination except small amount of Silicone. Sodium hydroxide was completely dissolved and removed from the substance using ethanol. Lithium carbonate powder after crystallization and ethanol purification is shown in Fig. 8.

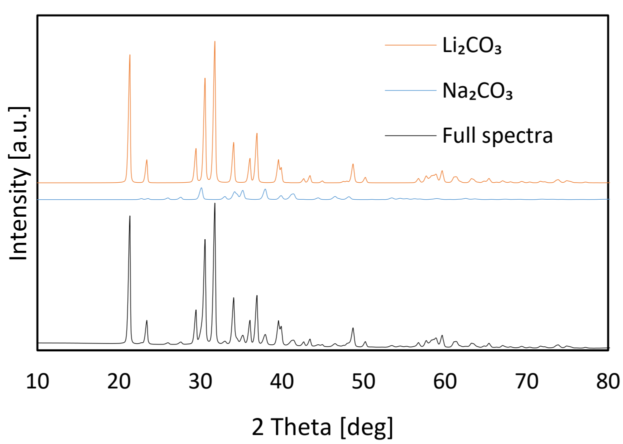



Fig. 7. XRD spectra of Lithium carbonate after crystallization and ethanol purification

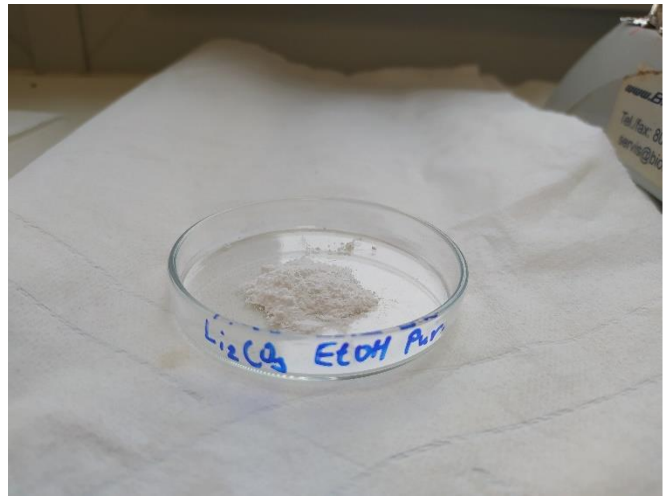


Fig. 8. Lithium carbonate after crystallization and purification and ethanol purification

## IV. CONCLUSION

It has been found that precipitation of lithium carbonate by reducing the solubility of lithium carbonate with increasing temperature yields a product of very high purity. However, practical application is not viable as obtaining a sufficient concentration for this method is challenging. In addition, a significant amount of lithium also remains in the water even when filtered at 100 °C, due to yield of temperature-driven precipitation is only 8.84 %. The method of crystallization of the mixture followed by washing with ethanol yields a product with a purity of 84.6 %. Precipitation followed by crystallization can also be used for low concentrations of lithium in wastewater. The resulting product is free of residual elements from the electrolyte after sodium hydroxide extraction with ethanol. Therefore, it can be said that the method of precipitating lithium using sodium carbonate to form lithium carbonate is an efficient recycling method that achieves an efficiency of lithium recovery around 90 %.

## ACKNOWLEDGMENT


This work was supported by the specific graduate research of the Brno University of Technology No. FEKT-S-23-8286. SEM and EDS were possible thanks to instrumentation of CEITEC Brno University of Technology.


## REFERENCES


[1] L. Martins, L. Guimarães, A. Botelho Junior, J. Tenório and D. Espinosa, "Electric car battery: An overview on global demand, recycling and future approaches towards sustainability", *Journal of Environmental Management*, vol. 295, 2021.

[2] E. Foreman, W. Zakri, M. Hossein Sanatimoghaddam, A. Modjtahedi, S. Pathak, A. Kashkooli, N. Garafolo and S. Farhad, "A Review of Inactive Materials and Components of Flexible Lithium-Ion Batteries", *Advanced Sustainable Systems*, vol. 1, no. 11, 2017.

[3] S. Kim, J. Kim, S. Kim, J. Lee and J. Yoon, "Electrochemical lithium recovery and organic pollutant removal from industrial wastewater of a battery recycling plant", *Environmental Science: Water Research & Technology*, vol. 4, no. 2, pp. 175-182, 2018.

[4] A. Zanoletti, E. Carena, C. Ferrara and E. Bontempi, "A Review of Lithium-Ion Battery Recycling: Technologies, Sustainability, and Open Issues", *Batteries*, vol. 10, no. 1, 2024.

[5] F. Larouche, F. Tedjar, K. Amouzegar, G. Houlachi, P. Bouchard, G. Demopoulos and K. Zaghib, "Progress and Status of Hydrometallurgical and Direct Recycling of Li-Ion Batteries and Beyond", *Materials*, vol. 13, no. 3, 2020.

[6] X. Luo, B. Guo, J. Luo, F. Deng, S. Zhang, S. Luo and J. Crittenden, "Recovery of Lithium from Wastewater Using Development of Li Ion-Imprinted Polymers", *ACS Sustainable Chemistry & Engineering*, vol. 3, no. 3, pp. 460-467, 2015.

[7] T. Tsuruta, "Removal and recovery of lithium using various microorganisms", *Journal of Bioscience and Bioengineering*, vol. 100, no. 5, pp. 562-566, 2005.

[8] J. Lemaire, L. Svecova, F. Lagallarde, R. Laucournet and P. Thivel, "Lithium recovery from aqueous solution by sorption/desorption", *Hydrometallurgy*, vol. 143, pp. 1-11, 2014.

[9] H. Bae and Y. Kim, "Technologies of lithium recycling from waste lithium ion batteries: a review", *Materials Advances*, vol. 2, no. 10, pp. 3234-3250, 2021.

# Impact of the Number of Measurements on Result Distortion

Jan Kucharik
*Department of Electrical and Electronic Technology, FEEC*
*Brno University of Technology*
0000-0002-5479-5959

Antonin Simek
*Department of Electrical and Electronic Technology, FEEC*
*Brno University of Technology*
0000-0002-1991-1029

***Abstract* — This study investigates the distortion of results due to the number of corrosion current measurements. Samples were prepared by pressing and subsequently sintering iron and magnesium powders. The samples were then immersed in an SBL solution at 37 °C. Measurements took place over 23 days. For two samples, measurements were taken each time, while the remaining samples were measured in sequence on their respective days.**

***Keywords—corrosion, corrosion current, simulated body fluid, solution, oxidation***

## I. Introduction

Corrosion of materials surrounds us in almost all human activities. Corrosion is the gradual degradation of metallic materials caused by a chemical reaction with the surrounding environment. The term is most commonly used in the sense of electrochemical oxidation of metals in reaction to an oxidizing agent, such as oxygen. The most well-known example of electrochemical corrosion is rusting, the formation of iron oxides. In this type of damage, oxides or salts of the original metal typically form. Corrosion can also occur in materials other than metals, such as ceramics or polymers, although in this context, the more common term is degradation. Material corrosion reduces its useful properties and structural integrity, such as strength, appearance, and permeability to fluids and gases.

## II. Corrosion

Corrosive processes occur when a metallic material reacts with the environment, especially with water and oxygen. As a result of electrochemical reactions taking place on the surface of the metallic material, a corrosion current is generated. This corrosion current can lead to faster damage to metallic surfaces, known as corrosion erosion. These reactions are typically induced by the presence of moisture and oxygen on the metal. In simple terms, the corrosion current is associated with the electrochemical behavior of the metallic material in an environment containing electrolytes, such as water or ions in the soil.

In explaining the causes and course of corrosive processes, it is based on the fact that a metal tends to transform into one of its more stable compounds found in nature. This manifests as a spontaneous process where the material tends to transition from the metallic form to the more stable forms of oxides, hydroxides, or salts. Each metal surface has a greater or lesser ability to react with components in the environment. In such a reaction, there is an exchange or association of electrons between the relevant reacting components [1].

Corrosive processes can be categorized based on whether the reaction occurs through the direct connection of two atoms or involves the reaction through ions of the reacting components in the corrosion process. A typical representative of the first type of corrosion process is chemical corrosion. This reaction occurs spontaneously for most metals. Very thin oxide layers are formed on the metal surface, which are not considered corrosion product layers in common technical practice. While this is a significant corrosion mechanism, it is not decisive. Corrosive processes associated with electrochemical reactions have greater importance, where the corroding system constitutes a system of galvanic cells. In this system, the electrode system usually consists of the corroding metal, and the corrosive environment acts as the electrolyte. This type of corrosive process is referred to as electrochemical corrosion.

### A. Chemical Corrosion

Chemical corrosion refers to a corrosive process in electrically non-conductive environments, such as certain gases and non-conductive liquids. The most common case of chemical corrosion is oxidation. In a simplified scheme, the oxidation of iron by atmospheric oxygen can be described by (1) [1]:

$$2\,Fe + O_2 \rightarrow 2\,FeO + 1/2\,O_2 \rightarrow Fe_2O_3 \qquad (1)$$

The rate of the corrosion reaction is highly dependent on temperature. With an increase in temperature from 10 to 20 °C, it can multiply several times. Corrosive reactions produce corrosion products that form a layer at the gas-metal interface. A continuous corrosion layer prevents direct access of the corrosive medium to the metal surface. If the contact between the metal and the corrosive environment occurs only through the diffusive oxide layer, the corrosion process gradually slows down - the layer has a protective character (e.g., $Al_2O_3$ on aluminum). The reaction rate is further dependent on the diffusion of reaction components (gas molecules, ions, electrons) through this layer. Diffusion is generally described by Fick's laws. The layer of corrosion products can act as a diffusion barrier, significantly slowing down the reaction. The ability of the layer of corrosion products to protect the metal (passivate it) is described by the Pilling-Bedworth rule. The equation for calculating this rule is given by (2) [2].

$$PB = \frac{V_{oxid}}{V_m} = \frac{M_{oxid} \cdot \rho_{oxid}}{M_m \cdot \rho_m \cdot n} \tag{2}$$

$V_{oxid}$ is the molar volume of the oxide layer, $V_m$ is the molar volume of the metal, $M_{oxid}$ is the molar mass of the oxide, $M_m$ is the molar mass of the metal, $\rho_{oxid}$ is the density of the oxide, $\rho_m$ is the density of the metal, and n is the number of metal atoms in the oxide [2].

### *B. Electrochemical corrosion*

Electrochemical corrosion refers to corrosion in a conductive environment - a liquid electrolyte. This type of corrosion is a result of electrochemical processes similar to those in a galvanic cell. An electric double layer begins to form at the interface between the metal and the electrolyte. Over time, a dynamic equilibrium is established, where the number of ions released by the metal equals the number of ions striking the metal surface. For each metal, the thus-formed potential difference at the metal-electrolyte interface has a specific value, known as the absolute (equilibrium) electrode potential. Its magnitude cannot be directly measured unless there is a disturbance in the equilibrium state. Therefore, relative potentials, i.e., potentials between individual electrodes, are determined. To enable electrode comparison, the hydrogen electrode was introduced as a standard, with its potential considered as zero. For individual metals, the so-called standard potential is then determined. This potential is defined as the relative potential of a given metal immersed in a solution of its salt with unit activity at a temperature of 18 °C, measured against the standard hydrogen electrode. The standard potentials of some metals are provided Fig. 1 [3].

Neglecting other factors influencing corrosion for now, it can be said that the more negative the standard potential of a metal (the less noble the metal), the more likely it is to be attacked by electrochemical corrosion.

The formation of electrode potential is the basis of electrochemical corrosion processes. However, this potential can only manifest itself when there is a disturbance in the equilibrium on the electrode. This occurs when depolarization takes place. Without depolarization, a portion of ions would transfer from the metal to the solution, and after reaching equilibrium, further dissolution would stop – the electrode would polarize.

The most common depolarizers in corrosion processes are atmospheric oxygen, hydroxide ions, and water molecules. Summarized reactions during oxygen depolarization are as follows, acidic environment (3), and alkaline environment (4)

$$O_2 + 4\,H^+ + 4\,e^- \rightarrow 2\,H_2O \tag{3}$$

$$O_2 + 2\,H_2O + 4\,e^- \rightarrow 4\,OH^- \tag{4}$$

Hydrogen depolarization can be described alkaline environment (5) acidic environment (6)

$$2\,H^+ + 2\,e^- \rightarrow H_2 \tag{5}$$

$$2\,H_2O + 2\,e^- \rightarrow H_2 + 2OH^- \tag{6}$$

Increasing strength of reducing agent

| Electrode | Reaction | SRP (at 298 K) |
|---|---|---|
| *Li | $Li^+ + e^- \rightarrow Li(s)$ | – 3.05 V |
| K | $K^+ + e^- \rightarrow K(s)$ | – 2.93 V |
| Ba | | |
| Ca | $Ca^{+2} + 2e^- \rightarrow Ca(s)$ | – 2.87 V |
| Na | $Na^+ + e^- \rightarrow Na(s)$ | – 2.71 V |
| Mg | $Mg^{+2} + 2e^- \rightarrow Mg(s)$ | – 2.37 V |
| Al | | |
| * Electrolytes ($H_2O$) | $H_2O(l) + e^- \rightarrow ½\,H_2 + OH^-$ | – 0.828 V |
| *Zn | $Zn^{+2} + 2e^- \rightarrow Zn(s)$ | – 0.76 V |
| Cr | $Cr^{+3} + 3e^- \rightarrow Cr(s)$ | – 0.74 V |
| *Fe | $Fe^{2+} + 2e^- \rightarrow Fe$ | – 0.44 V |
| Cd | $Cd^{+2} + 2e^- \rightarrow Cd(s)$ | – 0.40 V |
| Co | | |
| Ni | $Ni^{+2} + 2e^- \rightarrow Ni(s)$ | – 0.24 V |
| Sn | $Sn^{+2} + 2e^- \rightarrow Sn(s)$ | – 0.14 V |
| Pb | $Pb^{+2} + 2e^- \rightarrow Pb(s)$ | – 0.13 V |
| *$H_2$ | $2H^+ + 2e^- \rightarrow H_2(g)$ | 0.00 V |
| Cu | $Cu^{2+} + 2e^- \rightarrow Cu(s)$ | 0.34 V |
| Fe | $Fe^{3+} + e^- \rightarrow Fe^{2+}$ | 0.77 V |
| Hg | $Hg_2^{2+} + 2e^- \rightarrow Hg(l)$ | 0.79 V |
| Ag | $Ag^+ + e^- \rightarrow Ag$ | 0.80 V |
| Hg | $Hg^{2+} \rightarrow Hg(l)$ | 0.85 V |

Fig. 1 The standard potentials of some metals [3].

Hydrogen depolarization can be described as an Alkaline environment (7) and an Acidic environment (8)

$$2\,H^+ + 2\,e^- \rightarrow H_2 \tag{7}$$

$$2\,H_2O + 2\,e^- \rightarrow H_2 + 2\,OH^- \tag{8}$$

In the actual electrochemical corrosion process, both partial, locally separated reactions occur continuously on the metal surface.Anodic - leading to oxidation (loss of electrons) (9). Cathodic - leading to reduction (acceptance of electrons and disturbance of the equilibrium established at the anode) (10).

$$Me \rightarrow Men^+ + n{\cdot}e^- \tag{9}$$

$$Men^+ + n{\cdot}e^- \rightarrow Me \tag{10}$$

The site where metal dissolution occurs is called the anode, and the site where excess electrons are neutralized is the cathode. The spatial separation of the anodic and cathodic processes is not a necessary condition for the initiation and development of electrochemical corrosion processes. However, if it exists, it accelerates the progress of anodic and cathodic reactions. Electrochemical corrosion can thus be defined as a process in which the material undergoes simultaneous oxidation, and some components of the solution undergo reduction. The system in which such processes occur forms a so-called redox system.

## III. Potentiodynamic tests

These are primarily used on metallic samples in environments with good conductivity (low values of polarization resistance are sufficient). The result of the measurement is the dependence of current density on the change in potential of the sample E against RE, which can be, for example, a saturated calomel electrode (SCE). The circuit also includes auxiliary (CE) and working (WE) electrodes [4]. A sample immersed in a corrosive environment against RE exhibits a certain potential E_corr. In this state, both anodic and cathodic currents are generated on the surface of the sample, which have the same magnitude because the sample is in equilibrium with the surrounding environment (the rate of oxidation and reduction is the same). After polarization (using an external voltage source) of the sample in the positive direction, the anodic current increases at the expense of the cathodic current, which gradually decreases to a negligible value. The same principle applies in reverse for polarizing the sample in the negative direction [4]. In real cases, current density is measured at a certain potential. The current is plotted in a logarithmic scale, and the potential in linear coordinates. The result is thus a potentiodynamic polarization curve (semilogarithmic). Using Tafel analysis (Fig. 2), it is possible to determine the slope of the anodic (ba) and cathodic (bk) parts of the curve. The linear segments intersect at the value of $E_{corr}$ and the corrosion current density $i_{corr}$, it is possible to obtain from the resulting intersection point the corrosion potential, indicating the metal's tendency to corrode, and the corrosion current density, indicative of the corrosion kinetics [4].

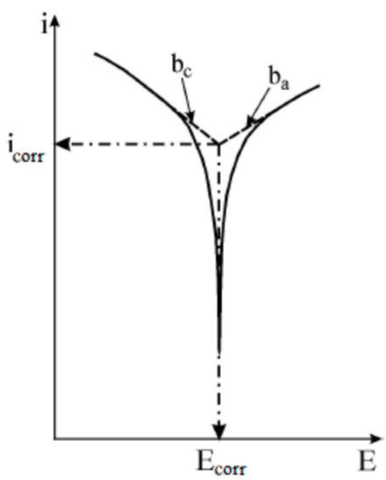



Fig. 2 Tafel analysis [4]

## IV. Sample preparation

For sample production, commercially available iron powder (Fig. 2) with grain size up to 50 µm and magnesium powder were selected. The magnesium powder was ground in a ball mill for 20 minutes with an amplitude of 1 mm to achieve a grain size of up to 500 µm (Fig. 3). The intentional difference in grain size was chosen to create internal porosity during the sintering of the samples. Subsequently, these metal powders were mixed in ratios of 80:20 weight percent (samles A) and 70:30 weight percent (samples B) with an iron predominance. This mixture was placed in containers and stirred for 4 hours using a rotary mixer.

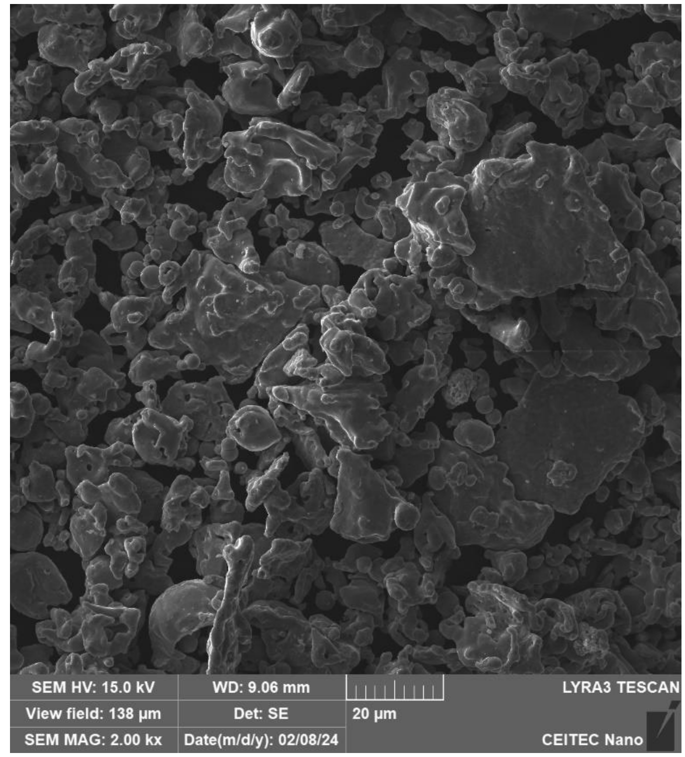



Fig. 3 Iron powder

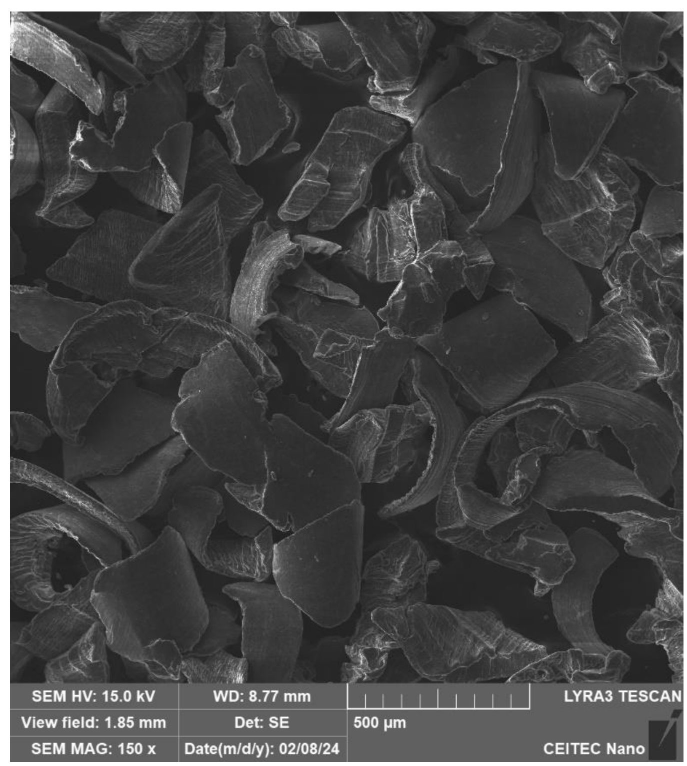



Fig. 4 Magnesium powder after grinding

The homogenized mixture was placed into a pressing mold with a diameter of 10 mm and then compressed at 100 $kPa \cdot cm^{-2}$ in a press. The resulting pellets were placed in a tubular furnace and subjected to a two-phase sintering process in a protective atmosphere of argon. The first stage was at a temperature of 450 °C. The temperature gradient was 5 °C per minute. The samples remained at this temperature for 120 minutes. The second phase was the sintering of metal powders at a temperature of 1000 °C. The samples remained at this temperature for 60 minutes. The temperature gradient was identical to the first phase, i.e. 5 °C per minute. Then they ware slowly cooled.

The fired samples were polished with 600 and then 1200 grit emery paper using ethyl alcohol instead of water to prevent the initiation of a corrosion reaction. The samples were drilled from the side and equipped with an M2 thread, into which a threaded rod, 50 mm in length, was screwed. This rod was covered with heat shrink tubing on most of its surface, except for the parts screwed into the sample and the exposed end for connecting to the potentiostat connector. The M2 thread represents only a negligible fraction of the sample surface, thus introducing minimal error into the measurement.

## V. Solution

The solution which was prepared was a simulated body fluid (SBF). This solution has ion concentrations nearly equal to those of human blood plasma. The SBF was approved in 2003 the Technical Committee ISO/TC150 of International Organization for Standardization. To prepare 1000 ml of SBF, 700 ml of ion-exchanged and distilled water and reagents which is shown in Table 1 were used. 1M-HCl was used to lower the pH. [5]

Table 1 Reagents used for preparing 1000 ml of SBF [5]

| Order | Reagent | Amount | Purity (%) | Formula weight |
|---|---|---|---|---|
| 1 | NaCl | 8.035 g | 99.5 | 58.44 |
| 2 | $NaHCO_3$ | 0.355 g | 99.5 | 84.00 |
| 3 | KCl | 0.225 g | 99.5 | 74.55 |
| 4 | $K_2HPO_4 \cdot 3H_2O$ | 0.231 g | 99.0 | 228.22 |
| 5 | $MgCl_2 \cdot 6H_2O$ | 0.311 g | 98.0 | 203.30 |
| 6 | 1.0M-HCl | 39 ml | - | - |
| 7 | $CaCl_2$ | 0.292 g | 95.0 | 110.98 |
| 8 | Na2SO4 | 0.072 g | 99.0 | 142.04 |
| 9 | Tris | 6.118 g | 99.0 | 121.13 |
| 10 | 1.0M-HCl | 0–5 ml | - | - |

Nominal ion concentrations of SBF in comparison with those in human blood plasma is show in table 2.

Table 2 Comparison of SBF with human blood plasma [5]

| | Ion concentrations ($mmol \cdot l^{-1}$) | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Ion | $Na^+$ | $K^+$ | $Mg^{2+}$ | $Ca^{2+}$ | $Cl^-$ | $HCO_3^-$ | $PO_4^{3-}$ | $SO_4^{2-}$ |
| Blood plasma | 142.0 | 5.0 | 1.5 | 2.5 | 103.0 | 27.0 | 1.0 | 0.5 |
| SBF | 142.0 | 5.0 | 1.5 | 2.5 | 148.8 | 4.2 | 1.0 | 0.0 |

placed in 100 ml of SBL and stored in the dark at 37 °C, simulating the temperature of the human body.

SBF solution is a chlorides, sulphate and phosphates containing solution (Table 1), and the increase in pH value due to the corrosion of the pure iron promotes the precipitation and deposition of phosphates [6].

Expected reaction of iron in SBF [6]:

$$Fe(OH)_2 + Cl^- \rightarrow FeClOH + OH^- \quad (11)$$

$$FeClOH + H^+ \rightarrow Fe^{2+} + Cl^- + H_2O \quad (12)$$

$$Fe^{2+} + O_2 + 3\,OH^- \rightarrow Fe(OH)_3 \downarrow + O_2^- \quad (13)$$

$$Fe(OH)_3 + 2Cl^- \rightarrow FeCl_2OH + 2\,OH^- \quad (14)$$

$$FeCl_2OH + H^+ \rightarrow Fe^{3+} + 2\,Cl^- + H_2O \quad (15)$$

$$2\,PO_4^{3-} + 3\,Ca^{2+} \rightarrow Ca_3(PO_4)_2 \downarrow \quad (16)$$

$$2\,PO_4^{3-} + 3\,Mg^{2+} \rightarrow Mg_3(PO_4)_2 \downarrow \quad (17)$$

$$2\,PO_4^{3-} + 3\,Fe^{2+} + 8\,H_2O \rightarrow Fe_3(PO_4)_2 \cdot 8\,H_2O \quad (18)$$

$$PO_4^{3-} + Fe^{3+} \rightarrow FePO_4 \downarrow \quad (19)$$

Expected reaction of magnesium in SBF [7]:

$$Mg(OH)_2 + 2\,Cl^- \rightarrow Mg^{2+} + 2\,OH^- \quad (20)$$

$$3\,Mg(H_2PO_4)_2 + 4\,H_2O \rightarrow \quad (21)$$
$$\rightarrow Mg_3(PO_4)_2 \cdot 8\,H_2O + H_3PO_4$$

$$H_2PO_4^- + 2\,OH \rightarrow PO_4^{3-} + H_2O \quad (22)$$

$$HPO_4^{2-} + OH^- \rightarrow PO_4^{3-} + H_2O \quad (23)$$

## VI. Analysis

Before drilling the hole, the samples underwent elemental analysis, and their surfaces were simultaneously mapped. It was found that a significant amount of magnesium evaporated during firing due to an inappropriately chosen firing profile. This fact was confirmed by the deposition of magnesium on the tube furnace wall. Specifically, for sample A, instead of 20%, only 1.4% of the weight percentage of magnesium was present, and for sample B, instead of 30%, only 1.2% of the weight percentage of magnesium was present.

Figures 4 and 5 show different defects in the surface structure. While cracks are not as prominent in sample A, sample B exhibits defects visible to the naked eye. This difference is caused by a larger volume of evaporated magnesium during firing.

The polished samples were weighed, and pieces with approximately equal weights were selected from the total number.

For the polished samples equipped with a measuring electrode, the corrosion current was measured. SBL was chosen as the measuring medium since research is concurrently being conducted on these samples to assess their suitability as a biodegradable bone substitute. The measured samples were placed in 100 ml of SBL and stored in the dark at 37 °C, simulating the temperature of the human body.

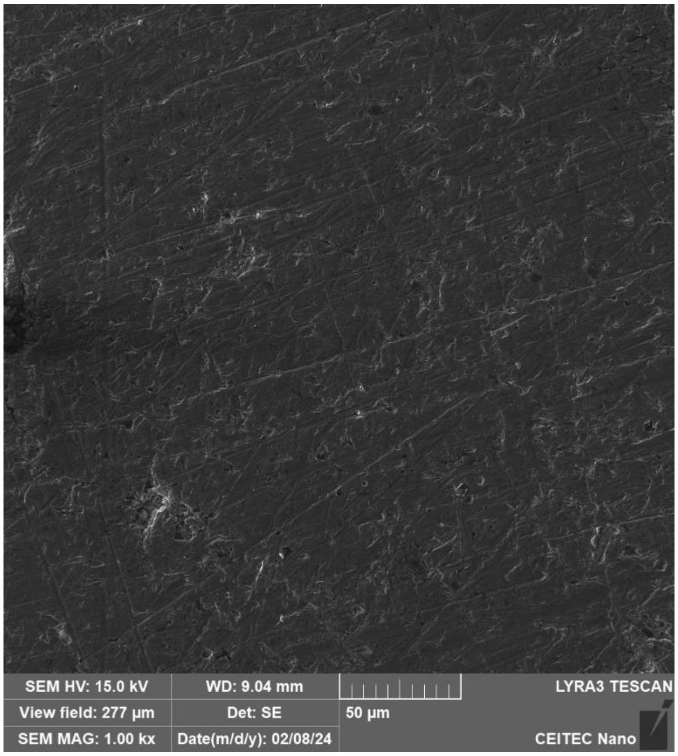



Fig. 5 Sample A after polishing

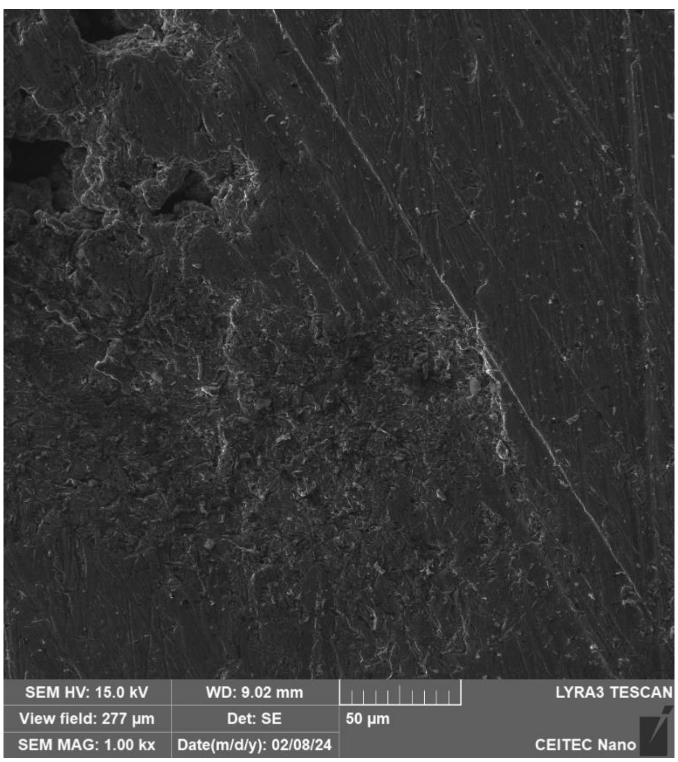



Fig. 6 Sample B after polishing

The measurement was carried out on the potentiostat in the linear polarization mode with measurement limits set at -9 V and 1 V, and a rate of change of the applied potential of 1 mV/s.

The samples were divided by composition into two groups, A1-A5 and B1-B5. Samples A1 and B1 were measured in each session. The remaining samples were measured in sequence on the respective days. Samples A1 and B1 were measured a total of 11 times, while the others were measured only 4 times, including the initial measurement.

## VII. Conclusion

The experiment demonstrates a direct correlation between the number of measurements and the reduction of the corrosion current of a given sample. When measuring the corrosion current using a three-electrode system potentiostat, there is an amplification of the oxidation reaction on the surface of the sample. This reaction gradually forms a passivation layer, which reduces the value of the corrosion current and thus slows down further oxidation reactions. For clarity, the dependence on the number of measurements has been converted to a dependence on days in the solution, as the samples are simultaneously exposed to chemical reactions with the solution. Figure 6 illustrates the corrosion current trends of samples labeled as A. From the graphical dependence of the corrosion current on the days in the solution, it can be inferred that the corrosion current of sample A1 has a significant decrease compared to other samples. This means that a larger passive layer is formed on the surface, which prevents further corrosion. Another influencing factor is that the sample is removed from the solution more frequently, and the oxidation factor in the air also contributes. The same behavior is expected for samples labeled as B, where again, B1 shows a substantial drop in the corrosion current for the same reason as sample A1.

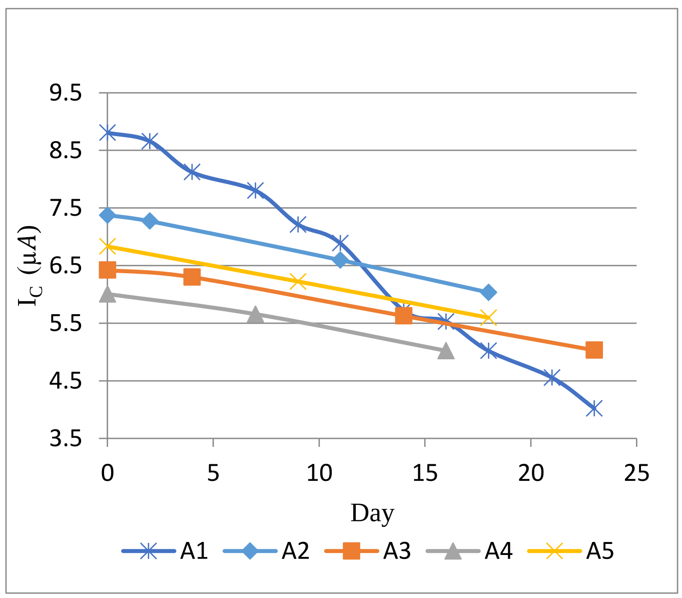



Fig. 7 Difference in corrosion current between samples A

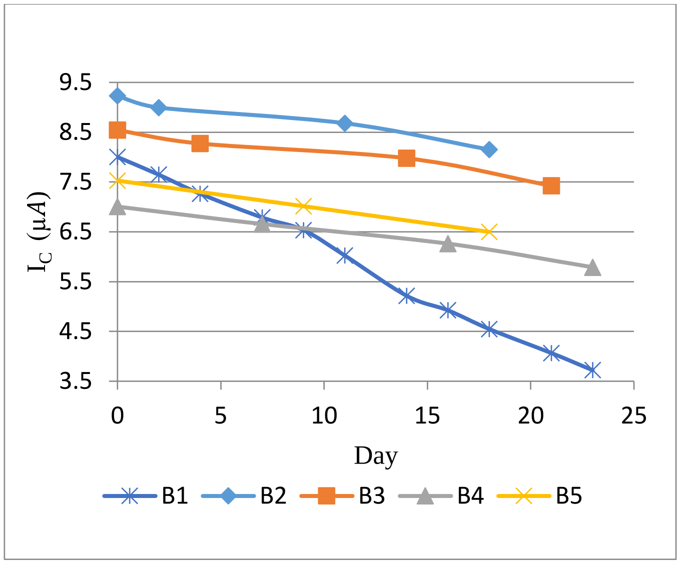



Fig. 8 Difference in corrosion current between samples B

The increased drop in corrosion current between some measurements may also be caused by varying lengths of time during which the sample was out of the solution. This duration depended on the dependency of installing the measuring electrode into the sample.

Recommendation: Based on the above, it is advisable to carefully consider the number and frequency of measurements when measuring the corrosion current using a potentiostat, as measurements can lead to significant distortion. Another option is to use cyclic voltammetry instead of the potentiodynamic method. In cyclic voltammetry, oxidation and subsequent reduction reactions occur, canceling out each other's effects.

## ACKNOWLEDGMENT

This work was supported by the specific graduate research of the Brno University of Technology No. FEKT-S-23-8286.

CzechNanoLab project LM2023051 funded by MEYS CR is gratefully acknowledged for the financial support of the measurements/sample fabrication at CEITEC Nano Research Infrastructure.

Acknowledgments go to Associate Professor Maria Sedlarikova for her methodological guidance and to Associate Professor Opravil from the Chemical Faculty for producing samples under challenging conditions.

## REFERENCES

[1] Mechanizmus koroze. Online. Dostupné z: www.ib.cvut.cz/doku.php?id=koroze:mechanizmus. [cit. 2024-03-09].

[2] MATĚJKA, L. Protikorozní ochrana hořčíkových slitin ZE10 a ZE41 pomocí fluoridových konverzních povlaků. Brno, Vysoké učení technické v Brně, Fakulta chemická, 2018.

[3] UNACADEMY. Electrochemical Series Chart. Online. Dostupné z: https://unacademy.com/content/jee/study-material/chemistry/electrochemical-series-chart/. [cit. 2024-03-10].

[4] HADZIMA, B.; LIPTÁKOVÁ, T. Základy elektrochemickej korózie kovov. Žilina: Žilinská univerzita v Žilině, 2008. ISBN 978-80-8070-876-4.

[5] SEDLAŘÍKOVÁ, M.; ZATLOUKAL, M.; KUCHAŘÍK, J.; ČUDEK, P.; FAFILEK, G.; DOLEŽELOVÁ, E. Corrosion Processes of Sintered Materials Based on Fe. Advanced Batteries, Accumulators and Fuel Cells -23th. Brno: Brno University of Technology, 2022. s. 52-55. ISBN: 978-80-214-6088-1.

[6] SEDLAŘÍKOVÁ, M.; ZATLOUKAL, M.; KUCHAŘÍK, J.; ČUDEK, P.; FAFILEK, G.; DOLEŽELOVÁ, E. Corrosion Processes of Sintered Materials Based on Fe. In Journal of Physics: Conference Series. Journal of Physics: Conference Series. IOP publishing, 2022. s. 1-9. ISSN: 1742-6596.

[7] HONGZHOU, P.; WEI, W.; HAOMIAO, J. Effect of Galvanic Corrosion on the Degradability of Biomedical Magnesium. 2021. Frontiers in Materials. 8. 10.3389/fmats.2021.767179.

PŘED VÍCE NEŽ 140 LETY VYNALEZL ZAKLADATEL ALEXANDER GRAHAM BELL TELEFON. AT&T MÁ OD TÉ DOBY DĚDICTVÍ OSMI NOBELOVÝCH CEN A VÍCE NEŽ 12 500 PATENTŮ PO CELÉM SVĚTĚ.

NAŠE JEDNOTKA, AT&T LABS NETWORK SYSTEMS, JE ZAMĚŘENA NA TRANSFORMACI PROSTORU SÍTĚ OSS POMOCÍ NOVÉ GENERACE TECHNOLOGIÍ AI, ANALÝZOU VELKÝCH DAT A DALŠÍCH DISTRIBUOVANÝCH SYSTÉMŮ.

PRO NÁŠ TÝM V BRNĚ HLEDÁME ZKUŠENÉ I ČERSTVÉ ABSOLVENTY SOFTWAROVÉHO INŽENÝRSTVÍ, KTEŘÍ CHTĚJÍ BÝT SOUČÁSTÍ DYNAMICKÉHO ROZVOJE V OBLASTI 5G/6G A SDN. STUDENTŮM IT NABÍZÍME INTERNSHIPS, ČÁSTEČNÉ PRACOVNÍ ÚVAZKY, TÉMATA A VEDENÍ BAKALÁŘSKÝCH A MAGISTERSKÝCH PRACÍ.

KONTAKT: KATEŘINA VELKÁ
KATERINA.VELKA@INTL.ATT.COM
AT&T GLOBAL NETWORK SERVICES
CZECH REPUBLIC
PALACHOVO NÁMĚSTÍ 2
625 00 BRNO

# AT&T

## INOVACE

1876
ALEXANDER GRAHAM BELL VYNALEZL TELEFON A ZALOŽIL BELL TELEPHONE COMPANY, DNEŠNÍ AT&T

1924
FAX

1939
PRVNÍ ELEKTRICKÝ DIGITÁLNÍ POČÍTAČ

1947
TRANZISTOR

1954
SOLARNÍ PANEL

1962
TELSTAR - PRVNÍ KOMUNIKAČNÍ SATELIT

1969
UNIX OS

1983
C++

1983
PRVNÍ OPTICKÝ KABEL MEZI USA, FRANCIÍ A VELKOU BRITÁNIÍ

1993
VIDEO KOMUNIKACE

# Synthesizing Submicron Particles for Li-ion Batteries: Spray-Drying & Electrostatic Precipitation


Ing. Hana Hálová
*Department of Electrical and Electronic Technology*
*Brno University of Technology*
Brno, Czech Republic
xhalov01@vutbr.cz

Ing. Ladislav Chladil, Ph.D.
*Department of Electrical and Electronic Technology*
*Brno University of Technology*
Brno, Czech Republic
chladil@vut.cz

Ing. Bc. Lukáš Preisler
*Energy Institute, Department of Power Engineering*
*Brno University of Technology*
Brno, Czech Republic
Lukas.Preisler@vut.cz



***Abstract*—The paper explores innovative methods for producing submicron particles for Li-ion batteries, focusing on spray-drying and electrostatic precipitation. Spray-drying synthesis faces challenges in generating sufficient quantities of product. To address this and also due to the high initial cost, an apparatus integrating an ultrasonic atomizer, drying chamber and modified electrostatic precipitator was designed. This setup enables precise control of parameters, ensuring the production of homogeneous, small-sized particles.**



***Keywords*— spray-drying synthesis, electrostatic precipitation, Li-ion, cathode material, electrostatic precipitator design**


## I. Introduction

When it comes to investigating the cathode material, it is essential to focus on not just the chemical composition, but also pay attention to the internal structure and synthesized particles formed during the synthesis. Although the composition of the active electrode determines the properties of the battery (such as voltage, and theoretical capacity), the specific surface area in contact with the electrolyte and the energy density of the battery [1], [2], [3], the particle size affects the rate of oxidation-reduction processes, e.g. the rate of charge/discharge, capacity, and coulombic efficiency. It is believed that smaller particles with larger interfacial areas allow the diffusion of more lithium ions, leading to high ion current and specific capacity. The potential of oxidation peaks is also inferred to increase as the particle size decreases. However, as the particle size decreases, a larger fraction of irreversible capacity may be formed due to the SEI layer on the particle surface [1]. For this reason, it is important to control the particle growth and distribution of individual precursors throughout the synthesis.

Given these effects on battery parameters, microstructure becomes a key factor that receives research attention. The research is focused on finding the ideal synthesis technique for each cathode material. However, a major problem, of the various laboratory techniques is the difficulty in transferring the results from the laboratory scale to the industrial one. [2]

One of the frequently used laboratory techniques for the synthesis of cathode material is a solid-state reaction, offering a low-cost and simple process [4]. However, the disadvantages associated with this technique include difficult particle size control, low homogeneity, and the need for high temperatures. The total synthesis time ranges from 30 to 60 hours [2], [5]. Another popular method for the synthesis of the cathode material is the sol-gel method, characterized by uniform particle distribution, precise control of stoichiometric coefficients, and high purity. However, the disadvantage is again the longer process time and the high initial cost of special equipment [4]. Another synthesis that can be used to prepare cathode materials for Li-ion batteries is hydrothermal [4]. This approach offers the significant benefit of producing highly pure and homogeneous materials. Nevertheless, it is limited by its applicability to specific materials, necessitates high-pressure conditions, and often requires several days for synthesis.

The relatively simple scalability of production is a strength of spray-drying synthesis, that can be applied to the solution or suspension. The advantage of this synthesis lies in the possibility of simple dissolution of precursors in aqueous or ethanol environment, which leads to precise and uniform distribution of elements in the synthesized product [2]. The use of spray drying synthesis also suggests the possibility of reducing the temperature and length of the subsequent calcination process [4].

## II. Spray-drying synthesis for cathode materials of Li-ion batteries

The spray-drying synthesis is a commonly used technique in practical applications – such as the manufacture of dried food, and oxide ceramics, but also the pharmaceutical and dairy industry. The principle of this synthesis is heating the droplets obtained by atomization of the solution while evaporating the solvent to form a homogeneous powder. [6]

The synthesis process consists of several steps, which include *droplet generation* (atomizing, spraying), *conversion from droplet to particle* by drying (solvent evaporation), and *particle collection* (separation of particles from gas) [6]. The procedure of spray-drying synthesis is shown in Fig. 1.

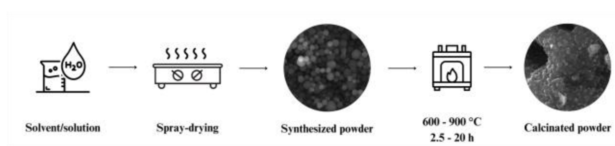



Fig. 1. Spray-drying synthesis with individual processes.

### A. Solution Preparation

To prepare the starting solution for powder formation using spray-drying synthesis, careful selection of precursors is required as all precursors dissolved in the solution will also be present in the final synthesized powder [2].

The most common solvent or liquid medium is water (used in [2], [7], but also ethanol [2] [8] or a combination of alcohol (ethanol) and water can be used [2] [9]. The selection of suitable solvent or liquid medium is influenced by price, safety, and toxicity, but also by the vaporization temperature which has to be usable in the used apparatus for spray-drying [2].

When preparing aqueous solutions, frequently used precursors are acetates [10] and nitrates [7] [11] for cations and ammonium salts for anions thanks to their low decomposition temperature [2].

Other organic substances (carboxylic acids, saccharides, synthetic polymers) but also carbon compounds (carbon nanotubes, graphene oxide, carbon black) are added to the solution [2]. These materials are added for various reasons, but mostly they work as reducing or complexing agents (citric acid) but can be also used as precursors transforming into carbon during subsequent heat treatment. Thus, the source of carbon to increase the electrical conductivity of the particles can be not only by mixing the particles with carbon, but also by mixing the particles with carbon or by adding precursors such as sucrose, glucose, and citric acid which are converted by heat into graphitic carbon. In the publication [12], the use of dextrin as a precursor is mentioned, which is used to reduce hygroscopicity of the material. [2]

When preparing the solution, it is important to pay attention also to the concentration [2]. High concentrations accelerate the onset of solubility limit surpassing solvent vaporization, potentially leading to crust formation on larger droplets. Adjusting solution concentration alongside other parameters like inlet temperature, flow rate, and atomization can optimize the drying process for specific priorities such as desired morphology, prevention of partial decomposition, or minimizing residual humidity. [2]

### B. Atomization of the Solvent/Solution

The first step to maintaining submicron particles is the atomization of the initial solution which increases the surface area. The particle size distribution of the synthesized powder is related to the initial spray in terms of the initial droplet size, however, that can be also influenced by particle deflation and agglomeration. [13]

That can be achieved by using several types of atomizers, such as ultrasonic nebulizers, two-fluid nozzles, pressure atomizers or rotating atomizers [13].

*Rotary atomizers* can handle flow rates of up to 200 tons/h [13] thanks to the drive of compressed air or an electrical motor thus providing the possibility of high performance. The droplets are formed by centrally feeding liquid into the atomizer, where it flows outward, dispersing into small droplets at the wheel's edge due to centrifugal force. Together with *the pressure nozzle* (capacity up to 400 l/h) [13] which forms droplets by pressurized liquid passing through a nozzle aperture is the most common type used. Another type of atomizer is *ultrasonic nozzle* whose advantage is the usage of a larger orifice that is less prone to clogging and a small particle size. [13]

### C. Droplet-to-particle conversion

Droplet-to-particle transformation is an important part of the process that involves the removal of the solvent from droplets atomized by the atomizer. Solvent removal can occur due to the *heat treatment using a hot furnace or a reactor*, h*eated carrier gas*, or thanks to the *solvent-diffusion process using a diffusion drier* [14].

The heating temperature and the whole process of heating must be precisely regulated to ensure that the solvent is thoroughly evaporated from the droplet [15]. The commonly chosen temperature in published research papers is between 190 – 220 °C for inlet temperature for aqueous solutions of nitrates or acetates [2], [7]. However, optimization of the process relies not only on temperature but also necessitates such as calculation of the minimal residence time to guarantee adequate drying duration.

### D. Particle Collection

Following the evaporation of the liquid component from the solution/suspension, it becomes imperative to collect the synthesized particles. This can be accomplished through the utilization of cyclone, filtration, wet scrubbing, or electrostatic precipitator, chosen based on the product and product amount being synthesized [14].

The specific method of particle collection affects not only the effective size of the collected particles but also the efficiency of the process. According to [14], gravity settling can capture particles larger than 100 μm with an efficiency of 40-50%. Higher efficiency is then achieved by cyclone capture - particles larger than 5 μm are captured with an efficiency of 85-95%. Filtration (90-99% efficiency for particles smaller than 2 μm) and electrostatic precipitators (90-99% efficiency for particles smaller than 10 μm) achieve the highest efficiency in particle collection. However, the purchase price of the equipment goes hand in hand with the higher efficiency of the process and it is therefore essential to take into account all the parameters and equipment requirements when setting up the spray-drying synthesis apparatus.

### E. Heat Treatment

Storing hygroscopic compounds or those prone to inorganic condensation after spray-drying is typically not recommended. For such substances, it's advisable to undergo immediate subsequent heat treatment to attain a stable and consistent intermediate state. However, also for other compounds formed by spray-drying synthesis, in most cases, further heat treatment is required to convert them to the final phase [2].

The thermal process may include precursor decomposition, solid-state diffusion, and crystallisation. The inorganic active material usually crystallises homogeneously in a single phase during spray-drying synthesis, but two phases can also be formed [2].

According to [16], [17], [18] the temperature of additional heat treatment is from 750 °C to 1,000 °C in the time range from 5 hours to 15 hours. Seenivasan [16] also uses preheating at 500 °C for 5 hours.

## III. SPRAY-DRYING APPARATUS DESIGN

When it comes to the usage of spray-drying synthesis in laboratory conditions, the process has to overcome several issues [2]. One of them is the high purchase price of spray-drying equipment and the unsatisfactory amount of synthesized powder. This is why spray-drying synthesis equipment is often manufactured and avoids the use of professional equipment, which tends to be expensive and only allows the synthesis of large quantities of powder. On the other hand, self-built spray-drying synthesis equipment often produces a very small amount of powder, which is not enough to create more electrodes. For this reason, when building a self-device, it is necessary to pay attention to the individual parameters that affect the amount of synthesized particles and the possible length of synthesis without the need for modification [2].

Parameters that significantly affect the spray-drying synthesis process are the design of the apparatus (type of atomizer, geometry of the drying chamber, and type of collector). These parameters affect the air velocity, the time for which the particle is dried, and the temperature. However, the process is also affected by the solution/solvent parameters - its concentration, the type of precursor with its boiling temperature, reactivity, agglomeration tendency, and many others [2], [6].

Based on the properties found, acetate precursors and aqueous solution will be used for the synthesis of cathode material, as in the work of [10]. An ultrasonic atomizer will be used to pulverize the solution, especially thanks to the formation of small particles. An insulated vessel will be used as the drying chamber, which will be heated using induction. The size of the drying chamber will be adjusted to ensure sufficient heating of the entire system, including the electrostatic precipitator which will be used to trap particles – scheme of apparatus is in Fig. 2.

### A. *Principles of Electrostatic Precipitator Design*

When it comes to collecting particles, it is important to adapt to the requirements that arise from the intended use. Since for laboratory use, it is important to gain smaller but still sufficient quantities for electrode production, it is advisable to consider the use of an electrostatic precipitator. The electrostatic precipitator is used in many industries for the removal of sub-micron particles but is best known for its use in the removal of fly ash during combustion (plate electrostatic separator). By using an electrostatic precipitator, particles smaller than 10 μm are captured with an efficiency of approximately 95% [19]. The use of an electrostatic precipitator also offers the possibility of increasing the production of cathode material, since, unlike e.g. filter papers, it can be run for a longer period of time without the need to clean or replace some of the components.

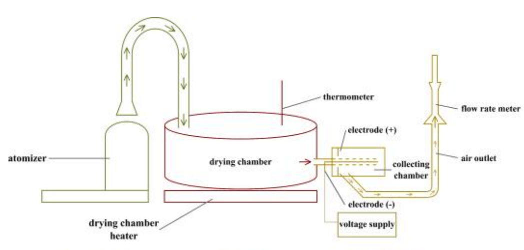



Fig. 2. Scheme of apparatus for spray drying synthesis.

For effective electrostatic precipitation, the following principles must be observed [19]:

- use of *DC voltage* to excite the electric field,
- most particles must be of the *same polarity*,
- there must be an electrical discharge in the gas - *corona formation*,
- the gas flow is *less than the critical value* at which the particles are stripped from the electrodes.

For the formation of corona discharge, the most curved surface is suitable - in this case the smallest possible radius of the wire (charging electrode). The corona discharge condition is expressed by the formula for reaching the initial electric field strength $E_0$. Initial *critical electric field strength* $E_0$ indicates the lowest electric field strength at which ionization occurs ($E_0$ rises with increasing gas temperature and gas pressure) – equation (1). [19], [20]

$$E_0 = k_1 \cdot \delta \cdot \left(1 + \frac{k_2}{\sqrt{\delta \cdot r_{char}}}\right) \tag{1}$$

Where $r_{char}$ represents charging electrode radius; $k_1$, $k_2$ empirical values for corona discharge. The relative density of the gas δ (see equation (2)) corresponds to the fraction of thermodynamic temperature ($T_0$), pressure under normal conditions ($p_0$), temperature (T), and pressure (p) in the aerosol. [19], [20]

$$\delta = \frac{p \cdot T_0}{p_0 \cdot T} \tag{2}$$

The onset of the effective corona discharge in an electrical separator is determined by the voltage that must be applied to the precipitator for even a very small current to begin to pass through it - this voltage is referred to as the initial *critical voltage* $U_0$ (equation 3).

$$U_0 = E_0 \cdot r_{char} \cdot ln\left(\frac{R_{dep}}{r_{char}}\right) \tag{3}$$

With increasing $R_{dep}$ (radius of the deposition electrode), a corona discharge occurs at a higher voltage; on the other hand, as $R_{dep}$ decreases, the risk of unintentional spark charge increases. Voltage conditions which limit corona at higher voltages (thanks to unwanted spark charge) are defined by equation (4). [19], [20]

$$U_s = k_1 ln\left(\frac{R_{dep}}{r_{char}}\right)\left(0.1 \cdot \delta \cdot R_{dep} \cdot k_2 \cdot \sqrt{0.1 \cdot \delta \cdot R_{dep}}\right) \tag{4}$$

The following Fig. 3 compares the spark and corona discharge areas. From the point of view of the design of the electrostatic precipitator, it is necessary to move in the corona discharge region, but with a voltage margin so that spark discharge does not occur unintentionally.

Particles smaller than 100 nm are charged mainly by diffusion and particles larger than 1 μm by polarization. Particles in the range from 100 nm to 1 μm are charged by diffusion as well as polarization - nevertheless, these particles are the most difficult to separate. The charge of particles $Q_P$ is defined by equation (5). [20]

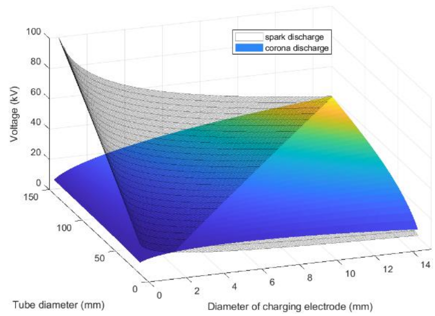



Fig. 3. Dependence of the voltage on the variables tube diameter and diameter of charging electrode for length of 0.4 m

$$Q_P = \left( \left(1 + 2 \cdot \frac{\lambda}{d_p}\right)^2 + \frac{2}{1 + 2 \cdot \frac{\lambda}{d_p}} \cdot \frac{\varepsilon_p - 1}{\varepsilon_p + 2} \right) \pi \cdot \varepsilon_0 d_p^2 E_{av} \tag{5}$$

Where $\lambda$ represents mean free path of the ions, $d_p$ particle size diameter, $\varepsilon_p$ relative permittivity of the separated particle, $\varepsilon_0$ permittivity of free space.

$E_{av}$ is the average electric field strength, which includes the difference in electric field strength in the electrostatic precipitator cylinder [19], [20] – see equation (6).

$$E_{av} = \frac{U}{R_{dep} - r_{char}} \tag{6}$$

The particle does not gain charge immediately - the charge increase is initially rapid, gradually slowing down and approaching asymptotically the value at saturation. The rate at which the particle gains charge is quite high - after 0.1s the particle gains more than 90% of the charge at saturation and more than 99% at 1 s. Hence, to sufficiently charge the particles, the particles need to remain in the active part of the electrostatic precipitator for at least 1 s as stated by Böhm [19].

To quantify the basic efficiency of the electrostatic precipitator ($\eta_p$), the calculation according to equation (7) was used. [19] [20]

$$\eta_p = 1 - e^{\frac{-w_p \cdot s_{us}}{V}} \tag{7}$$

Where $w_p$ represents speed of moving particles (drift speed) and $S_{us}$ electrode deposition area and V the volumetric flow rate of particles.

At the beginning of the synthesis, particles are deposited on clean electrodes, but as the synthesis time increases, a thicker layer of deposited particles is formed. For well-conducting particles, the particle charge is lost almost immediately from all locations on the particle surface, but for less conductive particles, the charge is lost only where the particle touches the electrode - a very slow loss of charge occurs. When a thick layer of trapped particles is formed, the charge removal of the particles becomes more difficult and there is an apparent specific resistance of the layer of synthesized particles and therefore a voltage drop. The voltage to be applied to the electrodes must then increase by this decrease.

During the ongoing deposition of particles on the collecting electrode (but also on the charging electrode), the particles are gradually removed spontaneously, but at a certain step it is necessary to proceed to a controlled cleaning of the electrodes so that further deposition of synthesized particles can occur again. This time is influenced not only by the time for which electrostatic precipitation takes place, but also by the surface area of the electrodes and the concentration of the solution. According to Böhm [19], electrode cleaning should occur after 60 to 140 minutes at the latest for smaller size separators. [19]

### *B. Design of Electrostatic Precipitator*

As pointed out in the previous section, the configuration of the electrostatic precipitator depends on several assumptions. However, an important factor to consider when assembling the separator is the actual location in the space, the availability of the individual components, and the cost.

The actual selection of the individual components of the electrostatic precipitator includes the choice of the diameter of the charging electrode, the diameter of the deposition electrode and the overall length. The selection of specific values must then follow the principles outlined in the previous section.

One of the most important conditions for separation is the formation of a corona discharge. At the same time, since the equipment will be in full operation in a vented hood, it is necessary to adapt the dimensions not only to commonly available materials, but also to use the most compact dimensions possible. For this reason, and also because of possible unwanted condensation, a maximum length of 0.4 m was adopted. The following Fig. 4 discusses the voltage dependence of the electrostatic precipitator and its effect on corona and spark charge formation for an electrostatic precipitator active length of 0.4 m and diameter of 9 cm. Fig. 5 then shows a comparison of the charging electrode diameter, voltage, and efficiency of the separation process. Since the highest efficiency is achieved by the separator with the largest charging electrode radius and at the highest voltage, it would seem that these are ideal design parameters. However, compared with Fig. 4, where the diameter of the charging electrode was compared with the critical voltage $U_0$ and the risk of an unwanted spark charge, it can be seen that the larger diameter of the charging electrode causes on the one hand the need for a higher voltage and at the same time reduces the voltage at which an unwanted spark charge is generated. That is why the larger diameter of charging electrode is unsuitable. The diameter of the tube is also an important parameter. As can be seen from Fig. 3, larger tube diameters have higher efficiency values. In agreement with this, Fig. 6 is plotted for a length of 0.4 m and a charging electrode width of 1 mm. It can be seen here that as the tube diameter increases, the difference between $U_0$ and the voltage at which spark discharges already occur increases.

## IV. CONCLUSION

In conclusion, the synthesis of submicron particles for cathode materials in Li-ion batteries demands careful consideration of manufacturing methods. Spray-drying synthesis offers promise, yet challenges persist in generating sufficient product quantities. To address these challenges, the apparatus comprising an ultrasonic atomizer, drying chamber, and modified electrostatic precipitator was designed. Through easy control of solution parameters and precise atomization,

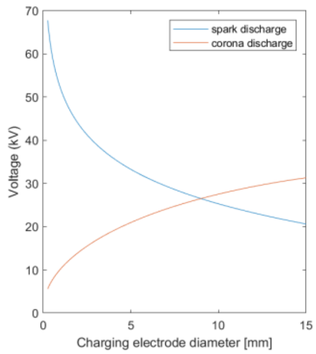



Fig. 4. Dependence of initial corona discharge and spark voltage on the diameter of the charging electrode.

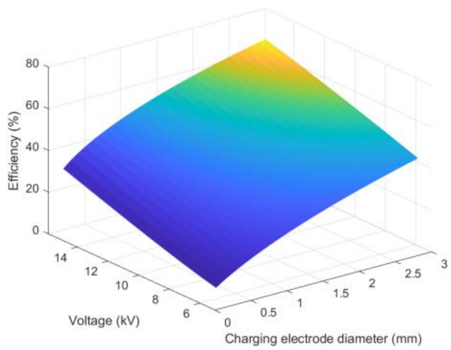



Fig. 5. Dependance of efficiency on the voltage and charging electrode diameter.

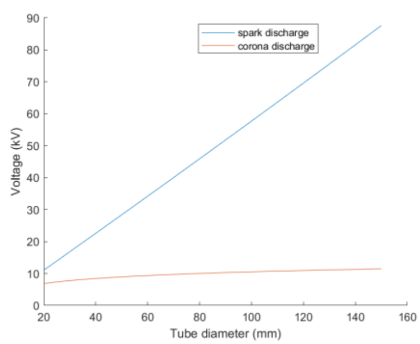



Fig. 6. Dependance of initial corona discharge and spark voltage on the diameter of the charging electrode.

spray-drying synthesis enables the production of homogeneous powder with small-sized particles with potential for reduced processing time and temperature.

Electrostatic precipitators play a crucial role in particle collection, with their design optimizing efficiency and facilitating scalability. The integration of spray-drying synthesis with tube electrostatic precipitator represents a promising approach for the efficient and scalable production of submicron particles for Li-ion battery materials.

## ACKNOWLEDGEMENT

This work was supported by the specific graduate research of the Brno University of Technology No. FEKT-S-23-8286.

## REFERENCES

[1] C. Lu and S. Lin, "Influence of the particle size on the electrochemical properties of lithium manganese oxide", *Journal of Power Sources*, vol. 97-98, pp. 458-460, 2001.

[2] B. Vertruyen, N. Eshraghi, C. Piffet, J. Bodart, A. Mahmoud and F. Boschini, "Spray-Drying of Electrode Materials for Lithium- and Sodium-Ion Batteries", *Materials*, vol. 11, no. 7, 2018.

[3] S. Yang, J. Park, J. Kim, M. Yoon, S. Wang, D. Jung and Y. Kang, "Spray-assisted synthesis of ant-cave-structured Ni-rich cathode microspheres with Li-reactive coating layer for high-performance Li-ion batteries", *Materials Today Chemistry*, vol. 35, 2024.

[4] T. Wani and G. Suresh, "A comprehensive review of LiMnPO4 based cathode materials for lithium-ion batteries: current strategies to improve its performance", *Journal of Energy Storage*, vol. 44, 2021.

[5] I. Nwachukwu, A. Nwanya, A. Ekwealor and F. Ezema, "Research progress in solid-state synthesized LiMnPO4 cathode material for Li-ion battery applications", *Applied Surface Science Advances*, vol. 18, 2023.

[6] A. Stunda-Zujeva, Z. Irbe and L. Berzina-Cimdina, "Controlling the morphology of ceramic and composite powders obtained via spray drying – A review", *Ceramics International*, vol. 43, no. 15, pp. 11543-11551, 2017.

[7] P. Duvigneaud and T. Segato, "Synthesis and characterisation of LiNi1−x−yCoxAlyO2 cathodes for lithium-ion batteries by the PVA precursor method", *Journal of the European Ceramic Society*, vol. 24, no. 6, pp. 1375-1380, 2004.

[8] Z. Chen, Q. Zhao, M. Xu, L. Li, J. Duan and H. Zhu, "Electrochemical properties of self-assembled porous micro-spherical LiFePO4/PAS composite prepared by spray-drying method", *Electrochimica Acta*, vol. 186, pp. 117-124, 2015.

[9] R. Mei, Y. Yang, X. Song, Z. An and J. Zhang, "Triple carbon coated LiFePO4 composite with hierarchical conductive architecture as high-performance cathode for Li-ion batteries", *Electrochimica Acta*, vol. 153, pp. 523-530, 2015.

[10] Y. SUN and Y. XIA, "Preparation and electrochemical properties of LiCoO2–LiNi0.5Mn0.5O2–Li2MnO3 solid solutions with high Mn contents", *Electrochimica Acta*, vol. 51, no. 26, pp. 5581-5586, 2006.

[11] G. Dong, H. Liu, L. Zhou, L. Chong, J. Yang, Y. Qiao and D. Zhang, "Investigation of various synthetic conditions for large-scale synthesis and electrochemical properties of Li 3.98 Al 0.06 Ti 4.96 O 12 /C as anode material", *Journal of Alloys and Compounds*, vol. 615, pp. 817-824, 2014.

[12] S. Choi and Y. Kang, "Kilogram-Scale Production of SnO 2 Yolk–Shell Powders by a Spray-Drying Process Using Dextrin as Carbon Source and Drying Additive", *Chemistry – A European Journal*, vol. 20, no. 19, pp. 5835-5839, 2014.

[13] J. O'Sullivan, E. Norwood, J. O'Mahony and A. Kelly, "Atomisation technologies used in spray drying in the dairy industry: A review", *Journal of Food Engineering*, vol. 243, pp. 57-69, 2019.

[14] A. Nandiyanto and K. Okuyama, "Progress in developing spray-drying methods for the production of controlled morphology particles: From the nanometer to submicrometer size ranges", *Advanced Powder Technology*, vol. 22, no. 1, pp. 1-19, 2011.

[15] R. Vehring, W. Foss and D. Lechuga-Ballesteros, "Particle formation in spray drying", *Journal of Aerosol Science*, vol. 38, no. 7, pp. 728-746, 2007.

[16] M. Seenivasan, J. Jeyakumar, R. Jose and C. Yang, "Spray-drying Al onto hydroxide precursors to prepare LiNi0.855Co0.095Al0.05O2 as a highly stable cathode for lithium-ion batteries", *Journal of Alloys and Compounds*, vol. 926, 2022.

[17] M. Ihalainen, M. Kortelainen, A. Mesceriakovas, T. Karhunen, S. Mesceriakove, D. Lindberg, J. Leskinen, T. Kankaanpää, J. Jokiniemi and A. Lähde, "Synthesis of solid NMC622 particles by spray drying, post-annealing and lithiation", *Advanced Powder Technology*, vol. 34, no. 10, 2023.

[18] Y. James Li, W. Chien, C. Tang, S. Wu, Y. Wu and C. Yang, "Electrochemical performance of spherical Li-rich LMNCO cathode materials prepared using a two-step spray-drying method", *Ceramics International*, vol. 48, no. 5, pp. 6302-6312, 2022.

[19] J. Böhm, *elektrické odlučování a odlučovače*, 1. Praha: Státní nakladatelství technické literatury, 1959.

[20] S. Kim, H. Park and K. Lee, "Theoretical model of electrostatic precipitator performance for collecting polydisperse particles", *Journal of Electrostatics*, vol. 50, no. 3, pp. 177-190, 2001.

# IIC to Modbus RS-485 converter in an industrial measurement system

1st Albert Mlčoch
*Department of Control and Instrumentation*
*Brno University of Technology*
Brno, Czech Republic
Albert.Mlcoch@vutbr.cz

***Abstract*—The design and application of a Modbus-IIC to Modbus RS-485 converter are presented in this paper. The system aims to gather data from numerous slow-sampling sensors. The proposed converter is based on the STM32 platform. Its design and functionality are showcased within a measurement setup involving SHT35 temperature and relative humidity sensors. Practical tests involved connecting and operating 13 sensors equipped with various types of IIC/RS-485 converters, including three instances of the proposed system, simultaneously within the setup. The designed hardware demonstrates functionality in laboratory settings.**

***Index Terms*—measurement system, IIC, MODBUS, RS-485, bus converter, sampling rate**

## I. Introduction

Nowadays commonly used Micro Electro-Mechanic Sensors (MEMS) use simple serial buses for communication and data transfer. The standard in the MEMS industry is communication via Inter-Integrated Circuit (IIC) or Serial Peripheral Interface (SPI). Often, the sensor allows communication via both IIC and SPI. Both serial buses are advantageous because of the small number of wires (2 wires for IIC and 3 wires for SPI) and the relatively wide range of data rates (from kbit/s to Mbit/s). The hardware required to operate the buses is very cheap and available. Bus handling can be implemented in software in virtually any microcontroller. Most microcontrollers support both buses at the hardware level.

The disadvantage of these buses is their sensitivity to noise and the resulting low bus range. Sensitivity to noise is due to the way the data are transmitted. Both buses differentiate logic signal levels based on the magnitude of the voltage on the conductor relative to the common ground. The maximum range is in metres at reduced communication speed. In industrial deployments, the range is only tens of centimetres.

Collecting data from many measurement points over a distance of tens of metres requires a more robust data network. Signals need to be converted from simple serial buses to one of the standard industrial buses, which are more resistant to interference and have a much longer range. Many communication standards are used in industry, such as CAN, Ethernet, Profibus, Modbus RTU, and others.

The Modbus protocol [1] is an open communication protocol. It has the advantage of simplicity and the possibility of choosing basically any physical interface. In practice, Modbus is used on the physical layers RS-232, RS-422, RS-485, or Ethernet (in the Modbus-TCP version). This paper demonstrates the design and use of an IIC to Modbus RS-485 converter. The use of the RS-485 interface requires the addition of a differential pair transceiver to the HW, similar to other differential buses (e.g. CAN).

## II. Description of the measuring system

The measurement system is designed to collect data from a relatively large number of slow-sampling sensors. The maximum number of sensors in the system corresponds to the maximum number of slave devices on the Modbus, that is, 247 connected devices. In practical tests, 13 sensors with IIC/RS-485 converters of different types, including three pieces of the proposed converter, were simultaneously connected and operated in the system.

In the experimental setup, the length of the line from the master to the slaves group is L = 20 m. All slaves are about 10 cm apart. The topology corresponds to Fig. 1. The maximum sampling rate $f_{s\,max}$ [S/s] depends on the maximum modulation rate. The modulation rate is approximately inversely proportional to the line length. The maximum bus speed corresponds to the specification of the physical layer of RS-485 [2]. Based on knowledge of the design of multichannel systems [3] and the parameters of RS-485 [4], the maximum sampling rate can be estimated from the relation:

$$f_{s\,max} \approx \frac{k}{L \cdot N \cdot D_{max}} \tag{1}$$

where: $k \approx 1.30 \cdot 10^8\ ms^{-1}$ is a defined or empirically determined transmission parameter, its value can be determined from the physical layer specification, $L$ is the line length, according to RS-485 specification maximum 1200 m, $N$ is the number of sensors in the system, $D_{max}$ is the length of the sent message in bits.

The theoretical minimum message length is 4 bytes - the device address, one data byte, and two Cyclic Redundancy Check (CRC) bytes. In practice, the shortest message is an error message of 5 bytes - device address, error instruction number, error number, and 2 bytes of CRC. The longest data message is defined by the Modbus protocol to be 256 bytes. Due to the nature of the application, such long messages are not used. The longest message is 21 bytes long and contains

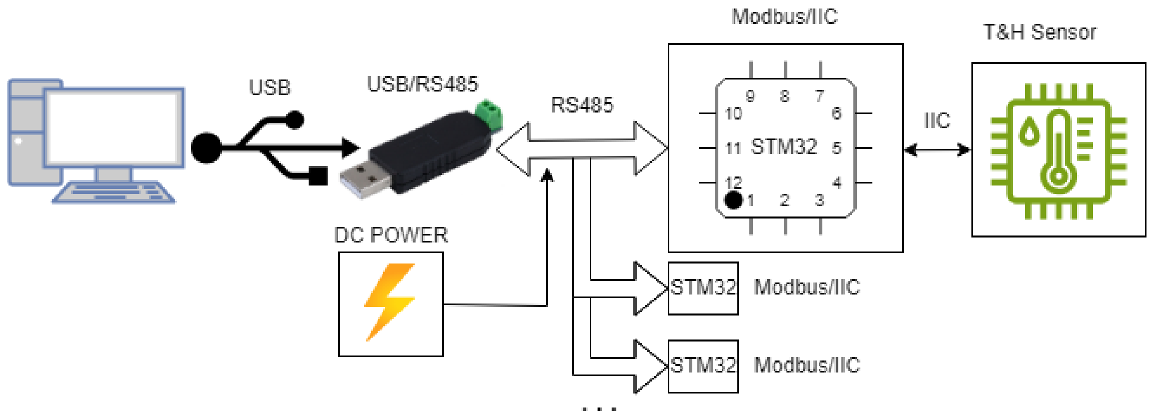



Fig. 1. Block diagram of the measuring system based on RS-485

| N [-] / L [m] | 10 | 50 | 100 | 500 | 1000 | 1200 |
|---|---|---|---|---|---|---|
| 5 | 9000 | 1800 | 900 | 180 | 90 | 75 |
| 10 | 4500 | 900 | 450 | 90 | 45 | 37 |
| 20 | 2250 | 450 | 220 | 45 | 22 | 18 |
| 50 | 900 | 180 | 90 | 18 | 9.0 | 7.5 |
| 100 | 450 | 90 | 45 | 9.0 | 4.5 | 3.7 |
| 200 | 220 | 45 | 22 | 4.5 | 2.2 | 1.9 |

TABLE I
APPROXIMATE MAXIMUM SAMPLE RATE $f_{s\,max}$ [S/S] FOR DIFFERENT LINE LENGTHS AND NUMBERS OF SENSORS ON RS-485, N [-] IS THE NUMBER OF SENSORS, L [M] IS THE LINE LENGTH, D = 288 [BIT].

the measured data. The message is in the format of device address, executed instruction number, length of data sent, maximum 16 data bytes and 2 bytes CRC.

The most common type of message is the periodic sending of all measured data. Each data send is preceded by a prompt from the master. The prompt is 8 bytes long. In addition to the length of each message, a gap of the duration of the send $D_{end} \geqq 3.5$ bytes must be included to represent the end of the message. The length of common communication for requesting sensor data can be defined as:

$$D \geqq D_{byte} \cdot (2 \cdot D_{end} + D_{master\ call} + D_{slave\ resp}) \quad (2)$$

Lengths of byte $D_{byte}$ consider 8 bit, the length of the end of the message consider $D_{end} \geqq 3.5$ bytes, the call to send data $D_{master\ call}$ is always 8 bytes long. The response length of device $D_{slave\ resp}$ is maximum of 21 bytes. For the maximum length of the data message, the communication length is based on eqv. 2 $D \geqq 288$ [bit]. Communication time is determined by recalculating the bit-rate. The communication time may be longer than the calculated time depending on the actual $D_{end}$ time and the response processing time of the microcontroller. Examples of approximate sample rates as a function of the number of sensors in the network and the length of the network are shown in Table I.

As mentioned above, the system is designed primarily to handle a large number of relatively slow sampling sensors. These include ambient temperature, humidity, illuminance, atmospheric pressure, surface elevation, and others. Typical

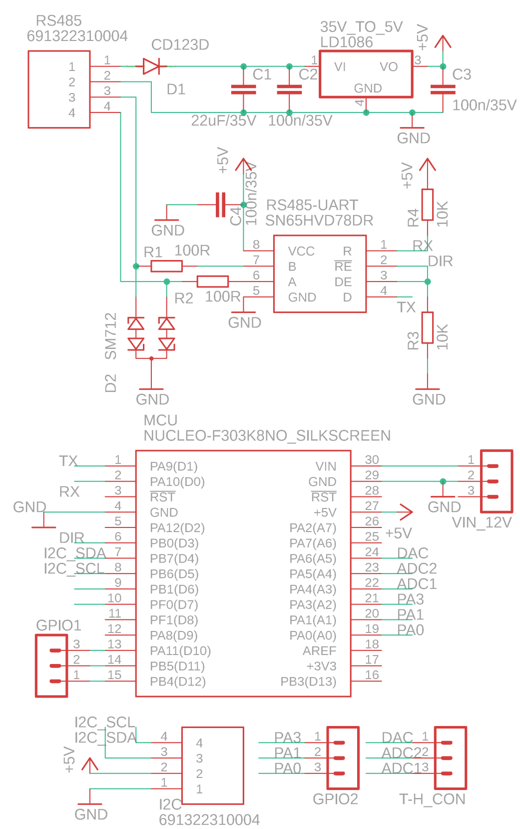


Fig. 2. Modbus-I2C converter sheme.

sampling periods for these quantities range from seconds to tens of minutes. Therefore, the speed of communication over the IIC bus is not critical. More important parameter of the IIC is the reliability of the communication. Therefore, it is possible to set the communication speed below the standard IIC speed of 100 kbit/s, and thus improve the reliability and range of the bus. Speeds of, e.g. 10 or 20 kbit/s can be used.

A block diagram of the whole system can be seen in Fig. 1. The control computer communicates through the USB interface with a USB to RS-485 converter. The RS-485 is a four-wire bus. It contains one pair of twisted wires for data transmission and two power wires. RS-485 is powered by an external $\pm 12$ V power supply. For laboratory conditions, this is the GW INSTEK GPD-3303 power supply. The power supply is used to power both the bus and the connected sensors. Between the bus and the sensor itself, a converter is designed with an STM32F303 microcontroller on the Nucleo-32 development board.

The optimal bus topology is a single long branch. To avoid reflections on the line, it is necessary to minimise the length of the branch lines from the line to the individual converters. For this reason, the converters are not designed as lateral branches from the main bus branch, but are ”right on the bus”.

## III. Used Hardware

A printed circuit board (PCB) was designed and manufactured to serve as a plug-in module to the Nucleo-32 development board. The scheme of the board is shown in Fig. 2. The PCB provides conversion from 12 V bus supply voltage to 5 V for powering the development board, powering the SN65HVD78 [5] converter and powering the IIC bus with attached sensor. The PCB contains integrated overvoltage protection on signal wires A and B and overcurrent and power supply reverse polarity protection. General-purpose input/output pins (GPIO) and power supply pins are routed from the PCB for future expansion of the module.

RS-485 is converted to the microcontroller UART using the SN65HVD78 converter. The SN65HVD78 allows communication speeds up to 50 MBd. The converter speed exceeds the maximum speeds used on the RS-485 bus.

## IV. Software solutions

Simultaneous operation of both peripherals, UART and IIC, is solved by two state machines, see Fig 3. Both state machines are called within one super-loop of the program. In one pass through the superloop, one state from each state machine is always executed. Measured data are regularly updated; the last measured value is always ready for immediate sending. Data are stored in two formats, 4 byte float and 2 byte unsigned int. Unsigned int values are multiplied by 10, thus the resolution is 0.1. The user selects which values should be read.

In the Decode message state, the values of the measured variables are overwritten in the structure of the message to be sent, preventing them from being overwritten by new measurements during message creation. Transmission errors are detected on the Modbus side by means of a 16-bit CRC (CRC-16); on the IIC side the data are checked by means of an 8-bit CRC (CRC-8).

Modbus messages are written to the hardware buffer of the UART interface. By means of interrupts, the individual bytes from the UART buffer are overwritten in the data structure of the Modbus device. The end of the message is determined by the minimum silent interval on the bus. The minimum interval is set by the system timer, there is no need to use interrupts. As soon as the received message has the appropriate length, a new message is acknowledged, and the Modbus state changes from Idle to Decode message. The message is decoded and CRC verified. A response message is immediately constructed, checking and sending the response follows in the next super-loop cycle in the Transmit state. The message and all associated flags are reset before transitioning back to the Idle state.

The state machine that controls the communication with the sensor is controlled by a timer. The STM32 microcontroller

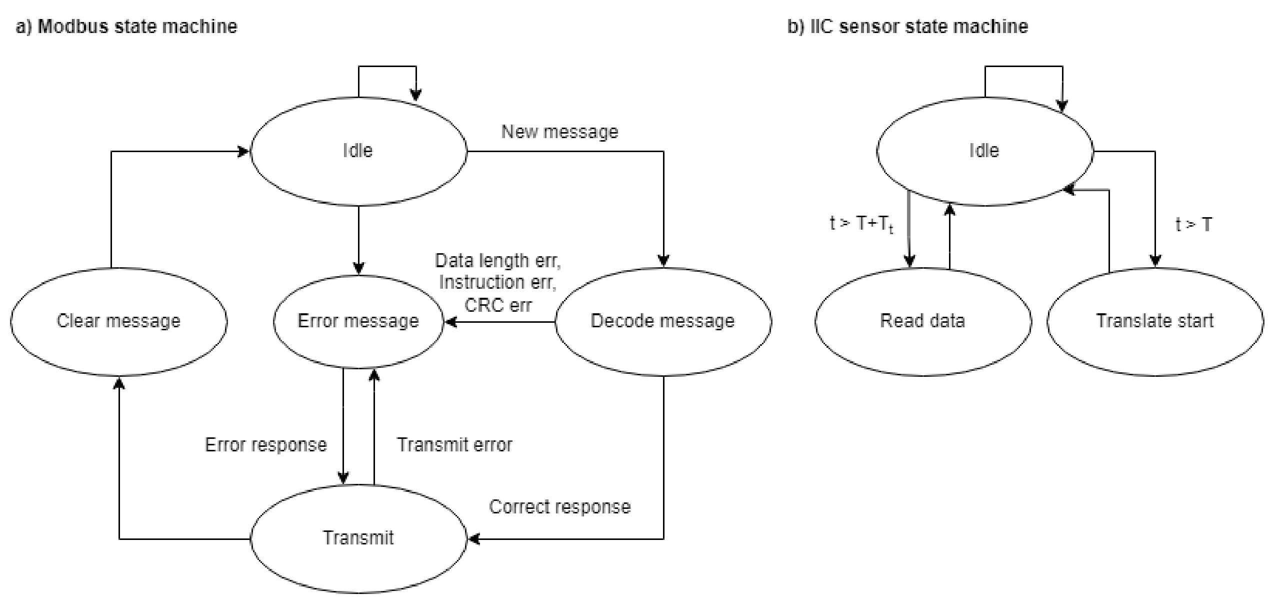



Fig. 3. Pseudoparalel state machines, a) Modbus state machine, b) IIC sensor state machine. $T$ - measurement period, $T_t$ - data translate time.

operates in master mode. After the set sampling time $T$, a command is sent to the sensor to start the conversion of the desired quantity, and the state machine returns to the Idle state. After the translation time $T_t$ has elapsed, the data are read out, checked with CRC-8, and transferred to the Modbus data structure. The state returns to Idle. In case of an incorrect read value, the error value is written to the Modbus structure. If there is no valid value at the time of the data request over Modbus, the Modbus device sends an error message of type 'invalid data'.

## V. Demonstration of data collection

The proposed Modbus-IIC converter has been tested in a measurement chain for SHT35 temperature and relative humidity sensors. Three transmitters were connected to the system. The data logger in the PC application collected temperature and relative humidity data from all connected transducers with a reading period of 1 minute for 3 weeks. The resolution of the stored data is 0.1 K a 0.1 % relative humidity. The PC application is a third party product and is not the subject of this work. In principle, it is possible to use any application that communicates over a serial link and uses the Modbus protocol.

Fig. 4 shows a 21-hour data segment. The sensors were stored freely in the laboratory. From the measured data, the window opening times at 17:35 and 7:51 are evident, manifested by a sharp drop in temperature and humidity due to the influx of dry air from outside. The data also show a gradual decrease in temperature and humidity at night by 60 mK/hour and 0.1 %/hour. The change in relative humidity in a room correlates with the number of people in the room. For example, it is possible to detect the time when workers leave for lunch. Slight differences in the humidity values of individual sensors may be due to offset.

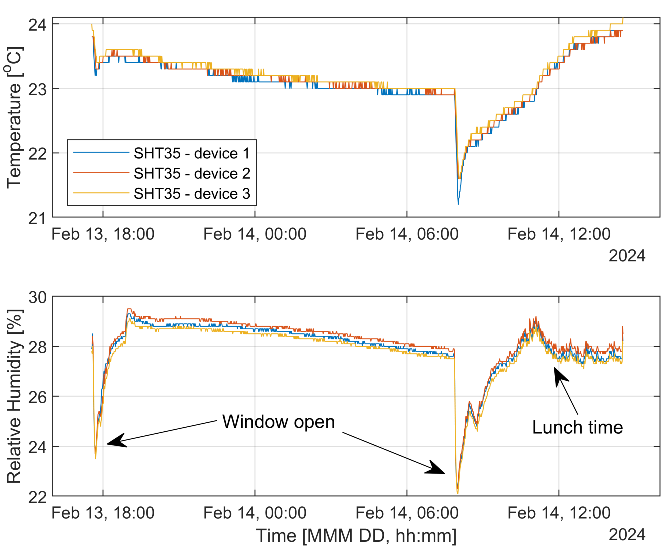



Fig. 4. Test measurement data.

## VI. Conclusion

The designed hardware is functional under laboratory conditions. The software solution is currently being extended to be usable for different IIC sensors. The as yet unused pins on the PCB will be used in the future for functions such as hardware switching of IIC sensor addresses, heating switching for humidity sensors, or transducer outputs for ease of user use, such as control LEDs.

## Acknowledgment

The completion of this paper was made possible by the grant No. FEKT-S-23-8451 - "Research on advanced methods and technologies in cybernetics, robotics, artificial intelligence, automation, and measurement" financially supported by the Internal science fund of Brno University of Technology.

## References

[1] *Modbus Protocol Reference Guide, Rev. J*, MODICON, Inc., Industrial Automation Systems, North Andover, Massachusetts 01845, 1996.

[2] T. Kugelstadt, "The rs-485 design guide," Texas Instruments, Tech. Rep. SLLA272D, 2021.

[3] M. Di Paolo Emilio, *Data Acquisition Systems: From Fundamentals to Applied Design*. New York, NY: Springer New York, 2013. [Online]. Available: https://link.springer.com/10.1007/978-1-4614-4214-1

[4] C. Kinnaird, "Use receiver equalization to extend rs-485 data communications," Texas Instruments, Tech. Rep. SLLA169, 2004.

[5] Texas Instruments, *SN65HVD7x 3.3-VSupply RS-485 With IEC ESD protection*, SN65HVD78, MARCH 2019. [Online]. Available: https://www.ti.com/product/SN65HVD78?dcmp=dsproject&hqs=td&#doctype2

# Image Reconstruction in Electrical Impedance Tomography through Multilayer Perceptron

Serge Ayme Kouakouo Nomvussi[1], Jan Mikulka[2]
*Department of Theoretical and Experimental Electrical Engineering Brno University of Technology*
Brno, Czech Republic
[1]244245@vut.cz, [2]mikulka@vut.cz

***Abstract*—This study introduces a novel image reconstruction algorithm designed to excel in challenging scenarios with noisy datasets. Comparative evaluations against established methods, the Total Variation technique and the Gauss-Newton algorithm, are conducted using key performance metrics including the correlation coefficient and structural similarity index. The Results demonstrate that the proposed algorithm displays variable performance in noise-free data compared to Total Variation but consistently outperforms it in the presence of noise. Furthermore, when contrasted with the Gauss-Newton algorithm, the proposed method consistently exhibits superior outcomes, particularly in scenarios involving noisy datasets, where the Gauss-Newton algorithm faces limitations. This study underscores the robustness of the proposed algorithm in noisy conditions, suggesting its potential for applications where accurate image reconstruction is critical.**

***Keywords—Multilayer Perceptron, Total Variation, Newton-Gauss, EIT.***

## I. Introduction

The field of Electrical Impedance Tomography (EIT) has emerged as a powerful non-invasive imaging modality, providing valuable insights into the internal conductivity distribution of biological tissues [1-3]. The pursuit of precise and efficient image reconstruction in EIT has led to the exploration of advanced computational techniques. This study forefronts this endeavor, focusing on the application of the Multilayer Perceptron (MLP) algorithm within the realm of EIT [4]. Traditional approaches [5 -9] to image reconstruction in EIT face challenges, particularly the ill-posedness of the inverse problem. In recent years, deep learning methodologies [10], including the Multilayer Perceptron neural network, have shown promise in capturing intricate relationships between measured voltages and internal conductivity distributions [11].

The primary objective of our research is to assess and elucidate the capabilities of the Multilayer Perceptron in the context of EIT image reconstruction. Leveraging the MLP's ability to learn complex patterns from data, our aim is to enhance the accuracy and efficiency of reconstructions. Throughout the study, we conduct a comparative analysis, pitting the performance of the Multilayer Perceptron against established methods such as Total Variation and the Gauss-Newton algorithm [12]. This comprehensive evaluation sheds light on the potential advantages and limitations of the MLP in comparison to traditional techniques.

The subsequent sections provide a detailed overview of the principles guiding Electrical Impedance Tomography, delve into the fundamental workings of the Multilayer Perceptron, and present a thorough analysis of our experimental results. This exploration at the intersection of deep learning and EIT holds promise for advancing the state-of-the-art in medical imaging, particularly in enhancing the accuracy and reliability of diagnostic tools. Additionally, the study introduces a novel algorithm for image reconstruction, emphasizing its efficacy, especially in the challenging context of noisy datasets. By employing established methods for comparative analysis, this research contributes valuable insights, paving the way for advancements in robust image reconstruction techniques, with potential applications across diverse domains.

## II. Methods

For this research, a synthetic dataset was generated using the EIDORS framework, a MATLAB-based tool designed for solving forward and inverse modeling in EIT. The dataset creation script comprises a main program that invokes three distinct functions, each defining and modifying relevant parameters for individual samples. EIDORS, requiring only a MATLAB installation, facilitates the simulation of EIT scenarios for diverse conductivity distributions.

The dataset comprises synthetic samples featuring circular and rectangular targets of varying conductivity and sizes strategically placed within a water-filled container, while the experimental setup, depicted in Fig. 1, entails a circular bucket surrounded by 16 electrodes, along with components such as a multiplexer, a measurement unit, a data analysis unit, and an image reconstruction unit. The water in the bucket possesses a conductivity of 0.04 S/m, whereas circular and rectangular objects inside exhibit conductivities ranging from 0.01 S/m to 0.12 S/m. The bucket, shown in black in Fig. 2, has a diameter of 100 mm, with circular brown objects spanning diameters from 6 to 10 mm, and rectangular targets measuring 5 mm in length with widths ranging from 2 to 5 mm. Electrodes surrounding the bucket facilitate current injection and voltage measurement, adhering to specific injection and measurement patterns, including adjacent injection and voltage measurement for the primary case and injection and measurement separated by two electrodes for the secondary scenario. Presumably, the objective of this setup is to perform electrical impedance tomography (EIT) for reconstructing images of the internal conductivity distribution within the container, including the circular and rectangular targets, offering a platform for validating and testing EIT algorithms or methodologies using synthetic samples with known conductivity distributions. The Finite Element Method (FEM) relies on a homogeneous cylindrical object 'd2c' comprising 1024 pixels arranged with a pattern of 16 complete electrodes. Each sample in the dataset features either a circular or rectangular target, or a combination of both shapes. The simulated data include an injected current of 1 mA at 20 kHz, resulting in measured voltage ranging from 0.43 to 28.26 mV. The reconstructed image's conductivity values fall between 0.04 and 0.29 S/m. The dataset generated with EIDORS frameworks comprises a total of 17745 samples.

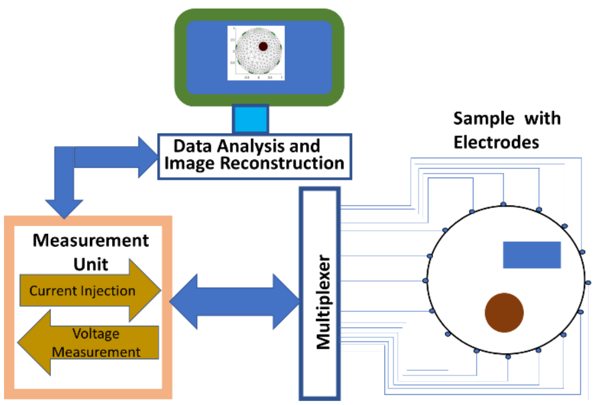



Fig. 1 EIT system

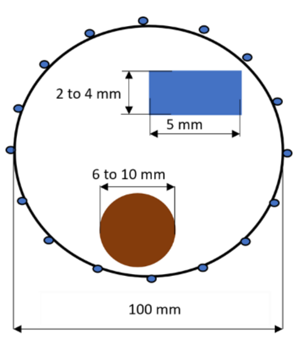



Fig. 2 Bucket and target size

To simulate realistic conditions, white Gaussian noise at varying levels is introduced during the voltage measurement process. The resulting dataset reflects a range of signal-to-noise ratios from 30 to 60 dB, providing a comprehensive basis for evaluating the robustness and accuracy of image reconstruction methods in the context of EIT.

The MLP architecture selected for this study, as illustrated in Fig. 3, is designed to address the complex challenge of reconstructing conductivity distributions in EIT. The input layer of the MLP aligns with the measured input voltage, as defined in Equation (1). To effectively capture intricate relationships within the data, multiple hidden layers can seamlessly integrate into the architecture. The exploration of the optimal number of neurons and layers in the hidden section reflects a meticulous approach to enhance the model's complexity and promote its ability to generalize patterns adeptly. It is important to mention that the conductivity of the original image is described in the Equation 2. The output layer boasts 1024 neurons, each actively contributing to the intricate task of reconstructing conductivity distributions. This architectural configuration meticulously balances model complexity and generalization, ensuring the MLP's efficacy in faithfully representing the nuanced conductivity variations observed within the imaged subjects.

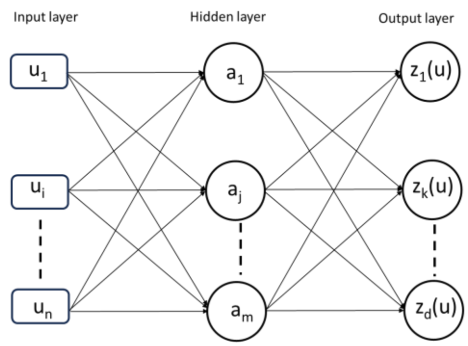



Fig. 3 Architecture of the new method

$$U = (U_1, \quad U_2, \dots, U_i) \qquad with\ i = 1,2, \dots 208, \quad (1)$$
$$\sigma = (\sigma_1, \quad \sigma_2, \dots, \sigma_j) \qquad with\ j = 1,2, \dots 1024. \quad (2)$$

The input layer is composed of neurons that serve as representatives for the characteristics inherent in the input data. Specifically, it encompasses 208 values, each corresponding to the measured input voltages associated with every sample during the forward problem-solving phase.

The Signal-to-Noise Ratio (SNR) intentionally added to the measured voltage is determined using the following equation:

$$SNR = 20\log_{10}\left(\frac{U}{V_n}\right), \qquad (3)$$

where $U$ is the boundary voltage without noise and $V_n$ is the noise voltage.

The equations for MLP are well known and the mean square error (MSE) of the proposed algorithm is computing using equation (4) where $T_i$ represents the target value, $Y_i$ corresponds to the network prediction, $M$ denotes the total number of responses in $Y$ and $N$ signifies the total number of observations in $Y$.

$$MSE = \frac{1}{2N}\sum_{i=1}^{M}(Y_i - T_i); \quad i = 1 \ldots 1024 \qquad (4)$$

The training process of a Multilayer Perceptron (MLP) begins by splitting the dataset into training, validation, and test sets. The model architecture, consisting of an input layer with 208 neurons, a hidden layer with variable neurons to be determined through training, and an output layer with 1024 neurons, is defined. Weights and biases are initialized randomly, and forward propagation is executed to generate predictions. The network training function updates weight and bias values using the scaled conjugate gradient method. This iterative process continues over the training dataset until convergence, with monitoring of validation set performance to mitigate overfitting. Simultaneously, the model's generalization ability is assessed through evaluation on both the test and validation datasets, enabling adjustments based on performance metrics. L2-Regularization is employed to enhance the model's robustness and performance. To reconstruct the image with the output vector predicted by the MLP, EIDORS frameworks is used.

To mitigate overfitting in the model, a penalty is introduced into the loss function, which is proportional to the square of the magnitude of the weights $W_j$ in the network, as described in equation (5). Mathematically, the regularization term L2 is expressed as the sum of the squares of all the weights in the network, multiplied by a regularization parameter, usually denoted $\lambda$.

$$Cost = MSE + \lambda\sum_{j} W_j^2 ; \ j = 1..number\ of\ weights. \ (5)$$

To assess the quality of the reconstructed image, a comprehensive set of metrics was employed, encompassing the correlation coefficient and structural similarity index (SSIM). These metrics played a pivotal role in quantitatively evaluating different facets of the reconstructed image, offering insights into its fidelity and overall quality.

The correlation coefficient (cc), as calculated in Equation (6), serves as a metric in assessing the degree of correlation between the recovered image $Y^*_{mn}$ and the original image $Y_{mn}$. It provides insights into the strength of the correlation, offering a quantitative measure of the quality of the reconstructed conductivity map. In Equation (6), $\overline{Y^*_{mn}}$ and $\overline{Y_{mn}}$ represent the mean values of the recovered and original images, respectively. A higher correlation coefficient means a stronger link, indicating a better quality of the reconstructed image.

$$\text{corr} = \frac{\sum_m \sum_n (Y^*_{mn} - \overline{Y^*_{mn}})(Y_{mn} - \overline{Y_{mn}})}{\sqrt{(\sum_m \sum_n (Y^*_{mn} - \overline{Y^*_{mn}})^2)(\sum_m \sum_n (Y_{mn} - \overline{Y_{mn}})^2)}}. \ (6)$$

The structural similarity index (SSIM), as utilized in Equation (7), serves as a metric for evaluating the quality of the reconstructed image. A SSIM value equal to 1 indicates complete identity between the original and reconstructed images. Conversely, a SSIM value less than 1 implies differences in quality between both images.

$$\text{ssim(x, y)} = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)}, \quad (7)$$

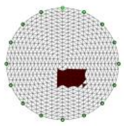
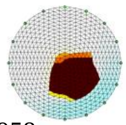
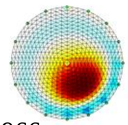
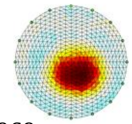
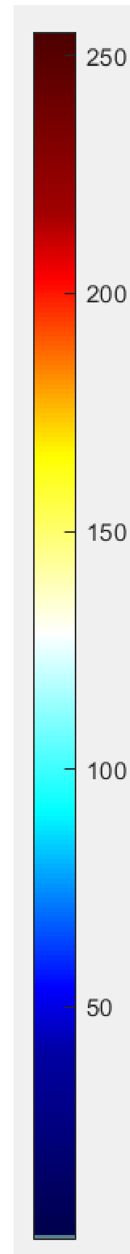
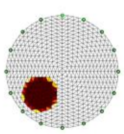
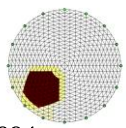
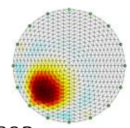
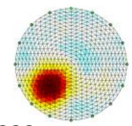
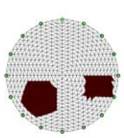
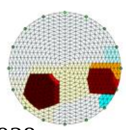
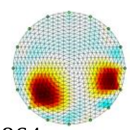
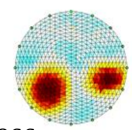
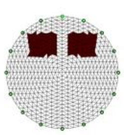
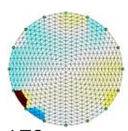
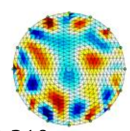
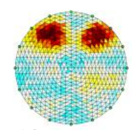
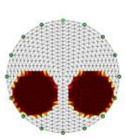
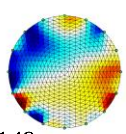
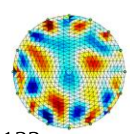
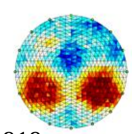
where $\mu_x, \mu_y$, $\sigma_x$, $\sigma_y$ and $\sigma_{xy}$ are the local means, standard deviations, and cross-covariance for images x, $y$. $C_1$ and $C_2$ are small constant which prevent division by zero.

## III. Results

The training process involved experimenting with different numbers of neurons for the hidden layer. After thorough testing, the configuration with 750 hidden neurons yielded the best validation mean square error of 0.0075. To mitigate the risk of overfitting, various regularization parameters were applied. Among these, the optimal parameter value identified was 0.001.

In Table I, a detailed comparison of the proposed method with the Total Variation and Gauss-Newton algorithm reveals interesting patterns. The correlation coefficient and structural similarity index exhibit varying trends across different scenarios. When analyzing noiseless data, the performance of the proposed algorithm compared to the Total Variation method varies; it sometimes surpasses the Total Variation method and sometimes falls short. However, when evaluating noisy datasets against the Gauss-Newton algorithm, the Structural Similarity Index consistently favors the proposed method. Moreover, in noise-free data, there are instances where the correlation coefficient of the Gauss-Newton algorithm outperforms that of the proposed method. However, overall, the comparative findings highlight the robustness of the new proposed algorithm, especially in scenarios involving noisy datasets, where it excels compared to established methods such as Total Variation and Gauss-Newton. It is noteworthy that, as mentioned at the end of Table I, the correlation coefficient and structural similarity index of the proposed method outperform those of the Total Variation and Gauss-Newton algorithm across the entire dataset, further affirming its superiority in various contexts.

TABLE I. COMPARISON RECONSTRUCTED IMAGES

| Model | Total Variation algorithm | Gauss-Newton algorithm | New method | |
|---|---|---|---|---|
|  |  cc = 0.958 ssim = 0.491 |  cc = 0.966 ssim = 0.507 |  cc= 0.969 ssim = 0.635 |   |
|  |  cc = 0.994 ssim = 0.999 |  cc= 0.893 ssim = 0.980 |  cc= 0.888 ssim = 0.984 | |
|  |  cc = 0.928 ssim = 0.987 |  cc = 0.864 ssim = 0.965 |  cc= 0.866 ssim = 0.962 | |
|  | 30 dB  cc = -0.179 ssim = 0.750 | 30 dB  cc = -0.210 ssim = 0.713 | 30 dB  cc = 0.713 ssim = 0.954 | |
|  | 60 dB  cc = 0.149 ssim = 0.455 | 60 dB  cc = 0.133 ssim = 0.659 | 60 dB  cc = 0.919 ssim = 0.893 | |
| cc and ssim of the dataset | cc = 0.549 ssim = 0.632 | cc = 0.487 ssim = 0.655 | cc = 0.865 ssim = 0.851 | |

In Fig. 4, correlation coefficients of the new method are presented for different datasets: 0.87653 for the Training dataset, 0.84031 for the Testing dataset, 0.86199 for the Validation dataset, and 0.86597 for the entire dataset. The negative center of the error histogram shown in Fig. 5 suggests that, on average, the predictions made by the neural network tend to be lower than the actual target values. In other words, the neural network is biased towards underestimating the targets.

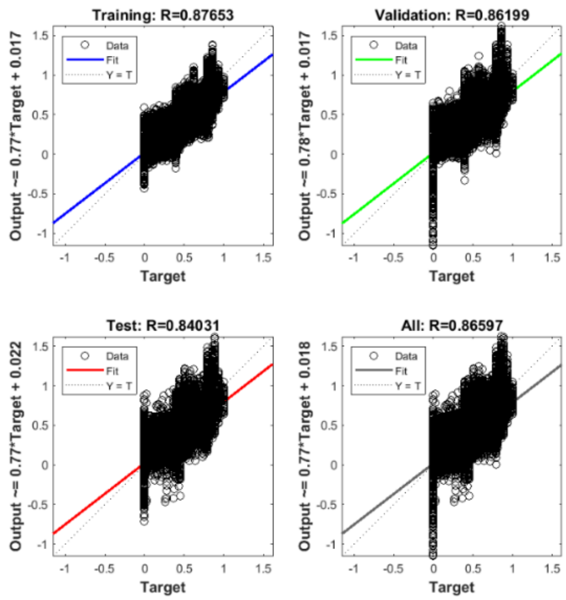



Fig. 4 Correlation coefficient of the dataset

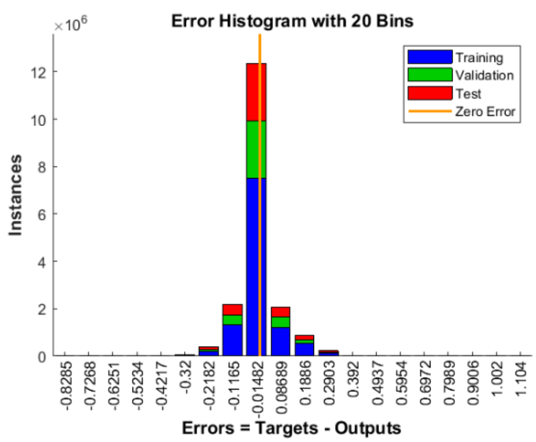



Fig. 4 Error Histogramm

## IV. Discussion

In this study, we propose a novel algorithm for image reconstruction, particularly focusing on its performance in the presence of noisy datasets. The algorithm is compared against two established methods: the Total Variation method and the Gauss-Newton algorithm. Performance metrics, including the correlation coefficient and structural similarity index are systematically evaluated across datasets with and without noise.

Our findings reveal that the new proposed algorithm exhibits variable performance compared to the Total Variation method for noise-free data, but consistently outperforms it in the presence of noise. When compared to the Gauss-Newton algorithm, the proposed algorithm consistently demonstrates superior results, especially in scenarios involving noisy datasets. Notably, the Gauss-Newton algorithm faces challenges in reconstructing images from noisy data. The both metrics used in this study are higher in the dataset for the new method than for the traditional methods.

The robust performance of the proposed algorithm, particularly in noisy conditions, holds promising implications for applications where accurate image reconstruction is crucial. This study contributes valuable insights into algorithmic advancements for image reconstruction and emphasizes the need for tailored solutions in challenging imaging scenarios. Future research directions may involve refining the proposed algorithm, exploring its applicability in diverse domains, and investigating the underlying mechanisms contributing to its robust performance in noisy conditions.

## Acknowledgment

The author, S. Kouakouo, extends gratitude to the Department of Theoretical and Experimental Electrical Engineering.

## References

[1] Harikumar, R., R. Prabu, and S. Raghavan. "Electrical impedance tomography (EIT) and its medical applications: a review." *Int. J. Soft Comput. Eng* 3.4 (2013): 193-198.

[2] Adler, Andy, and David Holder, eds. *Electrical impedance tomography: methods, history and applications*. CRC Press, 2021.

[3] Adler, Andy, and Robert Guardo. "Electrical impedance tomography: regularized imaging and contrast detection." *IEEE transactions on medical imaging* 15.2 (1996): 170-179.

[4] Rymarczyk, Tomasz, et al. "Optimising the use of Machine learning algorithms in electrical tomography of building Walls: Pixel oriented ensemble approach." *Measurement* 188 (2022): 110581.

[5] Farha, Maksalmina. "Combined algorithm of total variation and Gauss-newton for image reconstruction in two-dimensional Electrical Impedance Tomography (EIT)." *2017 International Seminar on Sensors, Instrumentation, Measurement and Metrology (ISSIMM)*. IEEE, 2017.

[6] Sun, Benyuan, et al. "An improved Tikhonov regularization method for lung cancer monitoring using electrical impedance tomography." *IEEE sensors journal* 19.8 (2019): 3049-3057.

[7] Zhao, Ruwen, et al. "A Second-Order Generalized Total Variation with Improved Alternating Direction Method of Multipliers Algorithm for Electrical Impedance Tomography Reconstruction." *Applied Sciences* 14.4 (2024): 1485.

[8] Liu, Shengheng, Jiabin Jia, Yimin D. Zhang, and Yunjie Yang. 'Image Reconstruction in Electrical Impedance Tomography Based on Structure-Aware Sparse Bayesian Learning'. *IEEE Transactions on Medical Imaging* 37, no. 9 (September 2018): 2090–2102. https://doi.org/10.1109/TMI.2018.2816739.

[9] Rudin, Leonid I., Stanley Osher, and Emad Fatemi. "Nonlinear total variation based noise removal algorithms." *Physica D: nonlinear phenomena* 60.1-4 (1992): 259-268.

[10] Zeng, Shihao, et al. "Deep learning based reconstruction enables high-resolution electrical impedance tomography for lung function assessment." *2023 45th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*. IEEE, 2023.

[11] Huuhtanen, Timo, and Alexander Jung. "Anomaly location detection with electrical impedance tomography using multilayer perceptrons." 2020 IEEE 30th International Workshop on Machine Learning for Signal Processing (MLSP). IEEE, 2020.

[12] Zhao, Ruwen, et al. "A Second-Order Generalized Total Variation with Improved Alternating Direction Method of Multipliers Algorithm for Electrical Impedance Tomography Reconstruction." *Applied Sciences* 14.4 (2024): 1485.

# Flexible nanofiber separator based on a zinc metal-organic framework

1st Tatiana Pisarenko
*Department of Physics*
*Faculty of Electrical Engineering and Communication*
*Brno University of Technology*
Brno, the Czech Republic
ORCID: 0000-0002-3527-7099

2nd Jana Holečková
*Institute of Physical and Applied Chemistry*
*Faculty of Chemistry*
*Brno University of Technology*
Brno, the Czech Republic

***Abstract*—In this paper, a flexible composite of polyvinylidene fluoride (PVDF) and zinc-based metal-organic frameworks (MOF) was successfully designed and synthesized. This material can serve as a separator or active filter to trap gases or heavy metals in wastewater such as uranium. The fabricated sample is in the form of a nanofiber mat with doped MOFs created by electrospinning. The results show successful incorporation of MOF nanoparticles in the fibers. Measurement was performed using scanning electron microscopy, Raman spectroscopy and X-ray photoelectron spectroscopy. Novel PVDF@Zn-MOF nanofibers is a promising material that can have a great impact in the field of water treatment and solve environmental questions in developing countries.**

***Index Terms*—zinc, PVDF, MOF, synthesis, electrospinning, Basolite Z1200, ZIF-8, $C_8H_{10}N_4Zn$**

## I. Introduction

Heavy metal contamination of water is a serious problem that threatens human health and the environment. Heavy metals such as mercury, lead, cadmium, uranium and arsenic are released into the water environment from a variety of sources, including industrial wastewater, mining, agriculture and fossil fuel combustion. Heavy metals do not decompose in the natural environment and can accumulate in the body. Its pollution also represents a significant threat to aquatic ecosystems [1]. Heavy metals can accumulate in fish and other aquatic organisms that are then consumed by humans. This can lead to chronic toxicity in humans and animals. Heavy metal pollution of water is therefore a global problem that requires urgent solutions. Where water contamination cannot be prevented, the possibility of filtering the water is an option [2]. It is therefore necessary to develop a material that can be able to provide filtration of such contaminated water under various challenging conditions. As a promising candidate, polyvinylidene fluoride doped with metal-organic structures appears to be a good solution [3].

Polyvinylidene difluoride, polyvinylidene fluoride, or simply PVDF, is a highly non-reactive thermoplastic and semi-crystalline fluoropolymer formed by the polymerization of vinylidene fluoride. Its chemical formula is $-(C_2H_2F_2)n-$. It exhibits several unique properties such as pyro-, tribo- or piezoelectric effect. It is very stable both thermally (melting point is $177\,^{\circ}\mathrm{C}$) and chemically (most halogens and acids) [4]. Its mechanical strength is also at a very good level. It is also biocompatible [4]. Hence, it can be used for a very wide range of applications such as medicine, filters, sensors, batteries and many more. Its disadvantages may be high cost, not easily recyclable and not biodegradable [5].

Metal-organic frameworks (MOFs) are hybrid materials that consist of metal ions and organic ligands. The metal ions are usually arranged in a lattice structure in which the ligands are coordinated to a central metal ion. MOFs are characterized by a large surface area, which can reach up to $7000\,\mathrm{m^2/g}$). This property makes MOFs ideal for applications where high adsorption capacity is important, such as gas storage and gas separation. MOFs are also porous materials, i.e. they contain pores of defined size and shape. The pores in MOFs are typically in the range of $0.5$ to $3\,\mathrm{nm}$ [6]. This porous structure allows MOFs to selectively adsorb gases and molecules based on their size and shape. The properties of the MOF can be tailored by selecting metal ions and ligands. This allows MOFs to be designed for specific applications. MOFs are thus a promising material for the adsorption of heavy metals from aqueous solutions as they offer several advantages over traditional materials [6]:

- **High absorption capacity:** the MOFs can absorb large amounts of heavy metals from aqueous solutions.
- **Selectivity:** MOFs can be adjusted to selectively absorb specific heavy metals.
- **Regenerability:** MOFs can be regenerated and reused after absorbing heavy metals.

MOF research is a relatively new field, and scientists are constantly trying to develop new combinations of this very promising material.

These two materials as a composite were chosen for their unique properties and especially their compatibility with each other. MOFs can thus occur as nanoparticles in PVDF. For filtration, the problem of this composite serving as a membrane

Research described in the paper was financially supported by the Ministry of Education, Youth and Sports of the Czech Republic under the project CEITEC 2020 (LQ1601), by the Internal Grant Agency of Brno University of Technology, grant No. FEKT-S-23-8228 and grant No. FCH/FEKT-J-24-8551. Part of the work was carried out with the support of CEITEC Nano Research Infrastructure supported by MEYS CR (LM2018110).

must then be solved. One possible solution is to create a polymer nanofibrous material, which can also achieve a very high flexibility.

For the development of a fine yet highly porous nanofibrous material, the electrostatic spinning method is very suitable. Electrostatic spinning, also known as electrospinning, is a process in which a strong electric field is applied to a polymer solution or melt to form very fine fibers in the tens or hundreds of nanometers. Filament formation occurs between two oppositely charged electrodes, one of which is in contact with the liquid, to which it transfers part of its charge. Due to the electrical forces, the surface tension of the solution is overcome, and the resulting jet stream (polymer solution) is further elongated and formed into fibers. The resulting fibers are finally accumulated as a cobweb on a substrate that is placed over a second electrode, usually grounded, or oppositely charged. This method is used for the production of nanofibers, which have a wide range of applications in various fields such as filtration, tissue engineering, electronics and others [7].

Therefore, this paper deals with the design of the precursor, the synthesis of the composite and the subsequent analysis of the nanofiber sample produced and a discussion on the applicability of the material fabricated in this way.

## II. Materials & Methods

This section is divided into two essential parts, namely the methods and materials used for the fabrication of the nanofiber composite as the first part and the methods and analyses used in the examination of the fabricated sample, where the success of the fabrication and the overall quality of the whole, relatively complex process, is examined. A fibrous mat was produced using electrostatic spinning and analysed using scanning electron microscopy (SEM), X-ray photoelectron spectroscopy (XPS) and Raman spectroscopy. All techniques used are non-destructive methods.

### *A. Solution preparation*

For the design and subsequent synthesis of the nanofiber membrane, polyvinylidene fluoride (Sigma Aldrich, St. Louis, USA) was chosen in the form of a pellets with a molar mass of $275.000\,\mathrm{g/mol}$. Dimethylformamide (DMF) and acetone (Ac) were used as solvent in a 7:3 ratio. A zinc-based MOF, otherwise known as 2-Methylimidazole zinc salt, was added to the solution. It has a zeolitic imidazolate framework (ZIF) constituted of metal ions and imidazolate anions. It is abbreviated as ZIF-8. It was a commercial product under the name Basolite® Z1200 (Sigma Aldrich, St. Louis, USA). The empirical formula for ZIF-8 is $C_8H_{10}N_4Zn$. The PVDF ratio was $15\,\%$ and the MOF ratio was $8\,\%$. The solution was mixed on an electric stirrer with $200\,\mathrm{rpm}$ at $80\,°\mathrm{C}$ for $24\,\mathrm{h}$. Temperature is particularly important here so that the solution does not begin to solidify before it is spun [8].

### *B. Fabrication of nanofibrous mat*

A nanofibrous mat was formed from the solution prepared as described in the previous section by electrospinning [9]. This is a critical step during fabrication because it is important to control several essential parameters that can affect the fabrication. The instrument used here was 4SPIN (Contipro, Brno, CZ). Electrospinning was performed in a 1 needle as emitter—solid collector cylinder configuration. This cylinder was covered with aluminum foil, which served as a substrate for the fibers that were deposited on it. The distance between the emitter and the collector was $20\,\mathrm{cm}$ and the high voltage was set to $50\,\mathrm{kV}$. Needle size of $17\,\mathrm{GA}$ was chosen as the optimum thickness. The solution flowed through the needle at $35\,\mathrm{\mu L/min}$. While the solution was being drawn onto the collector cylinder, the process of solvent evaporation was in progress and once the fiber reached the substrate, most of the solvent had already evaporated. In this way, several hundred nanometers thick fiber was wound onto the cylinder. The speed of the collector cylinder was set at $2000\,\mathrm{rpm}$ to control the nanofiber orientation, to control the thickness of the fibers, and to increase the β-phase of the piezoelectric PVDF. The nanofibrous mat was additionally dried to eliminate any solvent residue that may still be present in the material in the end of the fabrication process. The material was then hermetically sealed in a plastic bag. All set parameters are based on previous measurements and experience of successful PVDF nanofiber production.

### *C. Analytical instruments used for characterisation*

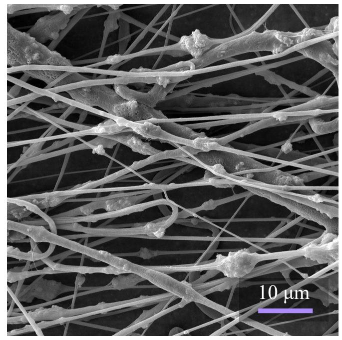
*1) Scanning electron microscopy (SEM):* The most fundamental element of the characterization was the scanning electron microscope, which allowed visual confirmation of the successful compatibility of PVDF and MOF and the state of the spun fibers. A Lyra3 microscope (Tescan, Brno, CZ) with an accelerating voltage of $10\,\mathrm{kV}$ was used. The focus of interest was on the formation and shape of the fibers and MOF. The study was performed both in a larger material width and within a single fiber. Since it is a polymer, a $20\,\mathrm{nm}$ gold layer was deposited on the material prior to its observation using an EM ACE600 coater (Leica, Wetzlar, DE). There were several reasons for this, the primary one is obviously for charge accumulation, however a thinner layer would have been sufficient for this. The problem with nanofibers tends to be that they are stimulated by an electron beam and can then move. Thus, carbon is quite successful in preventing both image drift and movement of the fibers themselves. The choice of gold for the coating instead of carbon was deliberate given that energy dispersive spectroscopy (EDS) and therefore elemental identification of the material was to be performed. In this case, it is not desirable to use carbon for the coating due to the fact that it would distort the resulting spectrum as PVDF is partly from carbon. The detector used was an X-Max 50 EDS (Oxford Instruments, Abingdon, UK). Elemental analysis was also performed with an accelerating voltage of $10\,\mathrm{kV}$. However, the plot is shortened on the x-axis to highlight the most important information occurring at lower energies. For the most accurate elemental ratios, information was collected from a $(500 \times 500)\,\mathrm{\mu m}^2$ area.

*2) Raman spectroscopy:* Another important method that has been chosen is Raman spectroscopy, which uses light scattering to determine the exact chemical fingerprint of the sample. More specifically, it determines the vibrational modes of molecules, their rotational states and other low frequency modes. It uses a monochromatic light source, a laser. Different materials may require different wavelengths of light source. In this work, a green laser with a wavelength of $532\,\mathrm{nm}$ has been chosen. The power value was set to $5\,\mathrm{mW}$, which was the optimum setting to obtain a strong enough signal and at the same time not to damage (burn) the sample, since it was a polymer. The integration time was $7\,\mathrm{s}$, the number of accumulations was $15\times$ and the lens magnification was set to $50\times$. The measurement was acquired from multiple locations to ensure consistency of results, as the MOF distribution may be less homogeneous for an essentially point-based method. The instrument used was confocal Raman imaging system alpha300 R (WITec, Ulm, DE).

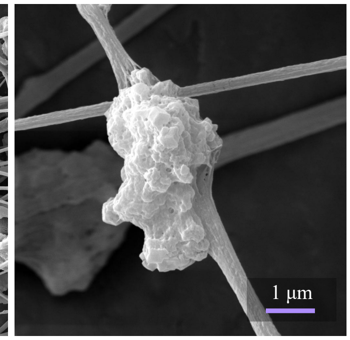
*3) X-ray photoelectron spectroscopy (XPS):* As the last instrument, the AXIS SupraTM X-ray photoelectron spectrometer (Kratos Analytical Ltd., Manchester, UK) was selected for the analysis of the accurate material composition and chemical state of the surface, as well as the overall electron structure and density of electron states. It uses the photoelectric effect where electrons are excited from the material after irradiation with X-rays. By measuring the kinetic energy of these emitted electrons, XPS provides valuable information about the atoms present and their chemical environment. The electron population spectra were identified by irradiating the material with a beam of X-rays. In this work, the wide spectrum in the range of $(1200\text{ to }0)\,\mathrm{eV}$ and the high-energy spectra of each element present in the material were obtained. These were the regions of carbon C1s on a range of $(294\text{ to }281)\,\mathrm{eV}$, oxygen O1s on a range of $(539\text{ to }527)\,\mathrm{eV}$, fluorine F1s on a range of $(692\text{ to }683)\,\mathrm{eV}$, elements commonly found in PVDF, and zinc Zn2p on a range of $(1029\text{ to }1017)\,\mathrm{eV}$, which represents the metal ion in the selected MOF. Both the wide spectrum and high-energy spectra were calibrated according to the carbon content to the $\mathrm{C}-\mathrm{C}$ chemical state at $284.8\,\mathrm{eV}$. For the high-energy spectra, background fitting and single chemical state fitting were performed. This method is capable of collecting data from the surface up to about $8\,\mathrm{nm}$. Thus, it is clear that in this case a sample with a gold-sputtered surface was not measured, as was the case in the II-C1 section.

## III. Results

First of all, it was necessary to determine whether the fibres were correctly spun, shaped and whether MOFs were present. This was based on similar previous successful fabrications from research earlier [7]. The excellent results were confirmed from the electron microscope images, where several fibers with many MOF dopants can be seen in Fig. 1a, resembling a knot on the fiber. It is also clear that these fibers are parallel aligned which is most influenced by the collector cylinder speed set at $2000\,\mathrm{rpm}$, which has already been mentioned in Sec. II-B. The speed of the cylinder and the dopant chosen has an effect on the width of the fibres, as has been shown in earlier publications [7]. However, their diameter, as can be seen, is inconsistent and ranges from hundreds of $\mathrm{nm}$ to units of $\mu\mathrm{m}$. Fig. 1b then shows a detail of one fiber in particular and the trapped dopant, where they can be seen to form cubic structures. It is evident that the width of the dopant is up to a multiple of the fiber diameter.




(a) Fibers




(b) MOF dopant

Fig. 1. Electron microscope images showing (a) the larger volume of the spun nanofibers and (b) a detail of the MOF structure successfully captured in the fiber.

Subsequently, energy dispersive spectroscopy was used to accumulate several frames from a $(500 \times 500)\,\mu\mathrm{m}^2$ sample area in the form of a map, which represents a relatively large area for sufficient determination of material composition. From Fig. 2, it is clear that all the important elements are present, carbon at $0.277\,\mathrm{keV}$, oxygen at $0.525\,\mathrm{keV}$, fluorine at $0.677\,\mathrm{keV}$ and zinc at $1.012\,\mathrm{keV}$. The table in the figure identifies the element weight wt% where the incorporation of zinc at $2.7\,\%$ in the form of MOF dopant can be confirmed. The weight of carbon and fluorine is, as is evident in larger amounts, due to the fact that these are components of the main material.

From the XPS results, a wide spectrum was first taken to provide an understanding of the overall elemental composition and bonding in the material. Four regions of interest are indicated in Fig. 3 marked in purple and further listed as high-energy spectrum. From the resulting signal intensities, the ratios of the elements can be concluded in a similar way to the EDS measurements. Thus, fluorine dominates the spectrum, but zinc is present in very small amounts, which was expected. The other peaks marked with dashed lines are mainly Auger components decay (O KLL, F KL1L1, F KL1L23, and F KL23L23). Furthermore, weaker bands of F2s and Zn3d were identified at 32.2 eV and 8.19 eV, respectively.

For a more detailed study of the elemental bonding, XPS high-energy spectra were obtained. For carbon in Fig. 4a, the $CF_2$ bond was identified, from the left, at an energy of $\sim$$291.3\,\mathrm{eV}$ followed by the very strong FC−OH bond at an energy of $289.3\,\mathrm{eV}$. Subsequently, a C−O bond at $286.7\,\mathrm{eV}$ and another strong peak representing a C−O/$CH_2$ bond at $284.8\,\mathrm{eV}$. The second, least prominent peak is the C−C/C−H bond at $283.5\,\mathrm{eV}$.

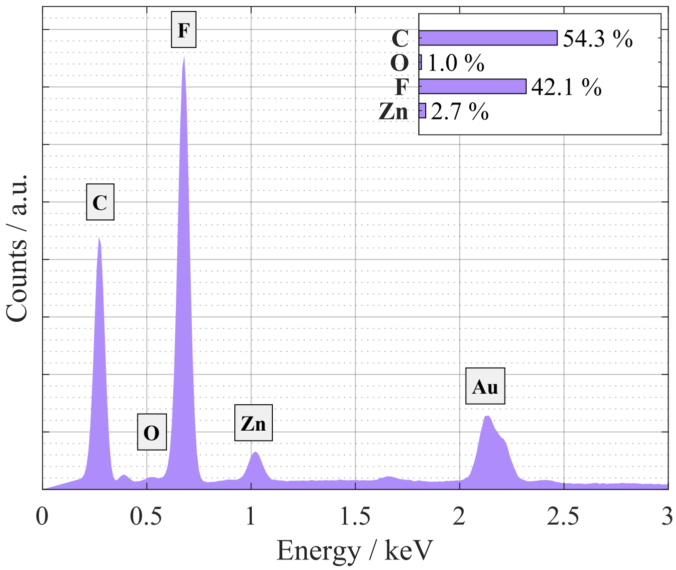



Fig. 2. Survey spectra from EDS showing all the important elements that the sample is composed of. The bar graph also shows the percentage of these elements in the material.

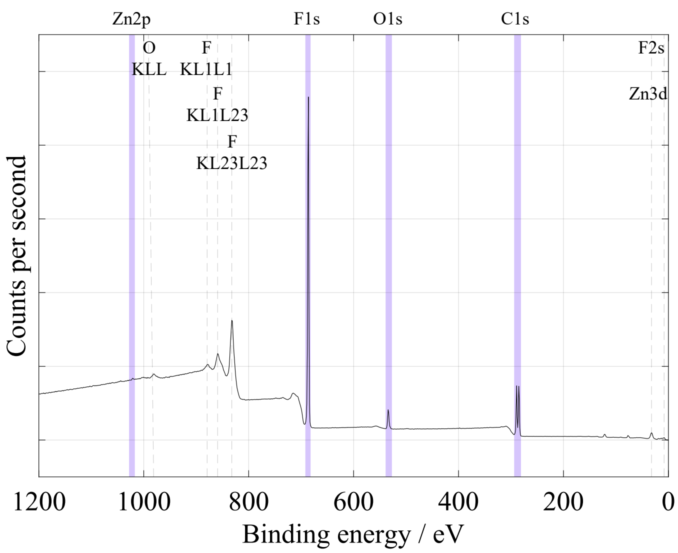
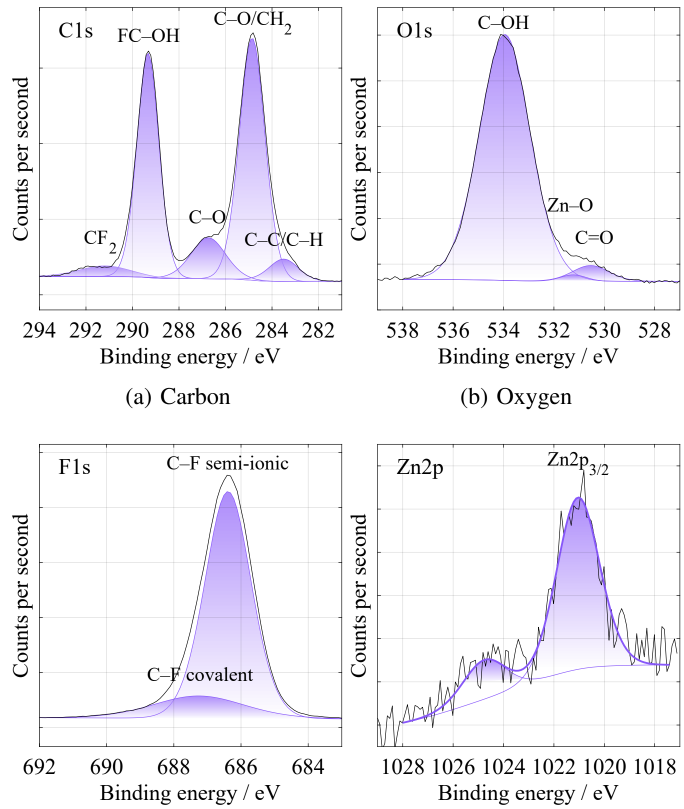



Fig. 3. Wide XPS spectra of the fabricated nanofibrous spec-imen. Highlighted are the most relevant elements, from left: zinc, fluorine, oxygen and carbon.

There are two distinct peaks on the high-energy band of oxygen in Fig. 4b. With the first m ost p rominent peak likely consisting of two possible components, namely C_OH and C_O. The C_O bond was then identified a t 530.7 eV. However, it is likely that this may be a Zn_O bond, which may also confirm t he s uccessful z inc binding.

For fluorine i n F ig. 4 c, t wo t ypical P VDF b indings are revealed, namely a C_F covalent at 687.3 eV and a much more prominent C_F semi-ionic at 686.4 eV.

The occurrence of zinc from Fig. 4d as seen, was the smallest of the four high-energy spectra, as evidenced by the weaker signal from the plot. A $Zn2p_{1/2}$ band was detected at 1021.0 eV.




(a) Carbon
(b) Oxygen
(c) Fluorine
(d) Zinc

Fig. 4. High-energy spectras of the most important elements in the material: (a) carbon, (b) oxygen, (c) fluorine, and (d) zinc from XPS spectroscopy.

The Raman spectroscopy plot in Fig. 5 show many peaks forming a comparatively complex spectrum. The most significant v alues w hich p romotes c rystallization o f $\alpha$ - a nd $\beta$-phases occur at $800\,\mathrm{cm}^{-1}$ and $840\,\mathrm{cm}^{-1}$ respectively. The $840\,\mathrm{cm}^{-1}$ value represents mixtures of $CF_2$ stretching, CC stretching and $CH_2$ rocking. The $\gamma$-phase is represented by a peak at $1431\,\mathrm{cm}^{-1}$, i.e. $CH_2$ deformation or scissoring, but some publications also attribute this peak to the $\beta$-phase. Subsequently, a strong combination of all these phases is presented at $2980\,\mathrm{cm}^{-1}$ representing $CH_2$, asymmetric stretching. It is also shown that the introduction of zinc can affect the crystallization of $\alpha$- and $\beta$-phases [10].

## IV. Conclusion

In this paper, the design and synthesis of a PVDF-doped zinc-based MOF material that can act as a filter for adsorption of heavy metals from wastewater, such as uranium or a gas separator, has been successfully described. It fulfils this function due to its high porosity. Thanks to the pores in the material, the MOF can trap gas or heavy metal molecules. The advantage here is the high tunability of the pores. PVDF@Zn-MOF is formed as a nanofibrous m embrane c reated b y elec-

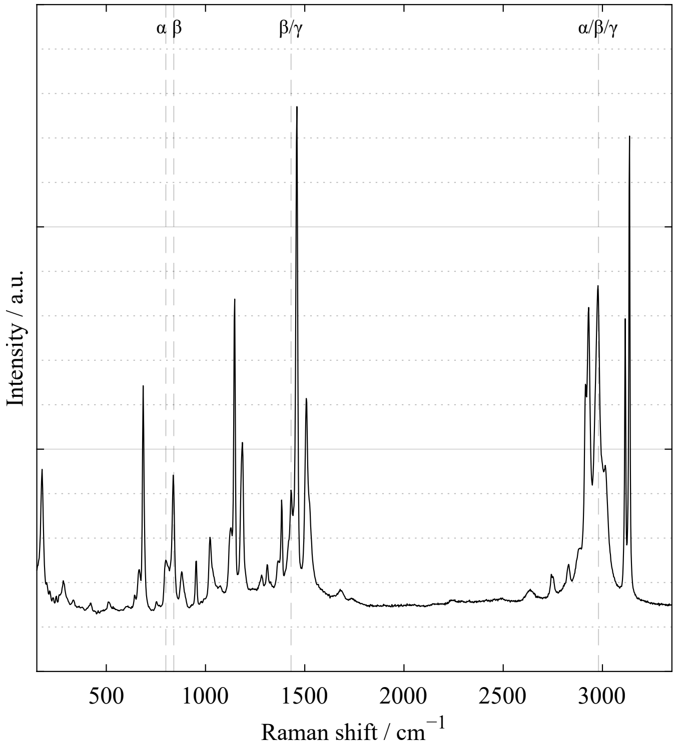



Fig. 5. Raman spectrum of PVDF@Zn-MOF nanofibrous membrane. The image shows a fairly comprehensive finger-print of the material.

trospinning. The results showed that according to the design and the chosen fabrication parameters, a high quality material was formed where MOF was successfully incorporated into PVDF fibers, which was confirmed by several methods. The next steps of this work offer the exposure, for example, to uranium contaminated water and the analysis of the amount of uranium trapped. A promising possibility here seems to be to actively control the membrane and its porosity in the future by means of an applied voltage and the piezoelectric effect that PVDF exhibits in the form of the β-phase.

## ACKNOWLEDGMENT

Authors would like to thank doc. Mgr. Dinara Sobola, Ph.D. and Ing. Nikola Papež, Ph.D., for their valuable comments and guidance through this work.

## REFERENCES

[1] H. Liu, T. Fu, and Y. Mao, "Metal–Organic Framework-Based Materials for Adsorption and Detection of Uranium(VI) from Aqueous Solution," *ACS Omega*, vol. 7, no. 17, p. 14430, 5 2022. [Online]. Available: /pmc/articles/PMC9089359//pmc/articles/PMC9089359/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC9089359/

[2] L. Liu, W. Yang, D. Gu, X. Zhao, and Q. Pan, "In situ Preparation of Chitosan/ZIF-8 Composite Beads for Highly Efficient Removal of U(VI)," *Frontiers in Chemistry*, vol. 7, p. 485261, 9 2019.

[3] A. P. Indolia and M. S. Gaur, "Optical properties of solution grown PVDF-ZnO nanocomposite thin films," *Journal of Polymer Research*, vol. 20, no. 1, pp. 1–8, 1 2013. [Online]. Available: https://link.springer.com/article/10.1007/s10965-012-0043-y

[4] A. E. Baumann, D. A. Burns, B. Liu, and V. S. Thoi, "Metal-organic framework functionalization and design strategies for advanced electrochemical energy storage devices," *Communications Chemistry 2019 2:1*, vol. 2, no. 1, pp. 1–14, 7 2019. [Online]. Available: https://www.nature.com/articles/s42004-019-0184-6

[5] Y. Huang, G. Rui, Q. Li, E. Allahyarov, R. Li, M. Fukuto, G. J. Zhong, J. Z. Xu, Z. M. Li, P. L. Taylor, and L. Zhu, "Enhanced piezoelectricity from highly polarizable oriented amorphous fractions in biaxially oriented poly(vinylidene fluoride) with pure $\beta$ crystals," *Nature Communications 2021 12:1*, vol. 12, no. 1, pp. 1–8, jan 2021. [Online]. Available: https://www.nature.com/articles/s41467-020-20662-7

[6] A. Peleš, O. Aleksić, V. P. Pavlović, V. Djoković, R. Dojčilović, Z. Nikolić, F. Marinković, M. Mitrić, V. Blagojević, B. Vlahović, and V. B. Pavlović, "Structural and electrical properties of ferroelectric poly(vinylidene fluoride) and mechanically activated ZnO nanoparticle composite films," *Physica Scripta*, vol. 93, no. 10, p. 105801, 8 2018. [Online]. Available: https://iopscience.iop.org/article/10.1088/1402-4896/aad749https://iopscience.iop.org/article/10.1088/1402-4896/aad749/meta

[7] T. Pisarenko, N. Papež, D. Sobola, Ş. Ţălu, K. Částková, P. Škarvada, R. Mackøa, E. Ščasnovič, and J. Kaštyl, "Comprehensive Characterization of PVDF Nanofibers at Macro- and Nanolevel," *Polymers*, vol. 14, no. 3, 2 2022. [Online]. Available: /pmc/articles/PMC8839271//pmc/articles/PMC8839271/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC8839271/

[8] C. Ribeiro, C. M. Costa, D. M. Correia, J. Nunes-Pereira, J. Oliveira, P. Martins, R. Gonçalves, V. F. Cardoso, and S. Lanceros-Méndez, "Electroactive poly(vinylidene fluoride)-based structures for advanced applications," *Nature Protocols 2018 13:4*, vol. 13, no. 4, pp. 681–704, 3 2018. [Online]. Available: https://www.nature.com/articles/nprot.2017.157

[9] P. Černohorský, T. Pisarenko, N. Papež, D. Sobola, Ş. Ţălu, K. Částková, J. Kaštyl, R. Macků, P. Škarvada, and P. Sedlák, "Structure Tuning and Electrical Properties of Mixed PVDF and Nylon Nanofibers," *Materials 2021, Vol. 14, Page 6096*, vol. 14, no. 20, p. 6096, 10 2021. [Online]. Available: https://www.mdpi.com/1996-1944/14/20/6096/htmhttps://www.mdpi.com/1996-1944/14/20/6096

[10] J. Ma, Q. Zhang, K. Lin, L. Zhou, and Z. Ni, "Piezoelectric and optoelectronic properties of electrospinning hybrid PVDF and ZnO nanofibers," *Materials Research Express*, vol. 5, no. 3, p. 035057, 3 2018. [Online]. Available: https://iopscience.iop.org/article/10.1088/2053-1591/aab747https://iopscience.iop.org/article/10.1088/2053-1591/aab747/meta

# Methods of generating resistance

1st František Rusnák
*Brno University of Technology*
*Department of control and instrumentation*
Brno, Czech Republic
0009-0009-4733-6517

***Abstract*—The subject of this paper is the design of a programmable resistive decade. The design of the resistive decade is aimed at future use in energy harvester testing. The theoretical part deals with several common ways of coding the desired resistance of the resistive decade. The design of optimized resistive decade and resistive matrix circuits is addressed. The design of the series resistive decade has been optimized by switching the resistances of the individual decade blocks in parallel. A generative machine learning algorithm was used for the design of the resistive matrix. In the final part of the work, the designs were compared with the existing solution.**

***Index Terms*—Programmable resistive load, Resistive decade, Electronic control of resistive decade, Energy harvester, Electronics optimization**

## I. Introduction

Resistive loads are an integral part of laboratory equipment. A full range of resistive decads, resistive loads and active power loads are available. The devices used to generate resistance can be defined based on parameters such as maximum voltage swing, maximum power, range of resistance generated, step fineness, as well as parasitic parameters such as accuracy, stability, and finally dynamic frequency characteristics.

The simplest resistor decade is used, for example, in the development of electronics as a resistor with a selectable value. However, their accuracy is limited and user manipulation is difficult. The opposite is true of high-precision resistor decads, which offer a wide range of adjustable resistance accuracy and parameter stability that make them suitable for calibration purposes or for laboratory measurements that require precise resistance. The basic concept of discrete values of generated resistance is based on the decimal system. By controlling the switches with a microcomputer, user handling can be facilitated and also the SCPI protocol can be implemented, which allows the instrument to be controlled remotely and thus automate the measurement process.

Different applications place different requirements on the parameters of the generated resistance. Resistance decads are also used as ballast loads to simulate the power consumption of the sensor electronics. This paper deals with the design and implementation of an optimized programmable resistive decade whose parameters can be better utilized as ballast load of micro-machines and Energy-Harvesters. According to a market survey [1], such devices have a very poor market representation.

### A. Evaluation of resistance decade requirements

According to [2] summary of piezo-element properties, their research is on the rise and their application in practical conditions is potentially great. The properties of some materials of natural origin are also discussed. The most commonly reported parameters of measured voltages of Energy-Harvesters are up to 30 V. In some cases, higher voltages of around 60 V are recorded and the highest voltages can be up to 112 V. The currents of the piezo-element vary by several orders of magnitude. The currents are lower than 12.2 $\mu A$. Occasionally, currents as high as 1.45 mA have been recorded. The estimation of the frequency parameters of the Energy-Harvesters is not mentioned too much. The numerical values are below 100 Hz. In many cases, Energy-Harvester interactions with humans are mentioned, such as the touch of a finger, the movement of a human knuckle, or the slamming of a fist. Assuming that energy is harvested from human joint motion, the effective frequency range is less than 10 Hz according to [3]. Energy-Harvesters that generate low values of voltages and currents can be connected together with ballast loads that use thin-film low-power resistors and low-power switching elements. It is also desirable to optimize the number of elements used. The reduced wiring of the resistive load will facilitate the use in field measurements such as [4]. Thus, the theoretical design objective is to optimize the circuitry of the resistive decade for the minimum number of components used with a range of generated resistance of 10 Ω - 10 $M\Omega$ and power in the order of tens of $\mu W$.

## II. Principles of variable resistance generation

Resistance decads that are based on passive resistors are distinguished by the simplicity of their circuitry and are the most common. Manual devices are dependent on user intervention. Therefore, a large proportion of resistive loads are equipped with switches that implement the function of selecting one of ten. This type of switching makes it easy to implement a user-friendly input of the resistance value, which is coded directly in the decimal system.

Switching individual resistors using manual switches or electronically controlled relays generates discrete output resistance values at the output of the resistor decade. For devices intended for calibration or measurement, the repeatability of the generated resistance is crucial. However, for other applications that use resistive decade, accuracy may not be the key feature, but a density of steps of generated resistance and a wide range is desired.

### A. Decadic coding principle

A user-friendly solution is to encode the requested value directly in the decimal system. A large number of resistive decodes that are based on the passive principle are internally wired according to the schematic in Figure 1. The principle of decadic coding has the advantage that current flows only through a constant number of contacts, thanks to switches that implement the function of selecting one of the ten. This topology is used for precision resistive decodes, because of the constant transient resistance of the switches in the on state.

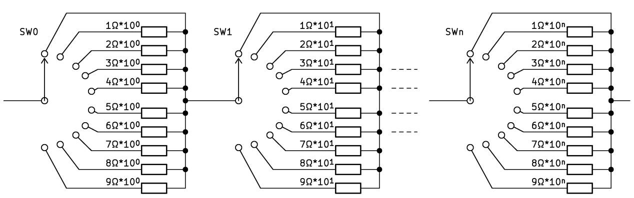



Fig. 1. Principle circuit of resistive decade with decadic coding [1]

### B. Binary coding principle

Less user-intuitive is the binary method of encoding the desired resistance value. The value of the output resistance increases with a constant absolute step as in decade encoding. The binary coding principle of resistance is the most efficient due to the number of discrete steps of the generated resistance to the switches used. Figure 2 shows a circuit that uses resistors $R_i$ whose values form a geometric sequence with a multiple of two. In user terms, directly entering the value of the desired resistance is relatively unintuitive, but it is an efficient method when automated relay switching is involved. Reference [5] presents an efficient use of a microcomputer for binary-based resistive decade control. This is a hybrid circuit where the decade is divided into decade blocks.

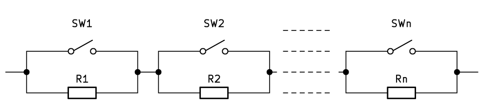



Fig. 2. Principle circuit of resistive decade with binary coding [1]

Binary coding does not provide a constant number of switches that are closed. When switching, the number of switches switched changes and therefore the total additive value of the switches transient resistance.

### C. Active principles and synthetic resistance generators

The generation of synthetic resistance is different from classical resistance decades. The core is the power active electronics which is controlled by the controller. The function of the controller is to provide control of the output current so that it is linearly proportional to the voltage $I = U \cdot R$. In this case, the quantity $R$ represents the gain constant of the controller. This ensures that the whole device appears externally as a purely resistive load.

There are different variants of active loads on the market. However, most of them are laboratory devices designed for power in the order of 100 W. Compared to other devices, the active resistance generator for low power has a very poor representation on the market [1].

## III. Parameters of Resistance Generators

The resistor decade consists of many resistors and switches. It can be up to dozens of components. The parasitic properties of each component affect the overall accuracy of the instrument. For resistors, their accuracy and temperature dependence are crucial. Especially for resistors with high values, their frequency dependence is also critical. Switches are also required to have low on-state resistance and high open-state impedance.

### A. Static parameters and precision

Static parameters such as accuracy, repeatability and independence from ambient phenomena provide basic information about the instrument. For laboratory instruments, overall accuracy is crucial. For other applications such as electronics development, simulating appliance loads using power ballasts, or even low-power ballasts, the accuracy requirement may not be critical. Generating resistance over a wide range will ensure the versatility of the resistive decade.

In general, the output resistance is formed by a series-parallel combination of resistors. In a purely series connection, the precision of the components with the highest resistance is most evident. In a parallel connection, the precision of the components with the lowest resistance value is most evident. In a combined series-parallel circuit, the total deviation is calculated using formula 2, which uses the partial derivative of equation 1, where the function f represents the formula for calculating the total resistance of the series-parallel combination.

$$Y = f(X_1, X_2, ..., X_n) \tag{1}$$

$$|\Delta_Y| = \sum_{i=1}^{n} \left| \frac{\partial Y}{\partial x_i} \right| \cdot |\Delta x_i| \tag{2}$$

For ballast loads that simulate the power draw of powered circuits, it is not the accuracy that plays a role, but the range of resistance generated, the remote communication interface to automate the measurement and last but not least the size, weight and shape of the device that would allow the use of small Energy-Harvesters for measurement and testing.

### B. Dynamic parameters

When an AC signal is applied to a resistive decade, the dynamic properties of the internal circuitry are revealed. Laboratory resistive decads are usually large and heavy. The size of the area of the conductors on which the electrical signal occurs is crucial. As the area of all resistors, switches and interconnecting wires increases, so does their parasitic capacitance. The resistors themselves have dynamic properties. At low values of resistance, parasitic series induction is most evident, and at high values of resistance, parasitic parallel capacitance is most evident [6].

In complex circuits, parasitic dynamic properties are of great importance, especially when generating higher resistance values. Even the transmitted resistive decrements quickly lose accuracy with increasing frequency [7]. For example, the accuracy of the MEATEST M632 is quoted as 0.01 %, but at a setpoint of 1 $k\Omega$ and a frequency of 10 $kHz$, the accuracy is quoted as 5 %, and when the resistance is increased to 100 $k\Omega$, this accuracy is only achieved up to 100 $Hz$ [1].

To suppress parasitic properties, it is advisable to choose SMD resistors with low parasitic properties, which are, however, more expensive. Another way to suppress parasitic properties is to reduce the number of components used, which, however, reduces the number of discrete steps of the generated resistor. In this case, the waveform of the generated resistance must be optimized for the needs of the application, in this case for use as a ballast.

### *C. Course of generated resistance*

Electrical resistance generators do not generally provide a continuous resistance value. The range of resistance values that can be set represent a set of discrete values of the generated resistance. The difference of two adjacent values is called a step. The step size can be evaluated in terms of absolute step and relative step.

In more sophisticated circuits, the step can be so small that the desired set of generated resistance values redundantly covers the required range. The generated resistance curve may then appear continuous for specific applications. In applications that require a wide range of generated resistance and relatively low accuracy, redundancy is caused by oversized resistor decade circuitry. The maximum possible number of all combinations is expressed by the formula 3, where R is the number of bistable switches and C is the total number of combinations.

$$C = 2^R \tag{3}$$

The ratio 4 between the actual number of useful combinations U and the number of combinations C can be used in assessing the hardware utilization of resistive decades.

$$Q = \frac{U}{C} \tag{4}$$

In binary combination, Q=1 is achieved. In decadic coding, $Q = 9.77 \cdot 10^{-3}$ is achieved with one block of resistive decade and $Q = 9.09 \cdot 10^{-9}$ with four blocks.

The value of the output resistance for manual decrements increases with the absolute step as shown in Figure 3 However, the relative step is disproportionate in the limit, and it is undersized at low resistance values and oversized at high resistance values.

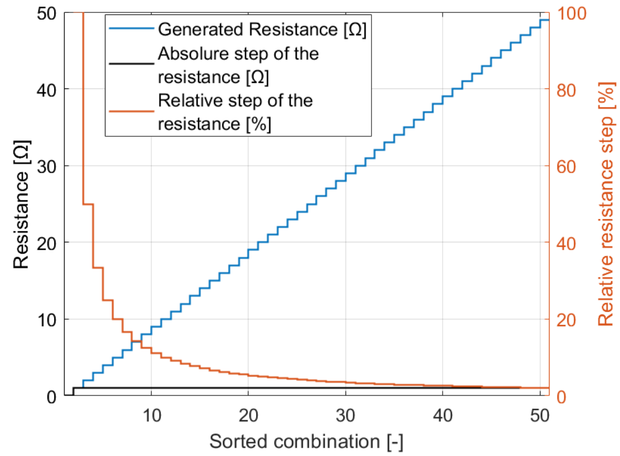



Fig. 3. Manual resistance decade characteristic chart

## IV. Design optimization of resistance generators

In the case of resistive decade deployment in an application that requires the generation of less accurate resistance, but over a wide range of values, a constant absolute step is not appropriate. A linear progression of the set resistance causes inefficient use of switch combinations where, on the one hand, the step is too large and, on the other hand, there is waste due to too small a step. The aim of the innovative solution is the efficient use of switchable elements. The solution may be to optimize the resistive decade to achieve a relative step of the resistive decade.

### *A. Decadic Resistive Decade Optimization*

The design is based on a decadic resistive decade circuit. In order to increase the number of generated steps, a circuit that uses parallel connection of resistors in each block was designed. In this case, however, the device loses user-friendliness in manual coding. The clarity of the input can be improved by automatic switching by the microprocessor. The benefit is to achieve multiple combinations of generated resistance with the same number of relays.

In a circuit that is to be controlled electronically, the manual switch must be replaced by relays. Commonly available relays switch one contact or switch between two contacts. Therefore, one SP10T switch must be replaced by ten SPST relays as shown in Figure 4.

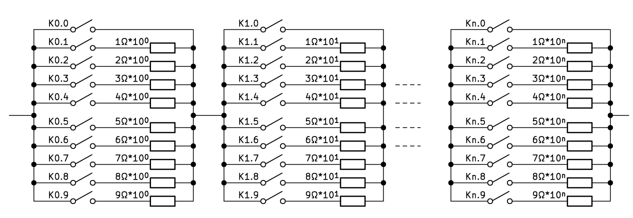


Fig. 4. Principle scheme of resistive decade with decadic coding using relays [1]

A script was created in Matlab that uses brute force computation to find the ideal solution for a parallel combination of switching resistors with the minimum number of relays used. The goal of the algorithm is to find a combination of resistors that would replace one block of serial decade encoding. The algorithm was run to search for the best combination of resistors out of a million random combinations and the combination of resistors for one block of resistor decade was found, which is shown in Figure 5 and the ordered output resistance waveform is plotted in Figure 6.

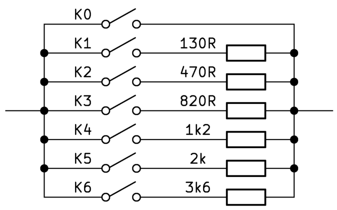



Fig. 5. Schematic of optimized resistive decade block with decadic coding [1]

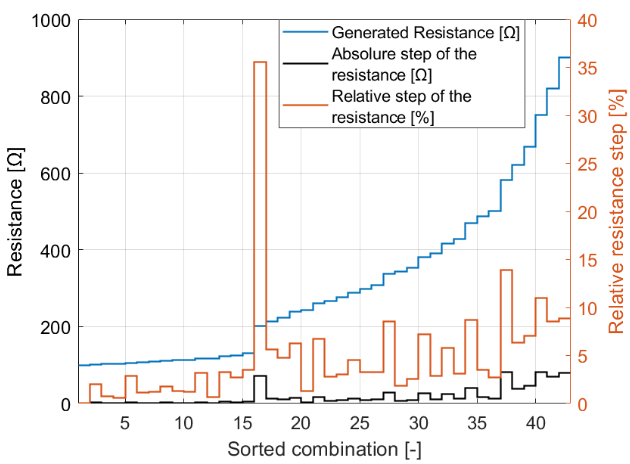



Fig. 6. Graph of the generated resistance of the designed block of the optimized decade

The decade block multiple is determined by the formula 5, where $R_M$ is the maximum resistance of the decade block and $\Delta R$ is the difference of two adjacent values of the total resistance.

$$D = \frac{R_M}{Max(\Delta R)} \tag{5}$$

The multiplier for the proposed combination is equal to $D = 12.1763$. This means that after the series connection of the individual blocks of the resistive decade, the circuit will contain the series connection of the blocks from Figure 7, where the resistor values of each subsequent block will be multiples of the constant $D = 12.1763$. In this case, however, it is preferable to use the constant $D = 10$, because this multiplication will always yield a resistance value that will belong to the resistor series E24.

With a lower number of resistors than 7, no solution was found that had a resolution suitable for decimal encoding. Thus, by combining the resistors in parallel, 3 resistors with switches were spared.

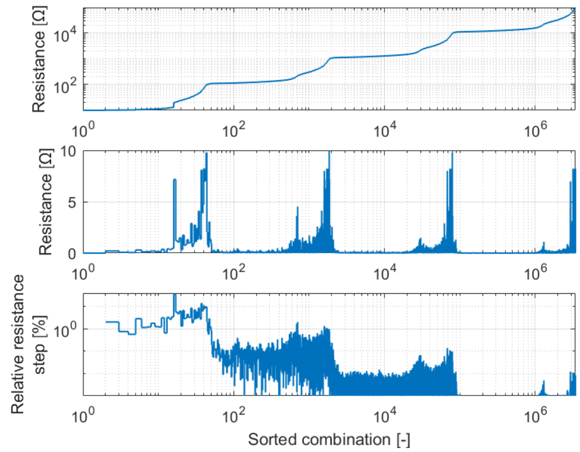



Fig. 7. Graph of the generated resistance of the optimized decade

By substituting the number of relevant combinations into the 4 equation, an efficiency of $Q = 0.34$ for a single decade block and $Q = 0.0127$ for a series connection of four decade blocks was achieved.

### *B. Resistive decade design using generative algorithm*

The aim of the design is to find an optimal circuit with a small number of hardware components. Genetic machine learning algorithms explore a wide space of possible serial-parallel circuits and search for the circuit that most closely matches user-defined criteria. A great advantage is the self-selection of the determining parameter of the resistive matrix quality evaluation. The algorithm favors resistive matrices with the lowest loss value, which is determined based on the evaluation conditions, and suppresses resistive matrices with high loss values. By actually designing the loss value determination conditions, the desired output resistance curve can be shaped to some extent. Therefore, the design is not bound by the characteristics exhibited by manual resistive decads. The desired characteristic of the output resistance is the set of values that are as close as possible to the geometric sequence. The waveform error is calculated using the equation 6, where $R_i$ is the ordered values of the generated resistance, n is the number of relevant steps, and DN is the desired number of decades of generated resistance.

$$Err = \sum_{i=1}^{n-1} \left(\frac{R_{i+1}}{R_i} - 10^{\frac{DN-1}{n}}\right)^2, where R_i \leq R_{i+1} \tag{6}$$

The computational matrix, which serves as a simulation model of a real resistive network, is formed by the bistable blocks in Figure 8. There are three types of blocks where parallel and series combination of resistor and switch is used.

The third option is left for a separate resistor without involving a switch.

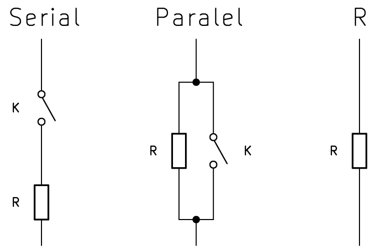



Fig. 8. Three types of blocks used to implement the resistance matrix [1]

A Matlab script was designed to implement the machine learning generation algorithm. The principle of the resistance matrix is shown in Figure 9. The resulting resistance is recalculated between points A and B. The problem is solved using a system of linear equations, where the value of each block $B_i$ is first calculated, which depends on the bistable state in which the switch is. Then the system of linear equations is solved, where the value of the individual voltages at the nodes $U_i$ is determined. The known values of the voltages are $U_1 = 1$ V and $U_n = 0$ V, where $U_n$ is the voltage of the last node. After solving the system of linear equations, all the voltages of $U_i$ are known. Then, the calculation of the total current flowing through point B is performed according to equation 7, where the voltage $U_i$ indexed based on the Uindex, which carry information about which voltage node the block is connected to. The coefficient Kcon determines whether the block $B_i$ is connected to the voltage node $U_n$ and therefore to point B.

$$I_B = \sum_{i=1}^{n_B} \frac{U_{Uindex_i}}{B_i} \cdot Kcon_i, where Kcon = \{0, 1\} \tag{7}$$

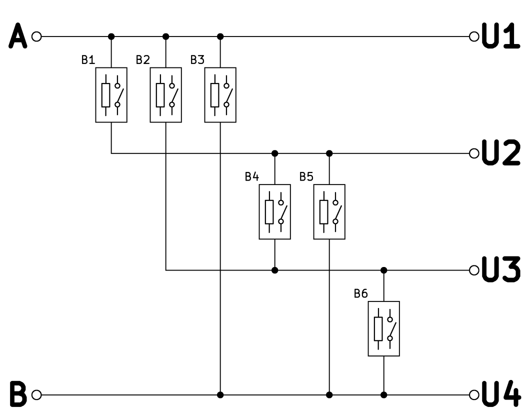



Fig. 9. Principle diagram of resistor matrix used for calculation of serio-parallel circuit of resistor blocks [1]

The resistance value 8 is then determined from the known voltage between points A and B and the calculated total current.

$$R = \frac{1}{I_B} \tag{8}$$

The resistance matrix is arbitrarily variable. The number of voltage nodes $U_n$ and individual blocks $B_n$ can be increased. However, as the number of blocks $B_n$ increases, the computational power requirements of the genetic algorithm search increase. Four scripts were run in parallel to find the optimal resistor matrix, which were completed after five days. The solution was searched for 15 $B_n$ blocks and 10 voltage nodes. The search range was set from 100 $\Omega$ to 1 $M\Omega$.

The algorithm generates the resistance matrix data. The data contains information about the type of switchable bistable block, the resistance value of the block and two indices that define which $U_x$ nodes the bistable block is connected to. Based on the acquired data, a series-parallel schematic of the resistors and switches was manually created and is shown in Figure 10.

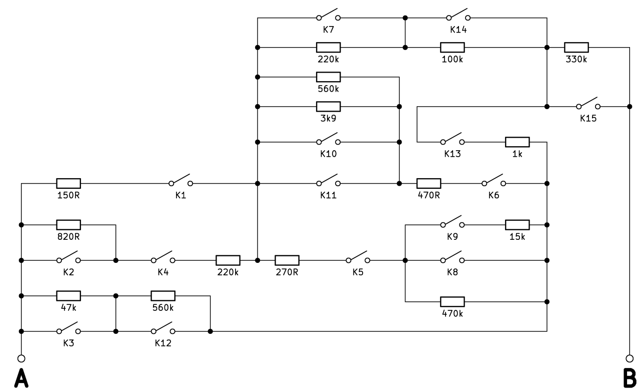



Fig. 10. Circuit diagram of resistor matrix generated by machine learning generation algorithm [1]

The generated circuit combination is not similar to any typical resistive decade circuit. Therefore, the generated resistance curve on the 11 plot is quite irregular. The absolute step values reach a large dispersion over the entire range of generated resistance. The relative step reaches a maximum value of 51.6 %. The hardware utilization efficiency according to the 4 formula is $Q = 0.93$. The value was calculated for the assumed 14 relays, due to the duplicated relays K10 and K11 in Figure 10. This doubling was generated by the genetic algorithm. To find a matrix without this defect, the algorithm must be replicated with different initial values.

### C. Comparison of proposals

Figure 12 compares the relative steps of the manual decade with the relative steps of the solved designs. The progressions are plotted by ordering the relative step values from highest to lowest.

The progression of the hand decade and the proposed resistance matrix drops below 1 % after approximately the same number of combinations. In this respect, the performance of the proposed resistive matrix is comparable to the resistive

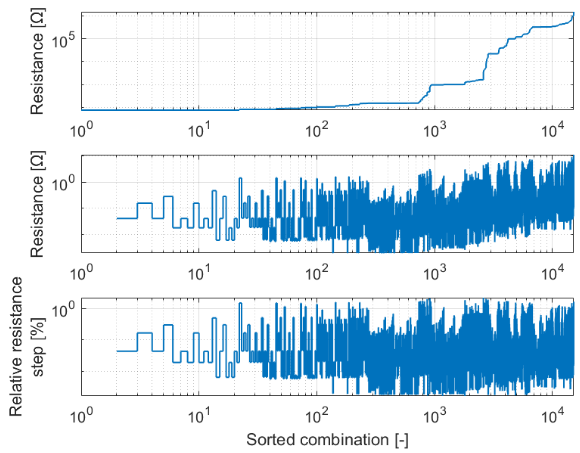



Fig. 11. Graph of the output resistance of a resistance matrix designed using a genetic machine learning algorithm

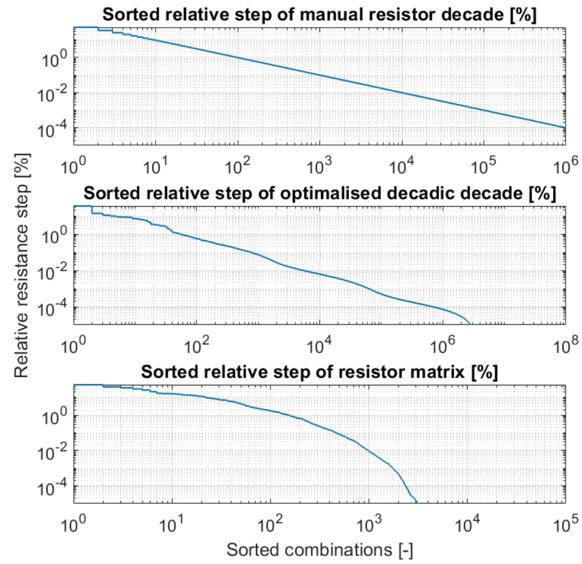



Fig. 12. Graph of the relative resistance step generated by the three coding principles

decade with binary coding. According to the relative step progression from the 11 plot, it can be said that the proposed resistive matrix contains high relative steps throughout the generated resistance waveform, while the binary resistive decade contains the highest steps at the beginning of the waveform. The advantage of the proposed design is that compared to the binary resistive decade, the current does not flow through a large number of switches at low generated resistance.

Comparing the progress of the optimized decadic resistance decay in the 12 plot, the relative jump curve drops below 1 % earlier. The whole relative step waveform does not drop too dramatically. Thus, the optimized resistive decade shows a relatively smooth step even at low values of the generated resistance.

## V. Conclusion

The aim of this work is the theoretical optimization of the existing solutions of resistive decades. The most common ones are manual resistive decads, which depend on user intervention. The most widely used coding principles are binary and decadic. Another type of resistive decads are electronically controlled instruments that allow automation of the measurement process.

Two ways of streamlining the resistive decade wiring have been proposed. The first is based on the concept of resistive decade coded decade. In this case, a significant improvement in the efficiency of the existing circuitry has been achieved. By designing an electrical circuit that uses a parallel combination of resistors, three switching elements and three resistors per decade block were saved. A hardware utilization ratio of $Q = 0.0127$ was determined for the four-block circuit, which is several times higher than the existing circuit, where $Q = 9.09 \cdot 10^{-9}$.

Another proposal is optimization using genetic algorithms that search for the optimal matrix of series-parallel connections of resistors and switches. In this case, a similar parameter $Q = 0.93$ was achieved as in the manual resistive decade encoded in binary. There is a non-constant number of switches in the circuit through which current flows, but these switches are not connected in pure series as in the existing circuit. This method allows custom shaping of the waveform of the generated resistance. This is an advantage in applications with atypical requirements.

## References


[1] RUSNÁK, František. Programovatelná odporová zátěž. Brno: Vysoké učení technické v Brně, Fakulta elektrotechniky a komunikačních technologií, Ústav automatizace a měřicí techniky, 2023, 69 s. Diplomová práce. Vedoucí práce: Ing. Jan Kunz, Ph.D.

[2] Sezer, N. and Ko¸c, M. (2021). A comprehensive review on the state-of-the-art of piezoelectric energy harvesting. Nano Energy, 80, 105567.

[3] Q. Meng, B. Li, H. Holstein, Recognition of human periodic movements from unstructured information using a motion-based frequency domain approach, Image and Vision Computing, Volume 24, Issue 8, 2006, Pages 795-809, ISSN 0262-8856,

[4] J.C. Cámara-Molina, A. Romero, E. Moliner, D.P. Connolly, M.D. Martínez-Rodrigo, D. Yurchenko, P. Galvín, Design, tuning and in-field validation of energy harvesters for railway bridges, Mechanical Systems and Signal Processing, Volume 208, 2024, 111012, ISSN 0888-3270, https://doi.org/10.1016/j.ymssp.2023.111012. (https://www.sciencedirect.com/science/article/pii/S0888327023009202)

[5] M. Helmy A. Raouf, Design, implementation, and characterization of a new resistance box fabricated with fifteen outputs decades, Measurement, Volume 101, 2017, Pages 206-210, ISSN 0263-2241,

[6] Tomáš Zedníček, https://www.doeeet.com/content/eee-components/passives/abc-of-clr/voltage-and-frequency-dependence-on-resistors/

[7] Pavliska, V. (2017). Programovatelná odporová dekáda.

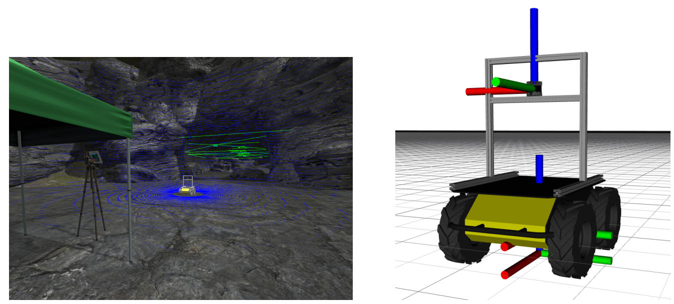
# Localization Accuracy of Autonomous Mobile Robots: A Comprehensive Evaluation of KISS-ICP Odometry

1st Milos Cihlar
*Dept. of Control and Instrumentation*
*Brno University of Technology*
Brno, Czech Republic
ORCID: 0000-0002-5478-9196

***Abstract*—This paper emphasizes the importance of accurate localization for the appropriate behavior of autonomous mobile robots. In particular, it provides a rigorous evaluation of the accuracy of the KISS-ICP algorithm, a lightweight lidar-based pose estimation algorithm known for its simplicity with minimal setup parameters. The algorithm works only with lidar data; unlike other more sophisticated SLAM algorithms, it does not use IMUs. The algorithm is shown to work under a variety of conditions and sensors.**

**To comprehensively evaluate how the algorithm performs under varying conditions, several sensors were used in the experiments. The evaluation included a series of 3D lidars, namely Ouster OS0, OS1, and OS2. These lidars are characterized by different field-of-view settings and operating modes with different point per row. In order to facilitate this evaluation, a data set with extensive robot traversals over a distance of more than 9 km in two different environments, each equipped with different sensor types, was prepared.**

***Index Terms*—Localization, SLAM, Outdoor navigation, ROS**



(a) simulated environment

(b) Husky with sensors description

Fig. 1. Simulated environment with lidar visualization and sensor layout on the Husky mobile robot

## I. Introduction

Localization in mobile robotics is a method used to estimate the state of the robot with respect to a given frame of reference. It is essential for effective navigation and operation of robots in their environment. Localization systems require computing power and sensors. Sensors are classified into two main categories. First, there are proprioceptive sensors. [4] These sensors measure the internal states of the robot. The most commonly used sensors include various types of encoders. These include optical, magnetic, capacitive, and inductive encoders. They also include gyroscopes and inertial measurement units (IMUs). Environmental properties are measured by exteroceptive sensors [4] such as cameras and lidars. These sensors provide essential data for localization tasks. They allow robots to navigate and interact effectively with their environment.

Localization is an essential and integral part of any robotic task in various types of vehicles, including autonomous unmanned ground vehicles (UGVs) and unmanned aerial vehicles (UAVs). Anchor-based localization [4] is one of the different types of localization methods. These localization techniques rely on external references. These include GPS (Global Positioning System), UWB (Ultra-Wideband), and Mocap (Motion Capture Systems).

While GPS may seem suitable for long-term localization, it has limitations such as susceptibility to interference in urban areas, under foliage, near buildings, or in other obstructed regions that can block satellite signals. In addition, all anchor-based localization methods have a common drawback: They require satisfactory line-of-sight to their beacon for accurate localization.

In contrast, localization methods that do not rely on external anchors are referred to as onboard or sensor-based [1] localization. These approaches use only onboard sensors for localization [4]. Onboard localization often uses simultaneous localization and mapping (SLAM) or odometry algorithms, which can be based on different combinations of sensors. Examples include lidar-based, visual-based, lidar-IMU, or their combinations and fusion techniques.

The development of 3D lidar sensors such as the Velodyne PUCK, HDL32, Livox MID-70 and MID-360, and Ouster OS series has resulted in significant developments in lidar-based pose estimation and SLAM algorithms. These sensors provide high-resolution point cloud data, enabling more accurate perception of the environment for mobile robots and autonomous vehicles.

The primary goal of SLAM algorithms is to estimate ego-motion, which refers to the robot's own motion relative to its environment, while minimizing the error from the true position. In addition, SLAM algorithms aim to construct maps of the environment, ideally preserving the quality of these maps over time. This preservation of map quality is critical for sustainable robot navigation and long-term localization performance.

Many modern 3D SLAM algorithms use the extraction of feature points, such as edges and surfaces, to estimate the position and orientation of robots [1]. One such algorithm is Lidar Odometry and Mapping in Real-time, or LOAM [5]. LOAM is divided into two parts: a high-frequency part for estimating the odometry, and a part for matching and registering the points into a 3D map. [5]

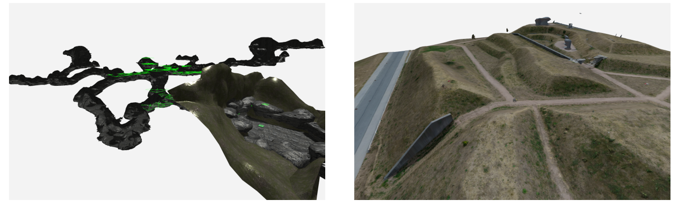
The LOAM SLAM algorithm has inspired many others [1], such as LIO-SAM [10], which integrates data from lidar and IMU, and LeGO-LOAM [6]. LeGO-LOAM is lightweight and operates on ground-based optimization using ground planes and features extraction. It then employs a two-stage Levenberg-Marquardt optimization method to solve the 6DOF transformation between two scans using planar and edge features. LeGo-LOAM-SC [7] builds on LeGO-LOAM and adds two types of loop closure.

Over the past decade, there has been a trend of increasing complexity in SLAM algorithm solutions [1]. However, KISS-ICP takes a different approach by not using feature extraction, learning methods, or loop closure. Unlike other algorithms [2] [8] [9] [10], it does not use pose graph optimization. Instead, KISS-ICP relies primarily on the ICP [3] algorithm, which has been in use for over 30 years and still achieves state-of-the-art results. The KISS-ICP algorithm is proof of this. It can be used in multiple environments, including autonomous cars, UAVs and UGVs [1].

The only algorithm that was superior [1] to KISS-ICP at the time of its inception was CT-ICP [2]. CT-ICP is a FULL SLAM method. Unlike KISS-ICP, CT-ICP not only estimates position, but also constructs a map and uses closed-loop methods to ensure long term stability of pose estimation. The core of CT-ICP is the introduction of combined continuity in scan matching and discontinuity between scans [2].

### A. Contribution

The objective of this paper is to evaluate the quality of Lidar SLAM odometry in different environments using different types of sensors. Specifically, the study focuses on comparing three types of sensors: OS0, OS1, and OS2, which can operate with varying data density. The aim is to evaluate the influence of not only the data density but also the field of view of the sensor on the localization accuracy. Furthermore, the localization accuracy of lidars with lower number of rows than the standard 128-row for Ouster lidars, will be investigated. The results of this evaluation can be used to properly adjust the covariance matrices for subsequent fusion with other localization methods.



(a) DARPA SubT Challange - qualification

(b) BUT park

Fig. 2. Environment where the dataset was collected

### B. Outline

The paper is organized as follows: Chapter II describes the creating dataset, sensor description, and environment description in which the dataset was collected. The following chapter III focuses on the experiments with the KISS-ICP algorithm and describes its setup in detail. The chapter IV discusses the evaluation of the data, including descriptions of the individual metrics. Finally, we present our conclusions in the last section V.

## II. Dataset

To evaluate the localization accuracy using the KISS-ICP algorithm. We have prepared a dataset specifically designed for testing the SLAM algorithm. The dataset consists of simulated data generated in the Gazebo Garden simulator. The advantage of the simulator is its ability to collect data from different sensors on identical trajectories. This capability allows us to evaluate localization quality using data that differ only in a single sensor's characteristics, a scenario that is unattainable in real-world testing.

Data collection included readings from 3D lidars, IMU sensors, wheel odometry, and ground truth information. Leveraging the ROS 2 bag functionality facilitated efficient data collection [12]. In total, the dataset includes over 9 kilometers of robot motion with multiple sensor configurations. The dataset comprises over 150GB of data.

This dataset provides a robust basis for evaluating the localization accuracy of the KISS-ICP algorithm under different sensor configurations. It contributes to a sophisticated understanding of its performance in real-world scenarios.

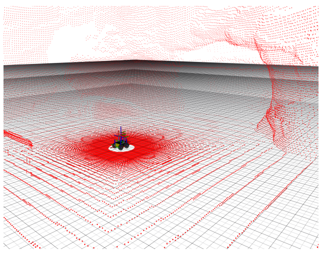
### A. Environment

The dataset was collected in two different environments, see Fig. 2. The first environment, referred to as "Cave" in this paper, is the qualification environment from the DARPA Subterranean Challenge. This environment simulates a cave environment characterized by obstacles, uneven and unstructured terrain, and significant slopes. In contrast, the second environment, referred to as "Park", is a model [13] of the actual environment near the Faculty of Electrical Engineering and Communication, Brno University of Technology. This environment is similar to a park, with predominantly well-maintained surroundings such as grass, gravel, and sandy paths. Unlike the challenging cave environment of the DARPA

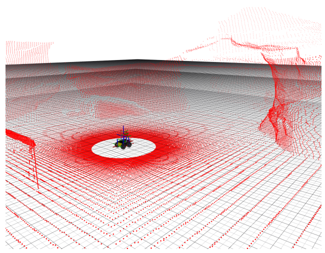
SubT challenge, the park contains significantly fewer feature points.

Within each environment, the robot traveled and collected data on three different paths, resulting in a total of six different trajectories recorded. These trajectories represent the variety of terrain and conditions encountered by the robot in both the ”Cave” and ”Park” environments. Detailed information on all experiments conducted is presented in Table II, providing a comprehensive overview for further analysis and comparison.

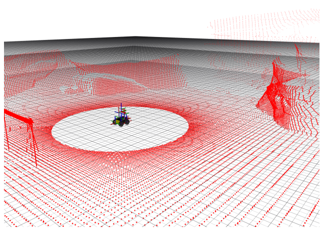
### B. Sensory Setup



(a) Ouster OS0



(b) Ouster OS1



(c) Ouster OS2

Fig. 3. Sensor visualization

Ouster 3D scanners were chosen for the experiment, specifically the OS0, OS1, and OS2 models, which differ primarily in their vertical field o f v iew ( VFoV). T he V FoV o f the OS2 model is four times narrower than that of the Ouster OS0. VFoV is inversely proportional to point density. The visualization of the points in Rviz for each sensor with different VFoV is shown in Fig. 3.

Each sensor can operate in several modes depending on the number of points it measures, i.e. its horizontal resolution. For this dataset, three modes were selected to operate at horizontal resolutions of 512, 1024, and 2048 points per row, respectively. All three sensors measure vertically 128 rows. The operating frequency was set to 10Hz. The basic specifications o f these sensors are listed in Table I.

TABLE I. Sensors overview

| Sensor | VFoV [°] | Range [m] | Max PPS [-] | Max Rate [Hz] |
|---|---|---|---|---|
| **Ouster OS0** | 90 | 100 | 5.2M | 20 |
| **Ouster OS1** | 45 | 200 | 5.2M | 20 |
| **Ouster OS2** | 22.5 | 400 | 2.62M | 20 |

### C. Data Description

The datasets were collected in two different simulated environments. In each environment, the robot followed three different trajectories for the ”Cave” environment and for the ”Park” environment. Point clouds from three 3D lidars with different VFoV were collected during each run in a given environment, as detailed in TABLE I. In addition, these sensors had different horizontal resolutions in subsequent runs, starting with 512 points for trajectory 1 in both the ”Cave” and ”Park” environments, and then progressing to 1024 and 2048 points, respectively.

The data was stored in ROS 2 bags with the following topics.

- `/tf_static` (`tf2_msgs/msg/TFMessage`)
- `/ground_truth` (`nav_msgs/msg/Odometry`)
- `/ouster_os0/cloud` (`sensor_msgs/msg/PointCloud2`)
- `/ouster_os1/cloud` (`sensor_msgs/msg/PointCloud2`)
- `/ouster_os2/cloud` (`sensor_msgs/msg/PointCloud2`)
- `/xsens_mti680g/imu` (`sensor_msgs/msg/Imu`)
- `/odometry` (`nav_msgs/msg/Odometry`)
- `/cmd_vel` (`geometry_msgs/msg/Twist`)

The *ground_truth* odometry message, which contains accurate information about the robot’s position and orientation published by the Gazebo simulator, can be used to evaluate localization accuracy. IMU and proprioceptive wheel odometry data are also included in the dataset. Static transformation information about the positions and orientations of the sensors relative to the base frame of the robot is also provided.

## III. Experiment

The long-term localization quality using the KISS-ICP algorithm was evaluated using a collected dataset. The KISS-ICP algorithm has only 7 adjustable parameters [1] that affect its performance, unlike other complex algorithms [2] [5] [6] [7] [9] [10]. For the purpose of this study, the original parameter settings [1] were preserved. The list of the values of all the parameters is shown in Table III.

In all, 23 experiments were performed with the sensor settings shown in Table II. Data acquisition was performed using ros2 bag [12]. The topic *nav_msgs/msgs/odometry* published by the KISS-ICP node was recorded. The aim was to evaluate the quality of the localization algorithm regarding VFoV, horizontal and vertical resolution.

## IV. Results

### A. Performance metrics

The evaluation of the lidar odometry is done by comparing the positions obtained by the KISS-ICP algorithm with the

TABLE II. Dataset overview

| environment | trajectory | sensor | horizontal res | length [m] | duration [s] | avg vel $[ms^{-1}]$ | scans | scan loss [%] |
|---|---|---|---|---|---|---|---|---|
| Cave | #1 | OS0 | 512 | 1153 | 1142 | 1.0 | 11220 | 1.75 |
| | #1 | OS1 | 512 | 1153 | 1142 | 1.0 | 11220 | 1.75 |
| | #1 | OS2 | 512 | 1153 | 1142 | 1.0 | 11220 | 1.75 |
| | #2 | OS0 | 1024 | 416 | 408 | 1.02 | 4047 | 0.8 |
| | #2 | OS1 | 1024 | 416 | 408 | 1.02 | 4047 | 0.8 |
| | #2 | OS2 | 1024 | 416 | 408 | 1.02 | 4047 | 0.8 |
| | #3 | OS0 | 2048 | 431 | 413 | 1.06 | 2366 | **42.7** |
| | #3 | OS1 | 2048 | 431 | 413 | 1.06 | 2366 | **42.7** |
| | #3 | OS2 | 2048 | 431 | 413 | 1.06 | 2366 | **42.7** |
| | | vertical res | | | | | | |
| Cave | #4 | 4 | 1024 | 349 | 346 | 1.01 | 3453 | 0.2 |
| | #4 | 8 | 1024 | 349 | 346 | 1.01 | 3453 | 0.2 |
| | #4 | 16 | 1024 | 349 | 346 | 1.01 | 3453 | 0.2 |
| | #4 | 32 | 1024 | 349 | 346 | 1.01 | 3453 | 0.2 |
| | #4 | 64 | 1024 | 349 | 346 | 1.01 | 3453 | 0.2 |
| Park | #1 | OS0 | 512 | 454 | 331 | 1.37 | 3303 | 0.21 |
| | #1 | OS1 | 512 | 454 | 331 | 1.37 | 3303 | 0.21 |
| | #1 | OS2 | 512 | 454 | 331 | 1.37 | 3303 | 0.21 |
| | #2 | OS0 | 1024 | 327 | 293 | 1.15 | 2617 | 10.7 |
| | #2 | OS1 | 1024 | 327 | 293 | 1.15 | 2617 | 10.7 |
| | #2 | OS2 | 1024 | 327 | 293 | 1.15 | 2617 | 10.7 |
| | #3 | OS0 | 2048 | 286 | 253 | 1.31 | 912 | **63.9** |
| | #3 | OS1 | 2048 | 286 | 253 | 1.31 | 912 | **63.9** |
| | #3 | OS2 | 2048 | 286 | 253 | 1.31 | 912 | **63.9** |

TABLE III. KISS-ICP parameters

| parameter | value |
|---|---|
| **Initial threshold** | 2m |
| **Min. deviation threshold** | 0.1m |
| **Max. points per voxel** | 20 |
| **Voxel size** | 1.0 |
| **Factor voxel size map merge** | 0.5 |
| **Factor voxel size registration** | 1.5 |
| **ICP convergence criterion** | $10^{-4}$ |

ground truth provided by the dataset. This approach makes it possible to compare odometry or SLAM that are based on different principles. It is a common approach to compare SLAM or odometry algorithms [18]. All trajectories share the same inertial frame. This facilitates direct comparison.

Popular metrics used to evaluate SLAM algorithms include Absolute Pose Error (APE) and Relative Pose Error (RPE) [11] [14] [16] [18]. APE is a metric based on comparing the absolute position of the robot at a given time stamp. It is useful for evaluating the global consistency of SLAM or odometry algorithms. When used as a absolute transformation from odometry to map frames, this metric is useful for evaluating localization precision. These frames and their definitions are described using the REP 105 [17] frame definition convention.

In contrast, Relative Pose Error (RPE) is a metric for comparing the local consistency of localization. RPE compares relative positions with respect to a reference trajectory. When KISS-ICP is used as one of the localization sources for fusion with other data, RPE is an appropriate metric for transformation from the base to the odometry frame.

The *evo: Python package for the evaluation of odometry and SLAM* [15] was used to compute the APE and RPE. Evo is a library designed for trajectory evaluation. It offers a range of functions such as metric evaluation, trajectory evaluation and visualization, export of graphs and results in various formats. It supports trajectory recording using ros2 bag, KITTI dataset [19], etc.

During the evaluation of APE and RPE, the mean, median and root mean square error (RMSE) were calculated to provide comprehensive insight into the performance of the localization algorithm.

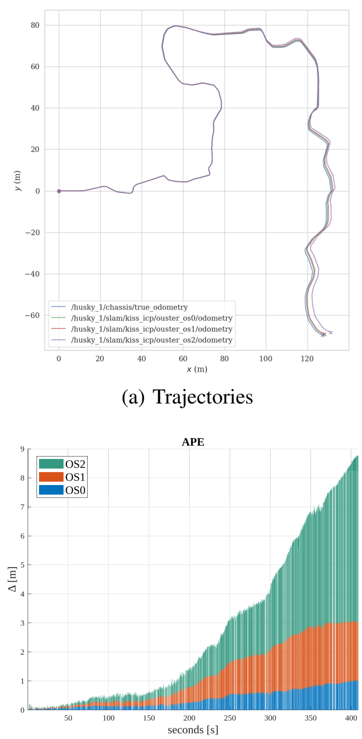
## B. Evaluation

The localization results of three instances of KISS-ICP lidar odometry along a reference trajectory are shown in the following Fig. 4. Specifically, it is trajectory #2 from the "Cave" environment with sensors OS0, OS1 and OS2. See TABLE II. The evolution of the APE during the experiment is shown in the subplot 4b. It is obvious that the absolute localization can easily and quickly deviate from the reference trajectory. Even a small miscalculation in orientation can lead to a massive error in the future trajectory. This is also evident from TABLE IV, which summarizes the mean and median values of RPE and APE.

More than the APE, we are interested in the accuracy of relative localization because it is not affected by multiplicative errors and better reflects the long-term accuracy of localization.

TABLE IV. Odometry evaluation

| Env. | Traj. | OS0 | | | | OS1 | | | | OS2 | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | RPE | | APE | | RPE | | APE | | RPE | | APE | |
| | | mean | median | mean | median | mean | median | mean | median | mean | median | mean | median |
| Cave | 1 | 0.0347 | 0.0298 | 15.3874 | 14.4067 | 0.0332 | 0.0271 | 44.4299 | 39.4062 | 0.0371 | 0.0309 | 14.3202 | 13.6992 |
| | 2 | 0.0294 | 0.0260 | 0.3984 | 0.2991 | 0.0326 | 0.0286 | 0.8748 | 0.6833 | 0.0372 | 0.0324 | 1.5347 | 0.8104 |
| | 3 | 0.0284 | 0.0247 | 0.0924 | 0.0795 | 0.0429 | 0.0281 | 8.3765 | 1.8791 | 0.0618 | 0.0368 | 3.5182 | 0.1353 |
| Park | 1 | 0.1012 | 0.0811 | 1.6171 | 1.4092 | 0.0849 | 0.0670 | 1.3316 | 1.2579 | 0.0889 | 0.0708 | 0.6590 | 0.6488 |
| | 2 | 0.0754 | 0.0588 | 2.2133 | 2.1037 | 0.0814 | 0.0591 | 0.4840 | 0.2782 | 0.0850 | 0.0517 | 3.1924 | 3.0398 |
| | 3 | 0.1374 | 0.0897 | 1.6284 | 0.5500 | 0.0919 | 0.0529 | 41.5783 | 14.1226 | 0.2087 | 0.0675 | 15.5273 | 2.9621 |




(b) APE

Fig. 4. KISS-ICP localization

The following Fig. 5 and 6 visualise the data from TABLE IV. They show the mean, median, and RMSE calculated from the RPE. The green bars represent the data from the experiments that were executed in the ”Cave”, while the orange bars represent the data from the experiments from the ”Park”.

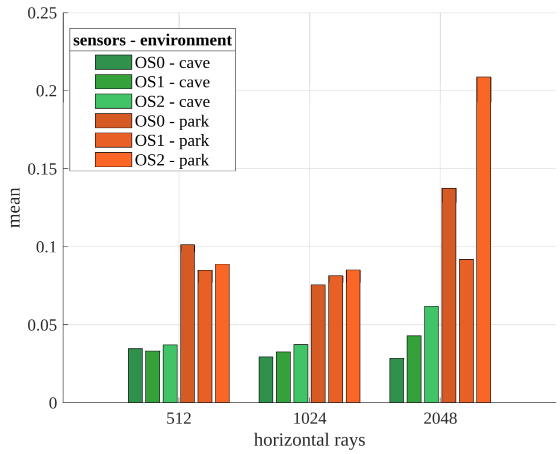
The experiment was designed to compare the behavior of lidar odometry with respect to sensor VFoV and row horizontal data resolution. Therefore, in Fig. 5, the data are grouped according to their horizontal resolution, and the individual bars illustrate the localization accuracy obtained with a specific sensor. In Fig. 6, the approach is reversed. The data is grouped by sensor type, and individual bars visualize their horizontal resolution.

The data clearly show that the accuracy of lidar odometry in ”Park” environments has a significantly higher relative error compared to data collected in ”Cave”. This highlights the profound influence of the environment on the localization accuracy. KISS-ICP is based on the Iterative Closest Point (ICP) algorithm, which minimizes the error between two point clouds. The characteristics of ”Cave” and ”Park” environments are significantly different. In the cave, the 3D lidar measures the roof of the cave. This results in a more spatially measured point cloud compared to the ”Park”, where the lidar only scans the terrain in the lower hemisphere. In addition, the ”Park” environment is much more monotonous than the ”Cave” environment. This further degrades the localization accuracy. It is easier to align two point clouds in a monotonous environment than in the unstructured terrain of a cave. This is because there aren’t as many variations in the points of the different features as there are in the cave. As a result, the comparison between the two point clouds will not be as accurate.

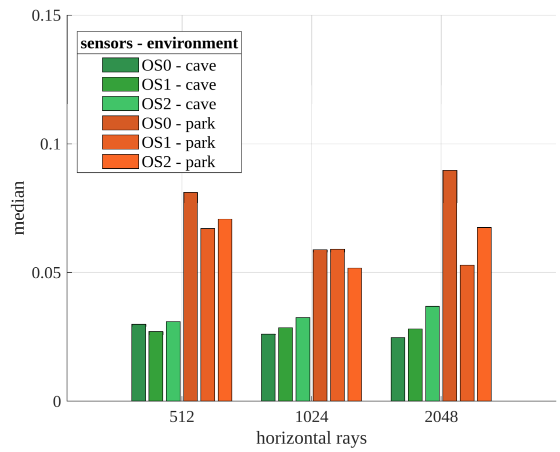
In contrast, the influence of the sensor’s field of view remains inconclusive. Fig. 5 shows approximately identical results for the OS0 sensor with a 90 VFoV and the OS2 sensor with a 22.5 VFoV. Similarly, the difference in point cloud density, i.e., horizontal resolution, is not obvious. The quality of the lidar odometry remains about the same with 512 points per row as with 2048 points per row. This is not surprising since both KISS-ICP algorithms filter the input point cloud twice. This approach is inspired by the CT-ICP algorithms. It makes the ICP algorithm more efficient and faster.

KISS-ICP lidar odometry performs equally well with sensors with small VFoV and relatively low horizontal resolution. The data from the TABLE IV shows a significant loss of scans from the OS2 sensor when stored in ros2 bags, about 50% for both environments. The mean and RMSE are significantly higher with the OS2 sensor in the park environment. However, no significantly worse localization is observed with the OS0 and OS1 sensors (with 2048 points per row) in a different environment and with a different VFoV.

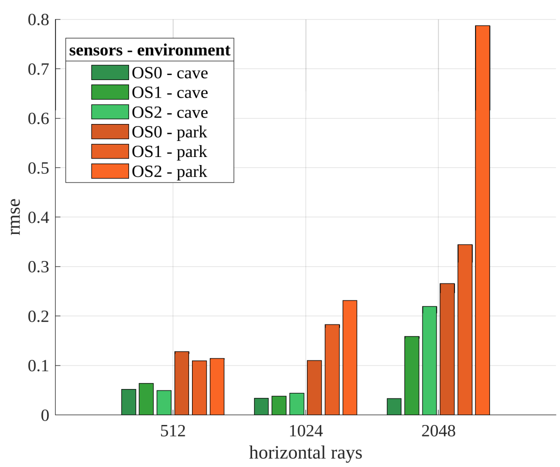
This can be attributed to the fact that a sensor with a lower VFoV filters more data at the input of the KISS-ICP algorithm because they are more densely distributed than sensors with a higher VFoV. With fewer points stored, the ICP algorithm is unable to correctly match the two point clouds, resulting in higher error rates. It’s clear that a low VFoV sensor, despite a high vertical point density, can lead to inaccurate localization by over-filtering points. After filtering, there isn’t enough spatial coverage of points left for the ICP algorithm to work properly. When the lidar data is lost or has a lower




(a) mean

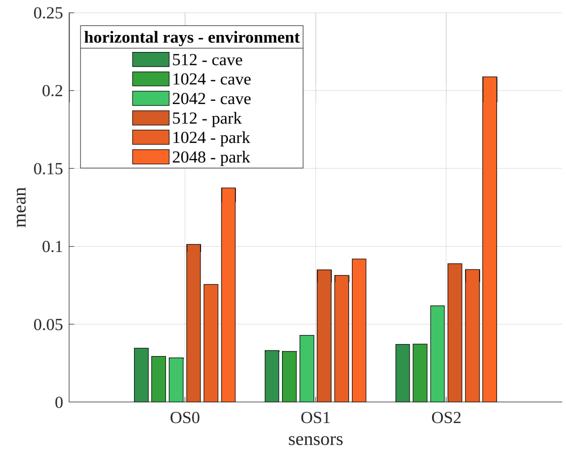



(b) median

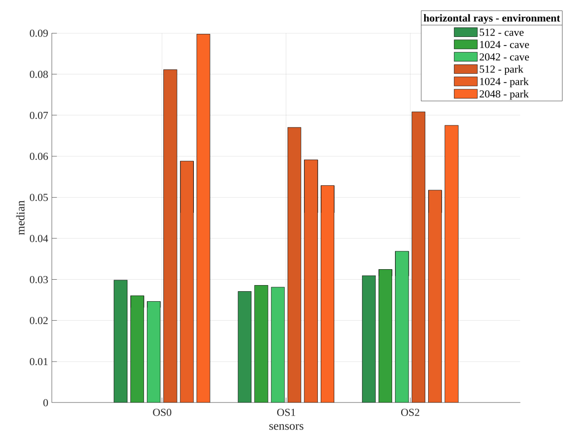



(c) rmse

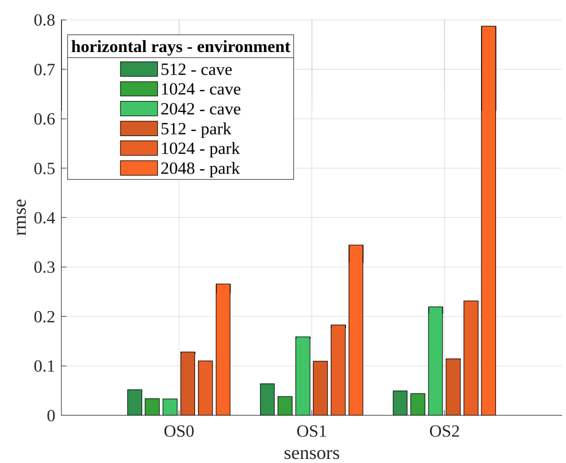
Fig. 5. RPE - grouped by number of points in a row




(a) mean




(b) median




(c) rmse

Fig. 6. RPE - grouped by sensors

frequency, the estimated position is significantly drifted. This is also indicated by a large RMSE value, see Fig. 5c and 6c.

Custom sensors with different numbers of rows were another group of sensors tested. Five sensors with a horizontal resolution of 1024 points per row and row numbers of 4, 8, 16, 32, and 64, as shown in TABLE II, were evaluated for localization quality. The motivation for this experiment was the determination of the localization quality of sensors with a lower point density than sensors such as Ouster. For instance, Velodyne Puck has exactly 16 rows. Fig. 7 and 8 shows the result.

The trajectories shown in Fig. 7 provide a clear illustration of how the trajectories degrade as the data becomes sparser. It's obvious that the localization performance decreases as the number of lidar rows decreases. As expected, localization using only 4 lidar rows gives the worst result. This result highlights the critical role of data density in ensuring accurate localization results. In contrast, the 64-row lidar gives the best result. The localization performance with both 64 and 32 rows is comparable to that of a 128-row sensor.

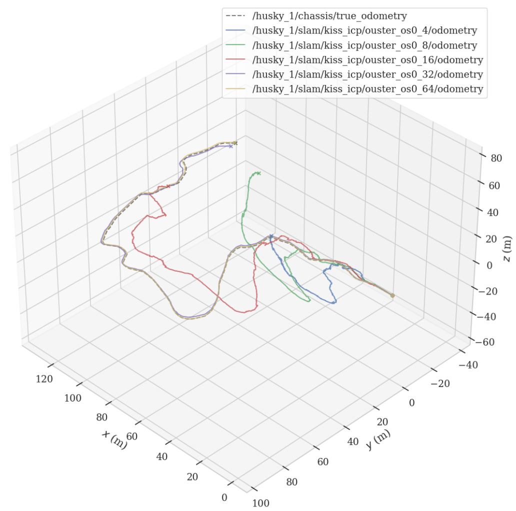
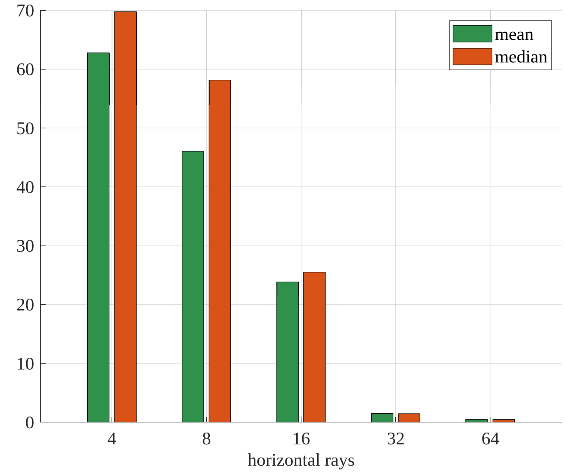
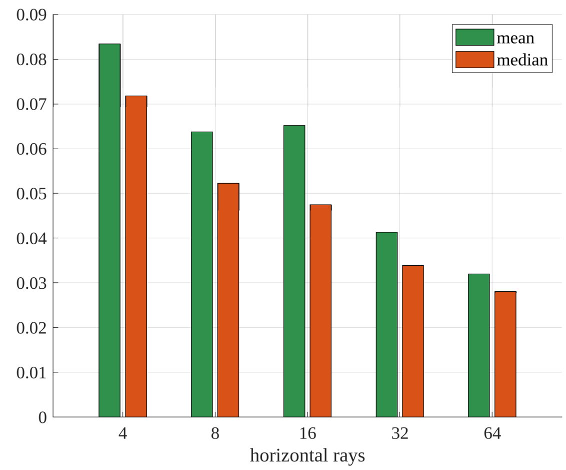



Fig. 7. KISS-ICP trajectory result with 4, 8, 16, 31 and 64 rows sensors

The 16-row lidar showed a noticeable jump in the absolute pose error (APE) 8. However, this may be due to random error, since the APE degrades rapidly even with small deviations of the robot from its actual position. The difference in the RPE values is not as significant, see Fig. 8. The 4-row lidar has only twice as much RPE error as the 64-row lidar. However, the number of points used by KISS-ICP is 16 times smaller. This implies that KISS-ICP is able to work with a very sparse point cloud.




(a) APE




(b) RPE

Fig. 8. KISS-ICP localization metrics with 4, 8, 16, 31 and 64 rows sensors

## V. Conclusion

The aim of this study was to evaluate the accuracy of localization using the KISS-ICP algorithm, a simple lidar odometry based on the ICP algorithm. A dataset of more than 9 km driven by the robot in two different environments using different sensors was collected for evaluation. The goal was to evaluate the lidar odometry with different VFoV sensors and data densities.

The results provided an evaluation of the localization accuracy. The performance of the algorithm is significantly worse in less structured environments and in areas where the lidar beams do not return any measurements due to the lack of obstacles in the range.

It was found that the KISS-ICP algorithm performs very well even with sparse data and does not require high sensor resolution due to dual subsampling. The only potential problem for the algorithm may be a low VFoV sensor with a lower frequency of incoming data.

Further research could focus on investigating the impact of adjustable parameters on specific types of environments, as well as a possible evaluation of localization accuracy through fusion with IMU and other odometry source data.

## Acknowledgment


The completion of this paper was made possible by the grant No. FEKT-S-23-8451 - ”Research on advanced methods and technologies in cybernetics, robotics, artificial intelligence, automation and measurement” financially supported by the Internal science fund of Brno University of Technology.


## References


[1] I. Vizzo, T. Guadagnino, B. Mersch, L. Wiesmann, J. Behley, and C. Stachniss, “KISS-ICP: In Defense of Point-to-Point ICP – Simple, Accurate, and Robust Registration If Done the Right Way”, IEEE Robotics and Automation Letters, vol. 8, no. 2, pp. 1029-1036, 2023.

[2] P. Dellenbach, J. -E. Deschaud, B. Jacquet, and F. Goulette, “CT-ICP: Real-time Elastic LiDAR Odometry with Loop Closure”, Proc. IEEE Int. Conf. Robot. Automat., vol. 2022, pp. 5580–5586.

[3] P. J. Besl and N. D. McKay, “A method for registration of 3-D shapes”, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 14, no. 2, pp. 239-256, 1992.
[4] Y. Shenghai, W. Han, and X. Lihua, “Survey on localization systems and algorithms for unmanned systems”, Unmanned Systems, 2020.
[5] J. Zhang and S. Singh, “Low-drift and real-time lidar odometry and mapping,” Autonomous Robots, vol. 41, no. 2, pp. 401–416, Feb. 2017, doi: 10.1007/s10514-016-9548-2.
[6] Shan, T., Englot, B. (2018). LeGO-LOAM: Lightweight and Ground-Optimized Lidar Odometry and Mapping on Variable Terrain. In 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) (pp. 4758-4765).
[7] G. Xue, J. Wei, R. Li, and J. Chen, “LeGO-LOAM-SC: An Improved Simultaneous Localization and Mapping Method Fusing LeGO-LOAM and Scan Context for Underground Coalmine”, Sensors, vol. 22, no. 2, 2022.
[8] Jens Behley, C. Stachniss (2018), ”Efficient Surfel-Based SLAM using 3D Laser Range Data in Urban Environments:, Robotics: Science and Systems XIV.
[9] Y. Pan, P. Xiao, Y. He, Z. Shao, and Z. Li, “MULLS: Versatile LiDAR SLAM via multi-metric linear least square,” in Proc. IEEE Int. Conf. Robot. Automat., 2021, pp. 11633–11640.
[10] T. Shan, B. Englot, D. Meyers, W. Wang, C. Ratti, and D. Rus, “LIO-SAM tightly-coupled LiDAR inertial odometry via smoothing and mapping,” in Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst., 2020, pp. 5135–5142.
[11] R. Kümmerle, B. Steder, C. Dornhege, M. Ruhnke, G. Grisetti, C. Stachniss, and A. Kleiner, “On measuring the accuracy of SLAM algorithms”, Autonomous Robots, vol. 27, no. 4, pp. 387-407, 2009.
[12] S. Macenski, T. Foote, B. Gerkey, C. Lalancette, and W. Woodall, “Robot Operating System 2: Design, architecture, and uses in the wild”, Science Robotics, vol. 7, no. 66, May 2022.
[13] P. Gabrlik, T. Lazna, T. Jilek, P. Sladek, and L. Zalud, “An automated heterogeneous robotic system for radiation surveys: Design and field testing”, Journal of Field Robotics, vol. 38, pp. 657 - 683, 2021.
[14] D. Prokhorov, D. Zhukov, O. Barinova, K. Anton, and A. Vorontsova, “Measuring robustness of Visual SLAM”, in 2019 16th International Conference on Machine Vision Applications (MVA), 2019, pp. 1-6.
[15] M. Grupp, ”evo: Python package for the evaluation of odometry and SLAM.,” , 2017.
[16] B. Garigipati, N. Strokina, and R. Ghabcheloo, “Evaluation and comparison of eight popular Lidar and Visual SLAM algorithms”, in 2022 25th International Conference on Information Fusion (FUSION), 2022, pp. 1-8.
[17] W. Meeussen, “Coordinate Frames for Mobile Platforms”, ROS - Robot Operating System. .
[18] R. Kümmerle et al., “On measuring the accuracy of SLAM algorithms,” Autonomous Robots, vol. 27, no. 4, pp. 387–407, Nov. 2009, doi: 10.1007/s10514-009-9155-6.
[19] A. Geiger, P. Lenz, C. Stiller, and R. Urtasun, “Vision meets robotics: The KITTI dataset,” Int. J. Robot. Res., vol. 32 no. 11, pp. 1231–1237, 2013.

# Jsme skupina E.ON, těší nás

Jsme jedním z největších energetických koncernů u nás i ve světě. Centrálu máme v Německu, ale najdete nás už v 15 zemích Evropy, včetně České republiky. Zakládáme si na tom, že naše energie můžou šetřit peníze i přírodu a čím dál víc využíváme obnovitelné zdroje.

## Kdo patří do naší rodiny

**E.ON Energie, a.s.**

- Obchoduje s elektřinou a plynem, zajišťuje marketing a komunikaci a stará se i o výrobu energií.
- Pro zákazníky připravuje řešení na míru v oblasti fotovoltaiky, tepelné techniky a elektromobility.

**EG.D, a.s.**

- Distributor energií, který vlastní a provozuje rozvodnou síť elektřiny zejména na jihu Čech a Moravy a rozvodnou síť plynu na jihu Čech.
- Zajišťuje připojení odběrných míst k síti a stará se o dopravu energií k zákazníkům.

**E.ON Česká republika, s. r. o.**

- Funguje jako podpora výše uvedeným společnostem. Zajišťuje jim služby, jako je účetnictví, právo nebo HR, a na starost má i zákaznickou péči.

## Na přírodě nám záleží

E.ON Energie je jedničkou ve výkupu zelené elektřiny v České republice.

EG.D pomocí bezpečnostních prvků na sloupech vysokého napětí chrání ptáky před úrazem.

V budovách E.ONu i ve všech dobíječkách elektromobilů využíváme zelenou energii.

e·on

# Nastartuj svou kariéru

**#spolujsmeeon**

**Studentské programy**

- Letní energetická akademie – poslední týden v červenci
- Letní technické brigády – od června do září
- Stipendijní program
- Diplomové/bakalářské práce
- Praxe a brigády

**Absolventské programy**
(plný úvazek)

- Trainee program – od září
- Junior technik – od září

**eon.cz/kariera**

E.ON Kariéra

E.ON Kariéra

# Obstacle Avoidance in UAVs: Using a Bug-Inspired Algorithm and Neural Network-Based RGB Camera Collision Prediction

1st Petr Raichl
*Dept. of Theoretical and Experimental Electrical Engineering*
*FEEC, Brno University of Technology*
Brno, Czech Republic
xraich02@vutbr.cz

2nd Petr Marcon
*Dept. of Theoretical and Experimental Electrical Engineering*
*FEEC, Brno University of Technology*
Brno, Czech Republic
marcon@vut.cz

3rd Jiri Janousek
*Dept. of Theoretical and Experimental Electrical Engineering*
*FEEC, Brno University of Technology*
Brno, Czech Republic
xjanou09@vutbr.cz

***Abstract*—Unmanned Aerial Vehicles (UAVs) are increasingly deployed in complex environments for various applications, necessitating advanced obstacle avoidance capabilities to ensure safety and mission success. Inspired by the simplicity and effectiveness of biological navigation strategies, this study introduces a novel approach to UAV obstacle avoidance, leveraging the principles of the bug algorithm combined with the predictive power of neural networks. We propose a hybrid model that integrates a lightweight neural network to predict potential collisions in real-time. Our methodology employs a two-stage process: first, the neural network assesses the immediate risk of collision; second, the bug algorithm-inspired decision-making process determines the UAV's maneuvering actions to navigate without crashing to obstacles. The approach was tested both in simulation and real outdoor experiments.**

***Index Terms*—UAV, Obstacle avoidance, Collision prediction, Artificial intelligence, Neural Networks**

## I. Introduction

In the rapidly evolving field of Unmanned Aerial Vehicles (UAVs), the ability to autonomously navigate through complex and dynamic environments is paramount. UAVs are employed in a wide array of applications, from aerial photography and surveying to search and rescue operations and agricultural monitoring, each presenting unique challenges in terms of navigation and obstacle avoidance. Traditional obstacle avoidance systems, while effective in certain scenarios, often struggle in environments characterized by unpredictability and the need for real-time decision-making.

Biological systems, particularly insects, exhibit remarkable navigation capabilities, efficiently avoiding obstacles in dynamic and complex environments. This observation has spurred interest in bio-inspired algorithms that mimic these natural strategies for robotic applications. Among these, the bug algorithm stands out for its simplicity and effectiveness in pathfinding and obstacle avoidance. However, the direct application of this algorithm to UAV obstacle avoidance is limited by its lack of predictive capability, which is crucial for navigating at higher speeds and in more densely populated environments.

To address these limitations, this study introduces a novel, hybrid model that combines the principles of the bug algorithm with the predictive power of neural networks. By integrating a lightweight neural network, our approach aims to predict potential collisions in real-time, leveraging sensor data that simulates the sensory perception mechanisms found in bugs. This predictive capability allows for the anticipation of obstacles and the planning of avoidance maneuvers in advance, significantly enhancing safety and efficiency.

The proposed model is trained on a diverse dataset of simulated environments, enabling it to recognize and react to a wide range of obstacle configurations. Through a two-stage process, the UAV first uses the neural network to assess the immediate risk of collision. Subsequently, the bug algorithm-inspired decision-making process guides the UAV in maneuvering around the obstacle, ensuring safe navigation in real-time.

The following sections will detail the development of this hybrid model, its evaluation, and the implications of our findings for the future of UAV navigation and obstacle avoidance.

## II. Related work

The literature showcases a diverse range of strategies for drone navigation and obstacle avoidance. Broadly speaking, these approaches vary based on the type of sensory data they utilize and the processing techniques applied to guide the aerial platform. For outdoor operation, a UAV typically employs GPS, range-finding, Lidars and visual sensors. These tools are instrumental in determining the drone's status, detecting obstacles, and facilitating the path-planning process.

In this work, we are focusing on Reactive collision avoidance. In the reactive collision avoidance methodology, the UAV dynamically determines and implements a collision avoidance strategy in real-time, utilizing data about its immediate surroundings captured through onboard sensors. This approach enables quick adaptation to abrupt environmental alterations, demanding minimal computational resources and pre-existing knowledge of the environment's layout. Due to these characteristics, reactive strategies offer significant advantages in changing environments over proactive planning methods, where the UAV relies on pre-mapped information to navigate around known obstacles during the planning phase [1], [2], [3].

Among the most known methods used for path planning and obstacle avoidance is Artificial Potential Field (APF). This method has gained great popularity because it is simple to implement and provides acceptable results. But also there are known drawbacks, including oscillation, Non Reachable Goals and Local Minima. The method has been modified several times to reduce the impact or eliminate known problems. [4] and [5] combine APF with bug algorithm.

A special algorithm family named the Bug algorithms is also used for its simple implementation and elimination of some problems discussed before. In Bug algorithms the state of the robot switches between 2 states: moving toward the goal, following the boundary of obstacles. If robot moves towards the goal and no obstacle is located between the robot and the goal, the robot stays in the moving forward state. If the robot identifies an obstacle in front of him then the state switches to following the boundary of the obstacle. Then it follows the obstacle until the condition for switching into a move forward state is true. The problem of this method is that the robot stays close to an obstacle instead of safe trajectory optimization. Also depending on existing following boundary strategy the algorithm can be non-optimal in terms of traveled distance. [6]. Many works adopted some version of bug algorithms, e.g. [7], [8], [9], [10]

Contrary to the classical approach for obstacle avoidance in UAVs, there is a growing body of research exploring the use of artificial intelligence, particularly deep reinforcement learning (DRL) [11], [12] and Supervised Neural Networks [13], [14], [15], [16]. This innovative approach leverages the capabilities of DRL to enable UAVs to learn from their interactions with the environment, improving their decision-making processes for navigating around obstacles autonomously. Through continuous learning and adaptation, UAVs can develop sophisticated strategies for obstacle avoidance that are not predefined by human operators, showcasing the potential of AI to enhance UAV navigational systems beyond traditional methods. As DRL presents innovative opportunities, it also has its own disadvantages. DRL algorithms require extensive training periods, large amounts of training data, and computational resources. Also transferring a trained model from a simulated environment where training is done to real-world applications can be challenging.

Supervised learning presents a feasible approach for training control policies, yet it crucially hinges on the quality of the expert guidance it seeks to emulate. This guidance could be derived from a human expert or predefined trajectories [17]. However, gathering a comprehensive dataset of expert trajectories or assessing the performance of partially trained policies can be both labor-intensive and risky. Furthermore, the discrepancy between the conditions experienced by the expert and those encountered by the agent—a phenomenon known as domain shift—may impede the ability of supervised learning techniques to generalize effectively across different scenarios. In [13] supervised learning is used to create system that navigates a small quadrotor helicopter autonomously at low altitude through natural forest environments using only a single cheap camera to perceive the environment. [trailnet] uses supervised learning to estimate position of UAV according to the trail and based on the estimated position control action is computed to make the UAV to follow the trail. [14] uses supervised learning to predict steering angle.

[18] combines output of DRL and Resnet18 neural network. One model is used for pathfinding and the second model for obstacle avoidance. The output of both algorithms is combined to change the direction of the quad-copter adaptively with indoor environments. However the experiment was taken only in a simulation environment.

As presented above, our approach aims to integrate the strengths of multiple methodologies, especially the bug algorithms in combination with AI abilities to process information from vision-based sensors, to optimize UAV obstacle avoidance.

## III. Methodology

### A. AI Model

Our development of a neural network model for collision prediction was inspired by work [19]. Their innovative approach utilized a Resnet8 architecture to predict collisions. In our work, we adopt a more complex neural network architecture Resnet18 to enhance the accuracy and reliability of collision prediction. The choice of Resnet18 was driven by its deeper structure, which offers a more profound capacity for feature extraction from the visual data captured during UAV flights. Also, this choice makes a compromise between neural network deep and inference time. For training, we used MATLAB 2023 with Deep Learning Toolbox. We employed transfer learning and fine-tuning techniques to train our model. The results of training are presented later.

### B. Dataset

For collision prediction training from images, we use a publicly available dataset from [19]. The dataset contains about 30,000 images annotated collision or not collision. The dataset was created from the camera which was mounted on the handlebar of bicycle. The dataset contains data from various situations recorded along different areas of the city - vehicles, pedestrians, and vegetation.

### C. Steering algorithm

To highlight a key differentiation in our approach compared to the work [19]. While this work integrated a specific AI model for steering UAV, our project uses a different approach for the steering mechanism. For steering UAVs to ensure safe navigation without collisions, we have adopted an approach inspired by the bug algorithms described above. This decision was motivated by the desire to explore alternative, potentially more effective methods for UAV real-time navigation in complex environments. Our approach emphasizes more procedural response to detected obstacles. The UAV proceeds to move forward during its flight. AI model scans the environment for potential obstacles. When there is potential collision detected, AI model immediately signals the UAV to stop its forward movement. Then the uAV initiates a twist maneuver around its yaw axis. While twisting, the AI model continues to analyze the environment. Once the model determines that there is no longer an obstacle in its forward path, the UAV proceeds to move forward again. The rationale behind exploring a different steering approach lies in the potential for enhanced safety and reliability. By implementing a method that temporarily stops the UAV and evaluates its environment before proceeding, we aim to reduce the risk of collisions.

## IV. SW/HW ARCHITECTURE

In this section we describe HW platform, cameras used for obtaining image data and SW stack we used for flying the UAV.

### A. UAV

Image of UAV used in our work is in Fig. 1. The size of the UAV is similar to the commonly known DJI Matrice 300. HW setup contains open source Pixhawk autopilot, onboard computer Intel NUC and siyi a8 mini camera.

Fig. 1. UAV used in this work

### B. SW stack

Figure 2 presents a diagram of our software architecture. Our platform seamlessly incorporates the PX4 flight stack as the chosen firmware for the Pixhawk autopilot system. PX4 stands out for its adaptability and reliability, offering comprehensive support for a wide array of UAV configurations. This flexibility is a cornerstone of our system's architecture, enabling us to tailor our operations to meet diverse application needs. One of the key features of PX4 that we leverage extensively is its software-in-the-loop (SITL) simulation environment. The SITL functionality is crucial for our development process, providing us with a robust framework for testing and debugging our control algorithms in a simulated environment before deployment in real-world scenarios.

For communication between the onboard computer and the PX4 system, our setup employs the MAVLink protocol, the industry standard for UAV communication. This protocol ensures a reliable and efficient exchange of command and control messages between the two systems, facilitating seamless integration and coordination of flight operations. To streamline the development process and avoid the complexity of directly handling MAVLink messages, we utilize the MAVSDK library. MAVSDK, a standard C++ library designed specifically for UAV applications, abstracts the intricacies of MAVLink communication into a more accessible and developer-friendly interface. This library acts as a middleware, translating the output from our ROS2-based software into MAVLink messages that the PX4 flight stack can interpret directly.

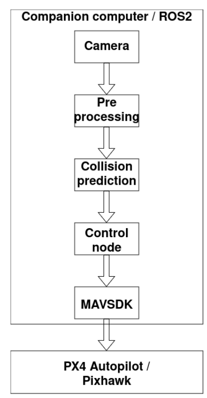



Fig. 2. Block diagram of SW architecture

The onboard computing capabilities are provided by an Intel NUC, equipped with the Ubuntu 22.04 Jammy Jellyfish operating system. This choice of operating system supports our software development needs, particularly for robotics applications. Our software is developed entirely in ROS2 Humble, an open-source robotics middleware, which offers a flexible framework for writing robot software. Within the ROS2 ecosystem, we have designed the whole pipeline that enables the control of a robot, specifically UAVs. One ROS2 node is dedicated to handling image data. This node's primary function is to receive image data from the camera and publish it at a predetermined rate, ensuring a steady flow of visual information for processing. We employ another ROS2 node specifically to load and manage neural network models. This node is intricately designed to receive the incoming image data and perform inference using the neural network models. The process includes several steps of image preprocessing, such as resizing and normalization, which are crucial for the models to function correctly. These preprocessing tasks are accomplished using the OpenCV DNN module, a comprehensive tool for deep learning in computer vision, enabling efficient image manipulation and preparation for inference tasks. Furthermore, our architecture includes a separate node dedicated to the control of the drone, leveraging the capabilities of the MAVSDK library. This library facilitates communication with the drone, allowing for precise control over its operations. We operate the drone in offboard mode, a configuration that enables remote control over its movements. The use of MAVSDK is critical in this aspect, as it ensures that control commands are published at the necessary rate to maintain the drone in offboard mode, thereby providing us with the flexibility to execute complex flight patterns and maneuvers based on the inference results from the neural network models.

## V. Experiments and Results

In this segment, we present both numerical and descriptive outcomes of our suggested approach. Initially, we assess Neural Network precision using various performance indicators. Subsequently, we examine our approach ability to steer UAVs safely, avoiding collisions with obstacles - in simulational and real outdoor experiment.

### A. Neural Network Training

The parameters and results of training of our model are shown in table I. The precision metric is metric from MATLAB training presenting cross-entropy loss function computed on testing data - meaning the data which were not used for training and validation during the training process. For a better understanding of how our model is working we present a confusion matrix, see Fig. 3.

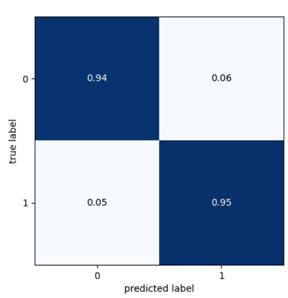



Fig. 3. Confusion matrix: 1 - collision, 0 - no collision

### B. Simulation

We conducted experiment in well well-known robotics simulator Gazebo together with PX4 SITL before we tested our approach on real UAV in outdoor. This simulation allows us to test the whole pipeline before even taking UAVs out of the laboratory. Even though there exists domain shift, we were able to verify that our approach is able to navigate UAVs without collisions - meaning that the neural network is able to predict collisions on simulated camera data and that the steering approach is able to move the UAV without collision.

### C. Outdoor experiment

We conducted an outdoor experiment to evaluate the performance of a model designed for predicting collisions of UAVs with obstacles. This test was aimed at detecting potential collisions with people, walls and trees under real-world conditions.

The experiment was successful for the most part. The proposed neural network model is capable of accurately predicting collisions with people and walls in real time and the steering approach is able to avoid collisions. Fig. 4 shows an example of predictions when flying towards a wall. Fig. 6 and 7 illustrate the dynamic process of collision detection and the preventive maneuvers executed by the UAV to ensure safe navigation.

However, the results regarding collision predictions with trees highlighted room for improvement. The model predicted collisions with trees at a late stage when the UAV was too close to them to initiate a safe maneuver, see Fig. 5.

Despite the described limitation, the proposed algorithm for steering the UAV displayed good performance. Whenever a collision was predicted, the algorithm successfully navigated the UAV away from the obstacles, preventing collisions.


Fig. 4. UAV flying towards a wall


Fig. 5. UAV flying towards a tree

## VI. CONCLUSION AND DISCUSSION

In summary, we successfully trained a neural network model to predict collision based on RGB camera input. Additionally, we implemented simple bug-inspired algorithm which in cooperation with neural network collision prediction enables safety navigation of UAVs without collisions. The functionality and reliability of the presented navigation approach were rigorously tested in simulation and outdoor experiments - the results were described above. However, there are areas for improvement. Especially, predictions of collisions with trees is delayed to be considered safe. To enhance the performance and reliability of the presented model, the extension of the training dataset with specific classes could improve classification accuracy and make navigation of UAVs safer.

| Model | Parameters [mil.] | Depth | Inference time [ms] | Precision [%] |
|---|---|---|---|---|
| Resnet18 | 11.7 | 18 | 17 | 96 |

TABLE I
THE RESULTS OF CNN TRAINING


Fig. 6. UAV flying towards a person and then avoiding collision


Fig. 7. UAV flying towards a group of people and then avoiding collision


## ACKNOWLEDGMENT

The research was supported of the general student development project being executed at Brno University of Technology.


## REFERENCES


[1] J. Park, N. Cho, and S. Lee, "Reactive Collision Avoidance Algorithm for UAV Using Bounding Tube Against Multiple Moving Obstacles", *IEEE Access*, vol. 8, pp. 218131-218144, 2020.

[2] M. Gnanasekera and J. Katupitiya, "A Time Optimal Reactive Collision Avoidance Method for UAVs Based on a Modified Collision Cone Approach", in *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2020, pp. 5685-5692.

[3] L. Lu, C. Sampedro, J. Rodriguez-Vazquez, and P. Campoy, "Laser-based Collision Avoidance and Reactive Navigation using RRT* and Signed Distance Field for Multirotor UAVs", in *2019 International Conference on Unmanned Aircraft Systems (ICUAS)*, 2019, pp. 1209-1217.

[4] Mei Wang, Zhiyong Su, Dawei Tu, and Xichang Lu, "A hybrid algorithm based on Artificial Potential Field and BUG for path planning of mobile robot", in *Proceedings of 2013 2nd International Conference on Measurement, Information and Control*, 2013, pp. 1393-1398.

[5] H. Wang, M. Cao, H. Jiang, and L. Xie, "Feasible Computationally Efficient Path Planning for UAV Collision Avoidance", in *2018 IEEE 14th International Conference on Control and Automation (ICCA)*, 2018, pp. 576-581.

[6] Y. Zhu, T. Zhang, J. Song, and X. Li, "A new Bug-type navigation algorithm considering practical implementation issues for mobile robots", in *2010 IEEE International Conference on Robotics and Biomimetics*, 2010, pp. 531-536.

[7] S. Mukherjee, R. Kumar, and S. Borah, "Computational Study of DFMB Algorithm for Unmanned Vehicle in Static 2D Environment", in *2020 International Conference on Computational Performance Evaluation (ComPE)*, 2020, pp. 126-131.

[8] B. P. Duisterhof, S. Li, J. Burgues, V. J. Reddi, and G. C. H. E. de Croon, "Sniffy Bug: A Fully Autonomous Swarm of Gas-Seeking Nano Quadcopters in Cluttered Environments", in *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2021, pp. 9099-9106.

[9] S. S, D. A. Nandesh, R. Raman K, G. K, and R. R, "An Investigation of Bug Algorithms for Mobile Robot Navigation and Obstacle Avoidance in Two-Dimensional Unknown Static Environments", in *2021 International Conference on Communication information and Computing Technology (ICCICT)*, 2021, pp. 1-6.

[10] E. F. Mohamed, K. El-Metwally, and A. R. Hanafy, "An improved Tangent Bug method integrated with artificial potential field for multi-robot path planning", in *2011 International Symposium on Innovations in Intelligent Systems and Applications*, 2011, pp. 555-559.

[11] Y. Chen, N. Gonzalez-Prelcic, and R. W. Heath, "Collision-Free UAV Navigation with a Monocular Camera Using Deep Reinforcement Learning", in *2020 IEEE 30th International Workshop on Machine Learning for Signal Processing (MLSP)*, 2020, pp. 1-6.

[12] S. Ouahouah, M. Bagaa, J. Prados-Garzon, and T. Taleb, "Deep-Reinforcement-Learning-Based Collision Avoidance in UAV Environment", *IEEE Internet of Things Journal, vol. 9, no. 6, pp. 4015-4030*, Mar. 2022.

[13] S. Ross, N. Melik-Barkhudarov, K. S. Shankar, A. Wendel, D. Dey, J. A. Bagnell, and M. Hebert, "Learning monocular reactive UAV control in cluttered natural environments", in *2013 IEEE International Conference on Robotics and Automation*, 2013, pp. 1765-1772.

[14] H. Xu, Y. Gao, F. Yu, and T. Darrell, "End-to-end learning of driving models from large-scale video datasets," in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recog.*, July 2017.

[15] N. Smolyanskiy, A. Kamenev, J. Smith, and S. Birchfield, "Toward low-flying autonomous MAV trail navigation using deep neural networks for environmental awareness", *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 4241-4247, 2017.

[16] M. Petho, A. Nagy, and T. Zsedrovits, "A bio-motivated vision system and artificial neural network for autonomous UAV obstacle avoidance", in *2020 3rd International Seminar on Research of Information Technology and Intelligent Systems (ISRITI)*, 2020, pp. 632-637.

[17] D. Gandhi, L. Pinto, and A. Gupta, "Learning to fly by crashing", in *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2017, pp. 3948-3955.

[18] N. -T. Duc, Q. -T. Hai, D. -N. Van, T. -H. Trong, and T. -D. Trong, "An approach for UAV indoor obstacle avoidance based on AI technique with ensemble of ResNet8 and Res-DQN", in *2019 6th NAFOSTED Conference on Information and Computer Science (NICS)*, 2019, pp. 330-335.

[19] A. Loquercio, A. I. Maqueda, C. R. del-Blanco, and D. Scaramuzza, "DroNet: Learning to Fly by Driving", *IEEE Robotics and Automation Letters*, vol. 3, no. 2, pp. 1088-1095, 2018.

# Hybrid Flyback Resonant Converter

Radek Tomanek
*Department of Power Electrical and Electronic Engineering*
*Faculty of Electrical Engineering and Communications*
*Brno University of Technology*
Brno, Czechia
Radek.Tomanek@vut.cz

***Abstract*—This paper describes a flyback converter that uses a series-resonant circuit to achieve zero-voltage switching to reduce electromagnetic interference to the environment and reduce switching losses in transistors.**

***Index Terms*—flyback, resonant converter, DC/DC converter, transformer-based power supply, switched-mode power supply, zero-voltage switching, ZVS**

## I. Introduction

Blocking converters are the most commonly used topology, especially in low power AC/DC adapters - typically up to 200 W. Their main advantage is simplicity and the ability to achieve a relatively high power density. In some applications, they take advantage of the fact that they allow for a relatively simple multi-output design [5], [6].

The main disadvantages of this topology include the turning on transistor under full voltage and turning it off under full current. This causes switching losses in the transistor. The steep edges of the current through the primary and secondary windings enrich its spectrum, which increases the AC resistance of the winding, and hence the losses in the winding. Last but not least, this hard switching poses an electromagnetic interference problem to the environment. Furthermore, in a conventional flyback, it is necessary to address the demagnetization of the leakage inductance, which is often solved in a lossy manner [3], [4], [7].

These drawbacks are eliminated by the hybrid flyback resonant converter due to zero-voltage switching (ZVS) and more favorable waveforms of currents through the primary and secondary windings, which will become apparent later. leakage inductance is used for resonance in this topology, so it is not an issue [2], [4].

## II. Topology

The topology of the hybrid flyback resonant converter is shown in Fig. 1. As you can see, the secondary side is connected in the same way as in a classic flyback converter. The primary side is excited from a transistor half-bridge ($Q_1$, $Q_2$) and a resonant capacitor $C_r$ is included in series with the primary winding of the transformer T [1]–[3].

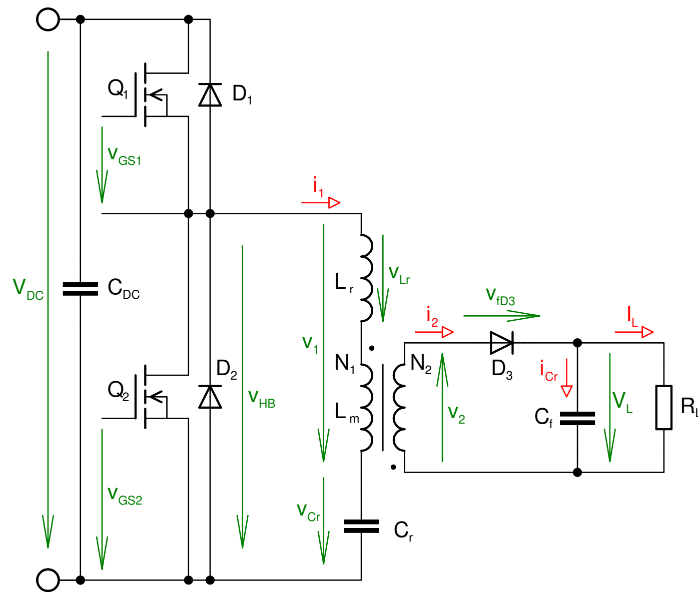



Fig. 1. Principle circuit diagram of hybrid flyback resonant converter [2], [3].

## III. Converter principle

The waveforms of important variables are shown in Fig. 2. They were obtained by simulation in Matlab Simulink with the parameters listed in TABLE I. The description of the behaviour is divided into individual time intervals that are marked in the waveforms. The current and voltage directions for the subtime intervals are also shown in the corresponding diagrams in Fig. 3 to Fig. 10.

### A. Interval $t_0 - t_1$

In this interval, the upper transistor is turned on and the magnetization of the magnetizing inductance of the transformer $L_\mathrm{m}$ takes place. The steepness of the current increase in this interval is determined both by the magnetizing inductance $L_\mathrm{m}$, which is approximately constant, and by the difference

Author gratefully acknowledge financial support from the Ministry of Education, Youth and Sports together with ECSEL—project No. 101007237/8A21002 Trusted European SiC Value Chain for a greener Economy (TRANSFORM) and the specific research programme of the BUT (project No. FEKT-S-23-8430). Author acknowledges Assoc. Prof. Pavel Vorel, Ph.D. (vorel@vut.cz) as my research supervisor and Assistant Prof. Jan Martiš, Ph.D. (martis@vut.cz) for his valuable comments and help while writing this paper.

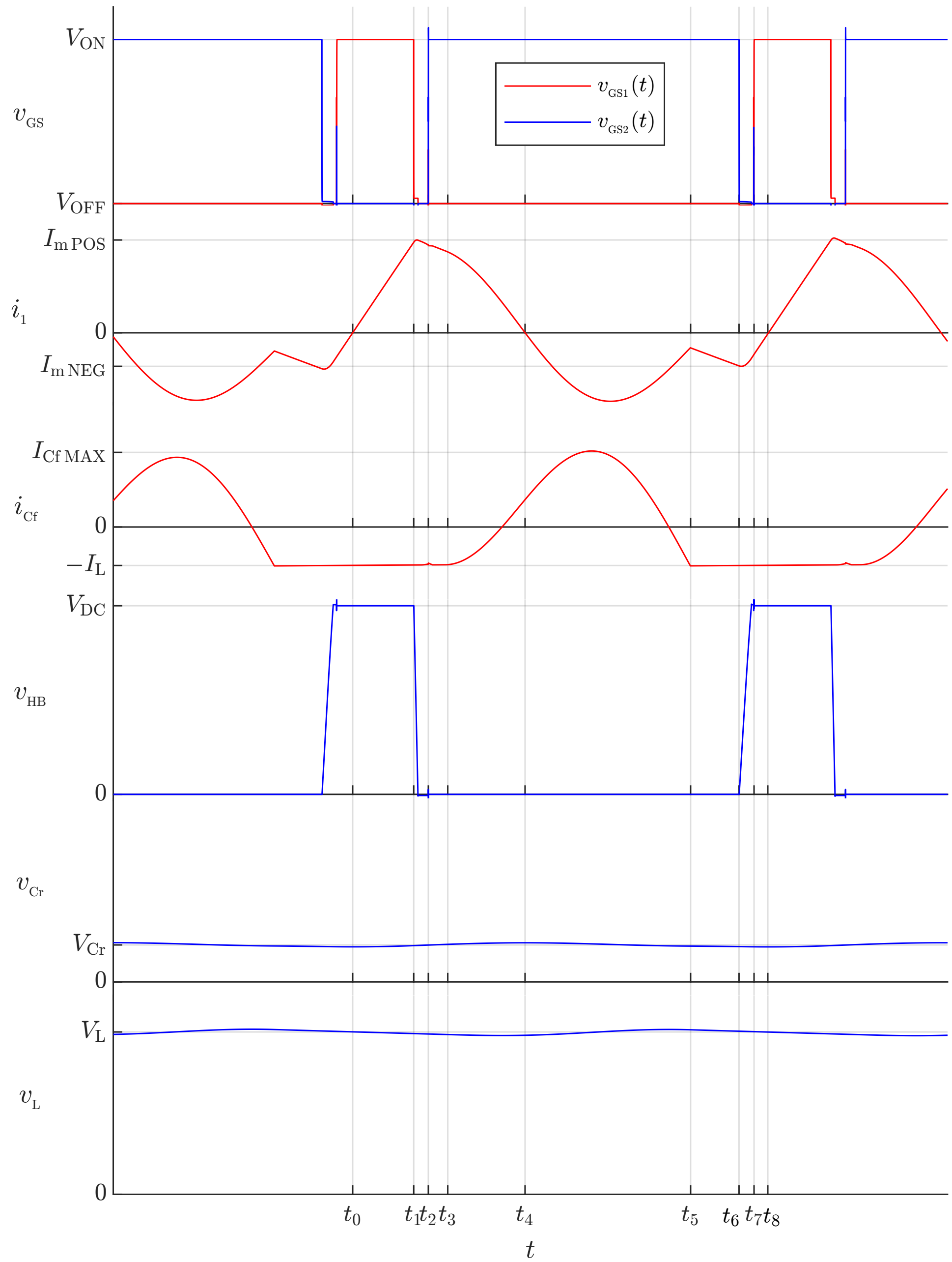



Fig. 2. Simulated waveforms with full load.

between the voltage on the resonant capacitor $v_{\mathrm{CR}}$ and the supply voltage $V_{\mathrm{DC}}$. If the converter supply voltage $V_{\mathrm{DC}}$ is significantly higher than the voltage change across the resonant capacitor $\Delta V_{\mathrm{CR}}$ during this interval, then the current growth $i_{\mathrm{Lm}}$ is nearly linear.

During the magnetization of the transformer, there is a positive voltage $v_1$ on the primary winding. This voltage is transformed into a positive voltage $v_2$. After adding the positive voltage $v_2$ with an arbitrarily large positive voltage of the load $V_{\mathrm{L}}$, the secondary diode $\mathrm{D}_3$ will be reliably polarized in the closing direction. Thus, no current can flow through the secondary winding in this interval, and therefore the transformer can only be viewed as a coil consisting of

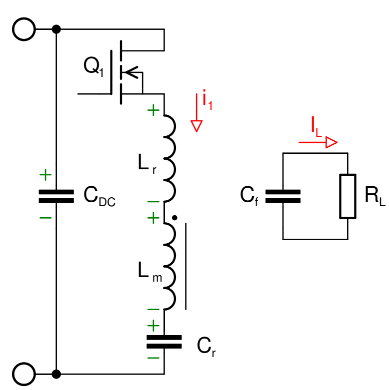



Fig. 3. Circuit behaviour during the interval $t_0$ - $t_1$.

TABLE I
PARAMETERS DURING SIMULATION.

| Name | Quantity | Value |
|---|---|---|
| Switching frequency | $f$ | 100 kHz |
| DC link voltage | $V_{\mathrm{DC}}$ | 300 V |
| Capacitance of the resonant capacitor | $C_{\mathrm{r}}$ | 48 nF |
| Resonant (leakage) inductance | $L_{\mathrm{r}}$ | 6.54 $\mu$H |
| Magnetizing inductance | $L_{\mathrm{m}}$ | 328.5 $\mu$H |
| Turns ratio | $n$ | 4.4 |
| Filter capacity | $C_{\mathrm{f}}$ | 15 $\mu$F |
| Parasitic capacitances of transistors | $C_{\mathrm{Q1}}$, $C_{\mathrm{Q2}}$ | 100 pF |
| Load resistance | $R_{\mathrm{L}}$ | 9 $\Omega$ |
| Duty cycle | $D$ | 20 % |
| Output voltage | $V_{\mathrm{L}}$ | 15 V |
| Output current | $I_{\mathrm{L}}$ | 1.67 A |

the primary winding only.

This interval is terminated by turning off the upper transistor $\mathrm{Q_1}$. Its switching time, thus the duty cycle $D$, is typically determined by control circuits based on the desired value of the magnetizing current $I_{\mathrm{m\,POS}}$.

### B. Interval $t_1 - t_2$

After turning off the $\mathrm{Q_1}$ transistor, the current direction through the primary winding of transformer $i_1$ remains unchanged, causes:

1. Discharge the parasitic capacitance of transistor $C_{\mathrm{Q2}}$ and charge the parasitic capacitance of transistor $C_{\mathrm{Q1}}$.
2. The voltage $v_{\mathrm{HB}}$ to drop steeply, causing the diode $\mathrm{D_2}$ to open.
3. From this point on, the lower transistor $\mathrm{Q_2}$ can be turned on, as there is almost zero voltage on it (only the voltage drop across the diode $\mathrm{D_2}$ in the pass direction).

The turn-on losses of this lower transistor are therefore almost zero.

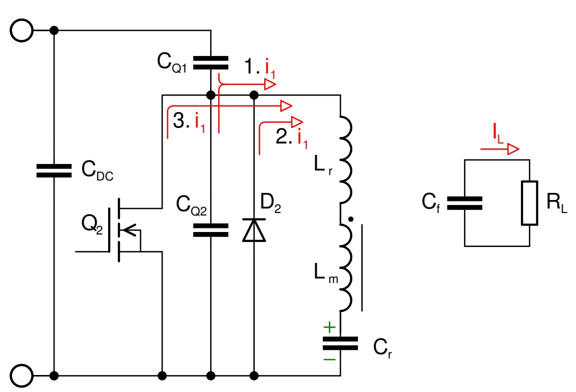



Fig. 4. Circuit behaviour during the interval $t_1$ - $t_2$.

### C. Interval $t_2 - t_3$

The current through the primary winding of the transformer $i_1$ still flows in the same direction, but the voltage on it ($v_1$) has changed its polarity, it is negative. The voltage of the resonant capacitor $v_{\mathrm{CR}}$ is applied across the switched transistor $\mathrm{Q_2}$ to

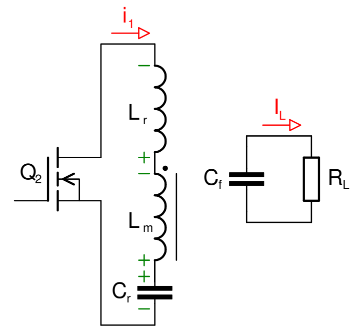



Fig. 5. Circuit behaviour during the interval $t_2$ - $t_3$.

the primary winding of the transformer. Thus, the current $i_1$ has stopped increasing and is decreasing.

This negative voltage $v_1$ of size $v_{\mathrm{CR}}$ is transformed to the secondary side. Its magnitude $v_2$ is given by the relation (1). But so far, the diode $\mathrm{D_3}$ has not been opened because the level of this secondary voltage does not exceed the magnitude of the voltage on the load $V_{\mathrm{L}}$ raised by the threshold voltage of the diode $v_{\mathrm{f\,D3}}$.

$$v_2 = (v_1 - v_{\mathrm{Lr}}) \cdot \frac{N_2}{N_1} \qquad (1)$$

However, due to the current $i_{\mathrm{Lm}}$, the voltage of the resonant capacitor $C_{\mathrm{r}}$ is still increasing. Therefore, the energy stored in the transformer is transferred to the $C_{\mathrm{r}}$ capacitor.

### D. Interval $t_3 - t_4$

After the voltage on the resonant capacitor $v_{\mathrm{CR}}$ increases enough to open the secondary diode $\mathrm{D_3}$, the energy starts to be transferred to the secondary side of the converter. During this interval, the current from the magnetizing inductance flows into both the secondary winding and part of it into the resonant capacitor $C_{\mathrm{r}}$, which is still being recharged.

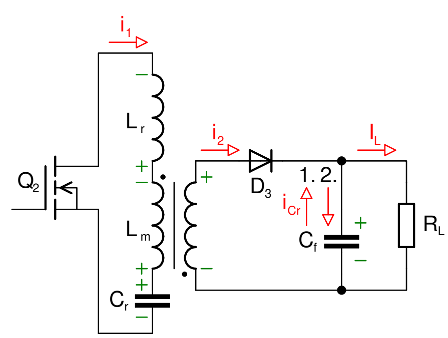



Fig. 6. Circuit behaviour during the interval $t_3$ - $t_4$.

The secondary winding is therefore subjected to an almost constant voltage from the filter capacitor $C_{\mathrm{f}}$, to which is added the voltage drop across the secondary diode in the pass direction $v_{\mathrm{f\,D3}}$, causing an almost linear demagnetization of the transformer.

As the current through the secondary winding increases, the current flowing from the capacitor $i_{\mathrm{Cf}}$ (1.) decreases until its direction changes, thus recharging the filter capacitor $C_{\mathrm{f}}$ (2.).

### E. Interval $t_4 - t_5$

In this interval, the energy is still transferred to the secondary side of the converter, but the direction of current

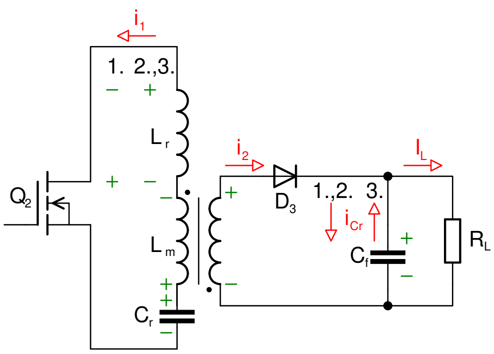



Fig. 7. Circuit behaviour during the interval $t_4$ - $t_5$.

through the primary winding $i_1$ and thus also through the resonant capacitor $C_{\mathrm{r}}$ has reversed. During this interval, the voltage across the output filter capacitor $C_{\mathrm{f}}$ goes past its maximum (2.), while the voltage across the resonant capacitor $v_{\mathrm{CR}}$ approaches its minimum. This causes the demising of current through the secondary winding (3.).

### F. Interval $t_5 - t_6$

Since the influence of the secondary winding is no longer applicable in this interval, the transformer can again be viewed as an inductance connected to a charged capacitor. For this short instant, the change in voltage $\Delta V_{\mathrm{CR}}$ across the resonant capacitor is again negligibly small compared to its DC component, causing an almost linear decrease in magnetizing current. Since the voltage on the primary winding of the transformer is negative at this point, $v_1$, energy is transferred from the $C_{\mathrm{r}}$ capacitor to the transformer.

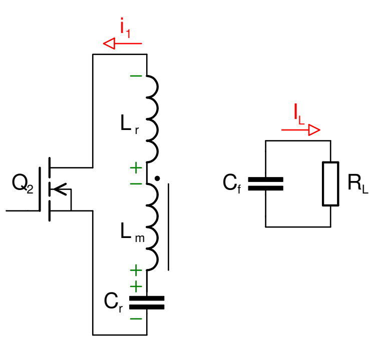



Fig. 8. Circuit behaviour during the interval $t_5$ - $t_6$.

### G. Interval $t_6 - t_7$

At the moment of switching off the lower transistor $\mathrm{Q}_2$ (1.), due to the magnetizing inductance current, the parasitic

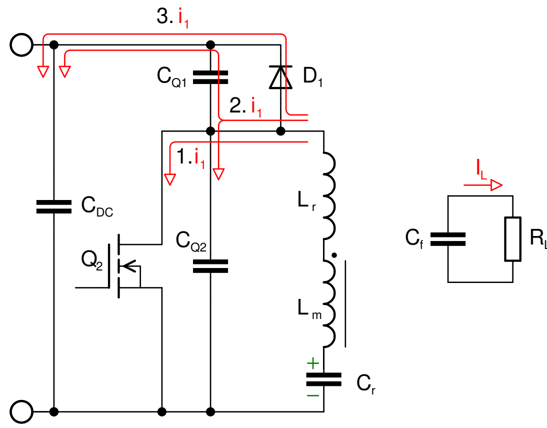



Fig. 9. Circuit behaviour during the interval $t_6$ - $t_7$.

capacitances of both transistors are overcharged (2.) and the diode $\mathrm{D}_1$ opens (3.).

After this interval, it is possible to turn on transistor $\mathrm{Q}_1$ again, because there is almost zero voltage on it (only a voltage drop on diode $\mathrm{D}_1$). At this point, there is still no fundamental change in the circuit behaviour, but the voltage drop across diode $\mathrm{D}_1$ is reduced.

### H. Interval $t_7 - t_8$

The voltage on the resonant capacitor $v_{\mathrm{CR}}$ is smaller than the input voltage $V_{\mathrm{DC}}$, a positive voltage is again applied to the primary winding. Since there is still a negative current through the primary winding $i_1$ during this interval, the energy stored in the transformer $\mathrm{T}$ and the resonant capacitor $C_{\mathrm{r}}$ is returned to the capacitor in the DC link $C_{\mathrm{DC}}$.

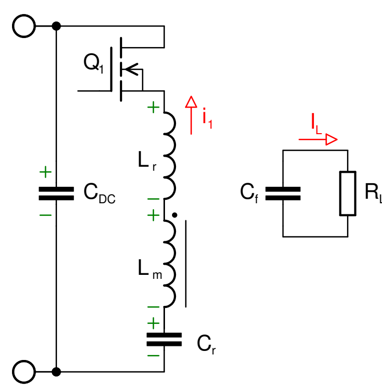



Fig. 10. Circuit behaviour during the interval $t_7$ - $t_8$.

At the end of this interval, the direction of the current through the primary winding of transformer $i_1$ will change, thus magnetizing the transformer again from the source with positive voltage. This ends the whole period and the same situation as in the interval $t_0 - t_1$ follows.

### I. Steady-state of the converter

The voltage transfer $k_{\mathrm{V}}$ for an ideal (lossless) converter is given by equation (2).

$$k_{\mathrm{V}} = \frac{V_{\mathrm{L}}}{V_{\mathrm{DC}}} = D \cdot \frac{N_2}{N_1} \cdot \frac{L_{\mathrm{m}}}{L_{\mathrm{m}} + L_{\mathrm{r}}} \tag{2}$$

Since the number of turns (N1, N2), the magnetizing inductance $L_{\mathrm{m}}$ and the resonant (leakage) inductance $L_{\mathrm{r}}$ are constants of the converter, the voltage transfer ku is changed only by duty cycle D during operation, not by the output current $I_{\mathrm{L}}$. Therefore, the leakage inductance $L_{\mathrm{r}}$ does not affect the softness of the converter. In a real converter, however, the output voltage $V_{\mathrm{L}}$ drops due to parasitic resistances of individual elements ($R_{\mathrm{DS\,on}}$ for $\mathrm{Q}_1$ and $\mathrm{Q}_2$, ESR for $C_{\mathrm{r}}$, etc.). [3].

The output current is given by equation (3), where $I_{\mathrm{m\,NEG}}$ is the initial magnetizing current level and $I_{\mathrm{m\,POS}}$ is the final magnetizing current level during magnetization [2].

$$I_{\mathrm{L}} = \frac{1}{2} \cdot \frac{N_1}{N_2} \cdot (I_{\mathrm{m\,POS}} + I_{\mathrm{m\,NEG}}) \tag{3}$$

The effect of the load current $I_{\mathrm{L}}$ on the primary current waveform $i_1$ can also be seen by comparing the time waveform $i_1(t)$ in Fig. 2, which corresponds to the nominal load current $I_{\mathrm{L}} = 1.67\,\mathrm{A}$, and in Fig. 11, where the converter is unloaded.

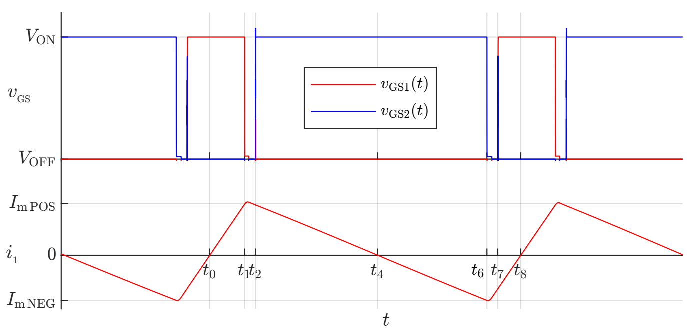



Fig. 11. Simulated waveforms without load.

The resonant frequency $f_{\mathrm{r1}}$ (4) applied at the time of flow of the secondary current $i_2$, that is, in the interval $t_3$ to $t_5$, is determined only by the leakage inductance $L_{\mathrm{r}}$ and the resonant capacitance $C_{\mathrm{r}}$ because the magnetizing inductance $L_{\mathrm{m}}$ is short-circuited from the AC point of view by the output capacitor $C_{\mathrm{f}}$ at this time.

$$f_{\mathrm{r1}} = \frac{1}{2 \cdot \pi \cdot \sqrt{L_{\mathrm{r}} \cdot C_{\mathrm{r}}}} \tag{4}$$

In other cases, the resonant frequency $f_{\mathrm{r2}}$ (5) is greatly reduced by the magnetizing inductance $L_{\mathrm{m}}$ [2].

$$f_{\mathrm{r2}} = \frac{1}{2 \cdot \pi \cdot \sqrt{(L_{\mathrm{r}} + L_{\mathrm{m}}) \cdot C_{\mathrm{r}}}} \tag{5}$$

## IV. Conclusion

The hybrid flyback resonant converter takes advantage of the advantages of a classic flyback converter in the form of simple secondary side wiring without the need for a separate output choke and a sufficient single-ended rectifier. Unlike a conventional flyback, it uses zero-voltage switching, which reduces interference to the environment, and it also reduces switching losses in the transistors. The converter is easily controlled using PWM modulation. While in a flyback converter the demagnetization of the leakage inductance has to be addressed, which in the simplest case is solved in a lossy manner, here the leakage inductance is used for resonance. Due to the fact that the time waveform of the current through the primary and secondary windings of the transformer does not contain steep edges, it also reduces the AC losses in the transformer windings compared to a conventional flyback converter. Unfortunately, the resonant current that is not involved in the power transfer in turn increases the effective value of the current through the primary winding, which in turn has a negative impact on the losses in the transformer.

## References


[1] R. Kanthimathi and J. Kamala, “Analysis of different flyback Converter topologies”, in *2015 International Conference on Industrial Instrumentation and Control (ICIC)*, 2015, pp. 1248-1252.

[2] Infineon Technologies AG, *Hybrid-flyback converter design with XDP digital power XDPS2201: Design Guide*, V1.0. 2021.

[3] A. Medina-Garcia, M. Schlenk, D. Morales, and N. Rodriguez, “Resonant Hybrid Flyback, a New Topology for High Density Power Adaptors”, *Electronics*, vol. 7, no. 12, pp. 1-12, 2018.

[4] Z. Ma, J. Yao, S. Wang, H. Sheng, S. Lakshmikanthan, and D. Osterhout, “Radiated EMI Reduction with Double Shielding Techniques in Active-clamp Flyback Converters”, in *2021 IEEE International Joint EMC/SI/PI and EMC Europe Symposium*, 2021, pp. 1064-1069.

[5] S. Ben-Yaakov and Y. Amran, “Multiple output flyback converters:: the role of output capacitors in shaping the currents of secondary windings”, in *21st Annual IEEE Conference on Power Electronics Specialists*, 1990, pp. 690-697.

[6] K. Soltanzadeh and M. R. Yousefi, “Analysis and design of two-switch flyback converter with double passive lossless snubber”, *IET Power Electronics*, vol. 11, no. 7, pp. 1187-1194, 2018.

[7] M. Patočka, *Magnetické jevy a obvody ve výkonové elektronice, měřicí technice a silnoproudé elektrotechnice*. V Brně: VUTIUM, 2011.

# Perceptual Omnidirectional Image Quality: Subjective Ratings by Diverse User Aspects

Marek Šimka
*FEEC, Dept. of Radio Electronics*
*Brno University of Technology*
Technická 3082/12, 616 00 Brno, Czech Republic
xsimka01@vut.cz

Ladislav Polak
*FEEC, Dept. of Radio Electronics*
*Brno University of Technology*
Technická 3082/12, 616 00 Brno, Czech Republic
polakl@vut.cz

***Abstract*—This paper deals with an unconventional experiment of subjective quality assessment of omnidirectional images. Within the subjective tests with 36 volunteers, the study was carried out in terms of different aspects of each user. Based on the assessments of 100 images by each respondent, it offers a new source of results for further investigation and development of virtual reality (VR) content or quality of experience (QoE). It focuses on the subjective ratings of various types of users such as people using dioptric glasses, with VR experience, or users with vision impairments. The output is thus the results of the subjective scores, including an analysis of their correlations with several objective quality metrics. The initial results of this work suggest that, for example, subjects with corrected vision using dioptric glasses exhibit similar subjective perceptions of the quality of omnidirectional images as average respondents without visual impairment.**

***Index Terms*—Omnidirectional images (360°), subjective assessment, image codecs, JPEG, JPEG XL, HEIC, AVIF**

## I. Introduction

The panoramic character of omnidirectional (360°) images offers users an immersive visual experience, presenting a comprehensive 360° view of the surrounding environment. However, as with any form of visual media, the perceived quality of omnidirectional images is subject to individual interpretation, influenced by various user aspects ranging from physical to personal preferences or character of content and many others [1].

This study focuses on understanding how different types of users subjectively evaluate the quality of 360° images. By conducting subjective tests and analyzing the results from the viewpoint of individual users, we aim to shed light on the nuanced relationship between user characteristics and their perception of image quality in the omnidirectional domain.

Throughout this investigation, we delve into the diverse aspects of users, considering factors such as age, eye impairments, and level of familiarity with omnidirectional virtual reality (VR) content. Our analysis goes beyond mere aggregate ratings, delving into how these user aspects influence subjective assessments. We seek to uncover whether certain user characteristics predispose individuals to perceive omnidirectional images differently and, if so, to what extent.

Research described in this paper was supported by the Internal Grant Agency of the Brno University of Technology under project no. FEKT-S-23-8191.

Understanding the interplay between user characteristics and perceptual evaluations of omnidirectional image quality holds significant implications for various fields utilizing such images. Whether optimizing immersive virtual environments, insights gained from this study can better inform the design and implementation of omnidirectional imaging technologies to suit users' diverse needs and preferences [2]. How users perceive image quality, what attributes they use to evaluate quality, or what influences their subjective quality of experience have been explored in a few papers.

The work of Zepernick et al. was concerned with determining the minimum or ideal number of respondents needed to subjectively evaluate the quality of VR content [3]. They tried to determine the number of subjects that would statistically produce reliable and credible data.

Another work [4] investigated whether the fact that the user is sitting or standing while observing VR content has an impact on the subjective perception of the quality of the omnidirectional video. They tried to find out, for example, from eye movements how the user behaves and if the perception of quality is different when sitting or standing.

The authors of [5] explored the quality of experience (QoE) of the omnidirectional video in the case of VR communication. They investigated the effects on QoE in terms of devices used, visual content characteristics, compression parameters, etc.

In [6], it was discovered a different influence on the perceived quality of 360° content. Specifically, they examined the field of view (FoV) and associated geometric distortions in terms of QoE. Analysis of the results indicated that FoV close to 110° had the highest preference among observers.

A different focus has the study [7] of consistency of 360° video quality assessment. Three repeated subjective tests were performed with standing and seated viewing in which participants rated the quality of the video. Among other findings, the summary results suggest that participants choose higher quality scale scores when standing compared to sitting scenarios. It also confirmed that the quality ratings of the visual stimulus were consistent across participants throughout the three sessions of each monitoring condition.

It is evident from these few mentioned papers that the field of image quality assessment (IQA) and, more specifically, the subjective perception of the quality of omnidirectional content is very broad and still offers an area for new research. This paper contributes by exploring a new perspective on the differences in user aspects in terms of subjective quality assessment. It focuses on the subjective ratings of diverse user types, such as people using dioptric glasses, with VR experience, or users with visual impairments. This proposes a new perspective on VR users and their subjective perception of image quality, which can help in the development of medical applications for diseased people or senior citizens, for instance.

## II. Subjective Experiment

### A. Image Dataset

In the case of subjective experiment, five reference images (*Biscayne*, *Flowers*, *Hokkaido 2*, *Telescope* and *Trains*, see in Figure 1) and their distorted variants from database[1] OMNIQAD were examined subjectively. A detailed description of features, resolution, and other specifications of each reference image is published in the paper [8]. A total of 100 images was evaluated, as there were 19 distorted variants for each reference image, and the original image was included too. Images were degraded with one still image codec (JPEG) and three emerging next-generation image compression algorithms, namely: JPEG XL [9], High Efficiency Image File Format (HEIC) [10], and AV1 Image File Format (AVIF) [11]. According to the target bit per pixel (bpp), the variants used in the test were distorted images in data streams from the lowest bits per pixel (bpp) values up to 1 bpp. Specifically, images in a range of 0.1, 0.2, 0.3, 0.5, 0.9 bpp (AVIF, HEIC, and JPEG XL), and 0.3, 0.5, 0.7, 0.9 bpp (JPEG).

### B. Setup and Conditions

For the experiment of subjective quality assessment of 360° images, a group of 36 observers (34 males and 2 females) consisting mostly of students were called to participate. The subject's age ranged from 19 to 29 years whereas the median age was 22 years. All participants had normal or corrected vision. This means that some people used correction with the dioptric glasses, i.e. they had on with the HMD during the test.

The Oculus Rift S[2] HMD has been used as a VR headset. It consists of an LCD panel with a resolution of $1280 \times 1440$ (per eye), a refresh rate of 80 Hz, and a Field of View (FoV) of 88°. It was connected to a laptop running Oculus software. The test stimulus were displayed in the app *Oculus 360 Photos*[3].

In contrast to other subjective experiments performed on traditional monitors or televisions, there is no need to consider environmental factors such as viewing distance, angle, and lighting conditions. The experiment was realized in a controlled environment of the laboratory at the Department of Radio Electronics (DREL), Brno University of Technology (BUT). Comfortable viewing of 360° images with the ability to freely rotate in different directions was provided to the participants by swivel chairs, which they utilized throughout the formal session.

[1] https://doi.org/10.5281/zenodo.7607071

[2] https://vr-compare.com/headset/oculusrifts

[3] https://www.oculus.com/experiences/rift/1004798886239391/

### C. Subjective procedure

According to the ITU [12] and [13] recommendations, the Absolute Category Rating with Hidden Reference (ACR-HR) methodology with a five-level rating scale (1-bad, 2-poor, 3-fair, 4-good, and 5-excellent) was adopted to assess the visual quality of 360° images. Hidden reference involves randomly including a reference among the tested images. This means that the observer rates the perceived quality of the stimulus independently, without knowing whether it is an original reference image or a degraded one. The structure of the experiment was divided into two sessions: pilot and formal.

In the pilot session, participants were introduced to the subjective evaluation and the experimental procedures, including familiarization with virtual reality (VR) technology. The underlying principle of the test was elucidated, ensuring participants understood the essence of their task. Basic demographic information, such as age, gender, and whether they wore glasses, was then collected to provide context for the subsequent analysis. Once the questionnaire and eye screenings were completed, participants were guided through a series of training exercises aimed at acquainting them with omnidirectional images and the process of providing perceptual quality ratings. Specifically, five sample images were presented to the participants as 360° content, consisting of compressed variants of the *Sights*[4] image from the database. These training images were carefully selected to represent typical content and distortions encountered in omnidirectional imagery. Importantly, these examples were distinct from the images used in the formal evaluation session to prevent affecting the subjective ratings. This structured approach ensured that participants were adequately prepared to contribute meaningfully to the study, enhancing the validity and reliability of the ensuing subjective assessments.

The formal session of the experiment adhered to a structured timeline, as depicted in Fig. 2. During this phase, each test stimulus, including the reference, was randomly presented one at a time. Participants were allotted a maximum viewing time of 15 seconds for each stimulus, although actual viewing durations varied depending on individual observer preferences. Following the viewing period, participants were prompted to provide their subjective ratings within a voting slot lasting up to 10 seconds. This slot featured a rating scale displayed on a gray background, allowing participants to report their perceived quality score efficiently. To record these ratings, participants verbally communicated their scores during the voting slot. To keep consistency across participants, stimulus viewing was controlled automatically, with switching occurring after 15 seconds. Alternatively, a slot may have been shorter if the participant reported the subjective score more quickly. This systematic approach aimed to minimize potential biases and variations, thus facilitating reliable and comparable subjective ratings. Overall, the duration of the formal part averaged around 19 min. Each participant had the opportunity to request a test interruption, break, and rest at any time.

[4] https://doi.org/10.5281/zenodo.7607071

a) *Biscayne* b) *Flowers* c) *Hokkaido 2* d) *Telescope* e) *Trains*

Fig. 1. Illustrative examples of the omnidirectional images from OMNIQAD database used in the subjective experiment.

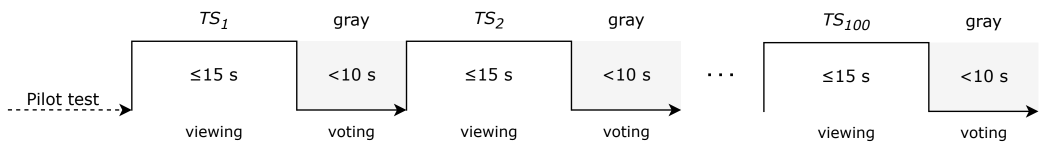



Fig. 2. The timeline of the subjective test quality (formal session) with the ACR-HR method ($TS_1$ represents a first tested stimulus, $TS_2$ a second tested stimulus). During the voting slot, the gray background with rating scale was displayed.

## III. Analysis of Subjective Ratings

### A. User Aspects

In the case of this work, the results of the subjective evaluation of the quality of the distorted omnidirectional images were analyzed in terms of differences between the subjects as users. More precisely, there was a heightened focus on what attributes or characteristics respondents communicated and also reported in the base questionnaire that was part of the experiment. Particularly interesting are the groups of users identified by terms characterizing the aspect: *Glasses*, *VR Experience*, *Eye Defect*.

The *Glasses* were users (18) who use dioptric glasses (with different diopter levels) in everyday life. The *VR Experience* group represents users (23) reporting in the questionnaire that they have used the VR glasses at least once or more in the previous time. The last type of unique user *Eye Defect* means the respondents (4) with another visual impairment (confirmed color blindness or astigmatism, etc.). In addition, part of the work mentioned the *Average* group for the results of all 36 subjects.

### B. Image Quality Assessment

The subjective scores from each participant are collected and used to obtain the mean opinion score (MOS) value as well as additional statistical data. The mean score for one stimulus (tested image) is calculated as [14]:

$$\bar{u}_k = \frac{1}{N}\sum_{i=1}^{N} u_{ik}, \tag{1}$$

where $u_k$ is subjective score of observer *i* for stimulus *k* and *N* is the number of observers.

The computed MOS scores are the main output of the subjective quality assessment. However, in this work, besides the individual MOS scores in different scenarios, the analysis was performed concerning objective quality metrics. We used various objective quality metrics to compare them with the subjective ratings. These were five widely used reference objective metrics: Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index Measure (SSIM) [15], Multi-Scale Structural Similarity (MS-SSIM) [16], Feature SIMilarity chrominance (FSIMc) [17], and Visual Information Fidelity in the pixel domain (VIFp) [18]. The PSNR, SSIM scores were obtained by the FFmpeg framework using the libraries *lavfi* and *libvmaf*. The MS-SSIM and VIFp metrics were computed in MATLAB using an available script[5]. The FSIMc metric was used according to [17].

## IV. Results

The first outputs of the data analysis are provided by Fig. 3, which contains the MOS values for the individual compressed reference images and separately also for their undistorted originals according to the already mentioned observed user types. In this case, the highest values can be seen for the Glasses group and, as expected, the lowest for the *Eye Defect* group. On the other hand, for the undistorted references, i.e. the originals, almost identical MOS values are seen for the *Average*, *Glasses*, and *VR Experience*.

A further point of view on the subjective test results was in terms of the individual distortions caused by the compression algorithms (see Fig. 4). Here, a total of four graphs are available for each compression algorithm separately. It can be noticed, for example, that for traditional JPEG (see Fig. 4 d)) the results are significantly close for all types of users and overlap with each other. Probably for this image compression, the sensitivity of each user to a given distortion is likely to be similar. We could also mention the AVIF and HEIC (see Fig. 4 a) and Fig. 4 b)) distortion cases, since here the characteristics of the MOS scores change more dynamically. Such as increased fluctuations in *Eye Defect* or changing spacing between the single user types depending on the level of distortion i.e. bpp.

Furthermore, an advanced approach to analyze the quality of omnidirecitonal images was used in this work considering the association with objective metrics. Commonly used statistical

[5]https://github.com/sattarab/image-quality-tools

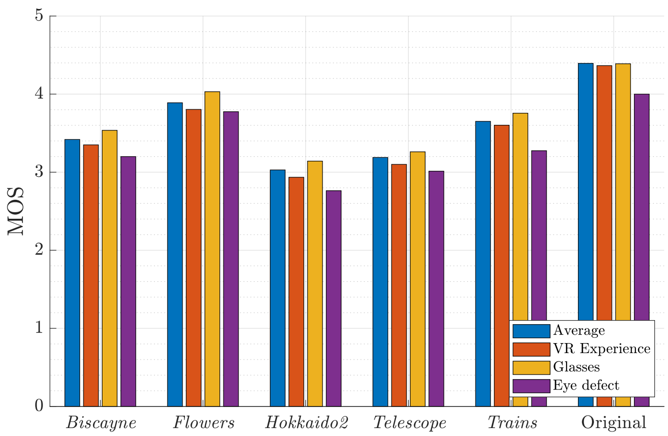



Fig. 3. MOS scores by diverse user types for specific reference images (compressed variants) and their undistorted originals.

methods in the IQA domain are Pearson linear correlation coefficient (PLCC), Spearman rank correlation coefficient (SRCC), and Kendall rank correlation coefficient (KRCC) [1], [19]. In this work, SRCC and KRCC were used to evaluate the level of correlation between subjective image quality scores and objective quality scores. The SRCC measures the monotonicity of the IQA metric prediction, i.e., the level to which the quality rating by metric corresponds to the relative sizes of the subjective scores. KRCC is also a non-parametric correlation coefficient that evaluates the number of matched and unmatched pairs in a data set. It is independent of any non-linear mapping between objective and subjective scores, as is the SRCC [20].

The detailed results for specific images and mainly for specific user types can be found in Table I. This contains the results of the correlation coefficients (SRCC and KRCC) for the previously mentioned five objective metrics and for the four user groups according to their diverse aspects. It is possible to compare the results in detail, perhaps in the case of specific reference images, to investigate how an image character has affected the perception of the users.

From the results of the average values for each type of image (last double column of the Table I), it can be seen that the highest correlation values were achieved by the *VR Experience* group for *Hokkaido 2* and *Trains* and by the *Glasses* users for *Flowers* and *Telescope*. Only in the case of the *Biscayne* image were the correlation coefficients highest for *Average*, the mean of all users, so here all types perceived quality very similarly regardless of aspect. Users using glasses showed a high degree of correlation with objective metrics similar to people who had VR experience, for example. So these are different results than could be seen in the actual MOS ratings on the previous graphs (Fig. 3 and Fig. 4). However, even here the expectation that the *Eye defect* perceived the quality of omnidirectional images to be at the lowest adequate and different from other users was confirmed.

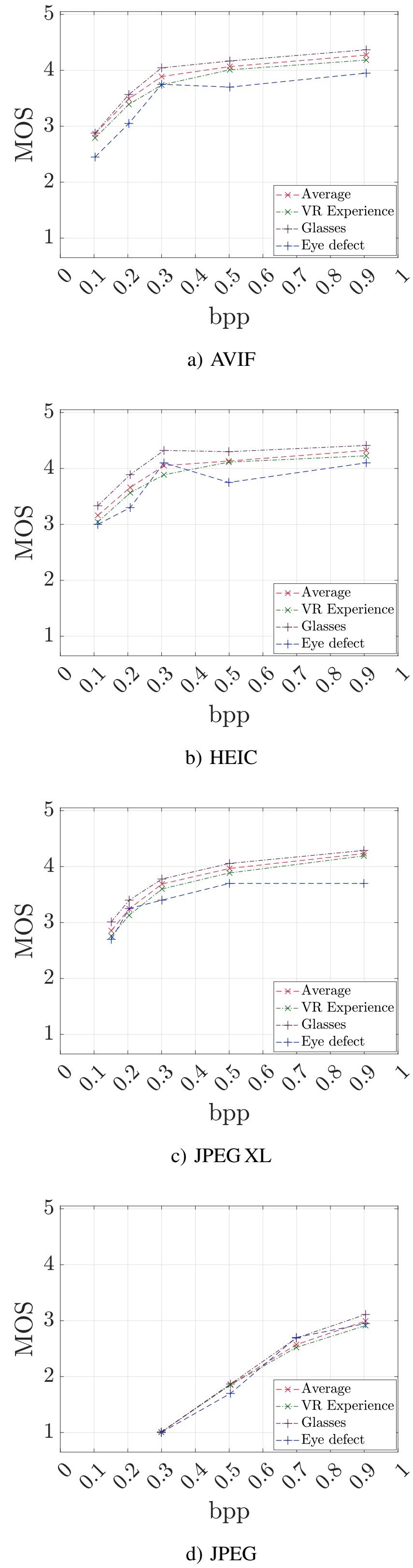



Fig. 4. Overall MOS ratings of diverse user types for different bpp values.

TABLE I

RESULTS OF SUBJECTIVE RATINGS RELATIVE TO OBJECTIVE METRICS IN TERMS OF SRCC AND KRCC FOR SPECIFIC USER TYPES ACCORDING TO REFERENCE IMAGES.

| **User Type** | **Image** | **PSNR** | | **SSIM** | | **MS-SSIM** | | **FSIMc** | | **VIFp** | | **Mean** | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | SRCC | KRCC | SRCC | KRCC | SRCC | KRCC | SRCC | KRCC | SRCC | KRCC | SRCC | KRCC |
| VR Experience | *Biscayne* | 0.8817 | 0.7323 | 0.9248 | 0.7918 | 0.9389 | 0.8156 | 0.9441 | 0.8204 | 0.9433 | 0.8156 | 0.9265 | 0.7951 |
| | *Flowers* | 0.7560 | 0.5977 | 0.8659 | 0.7397 | 0.9055 | 0.7397 | 0.9167 | 0.8012 | 0.9152 | 0.7634 | 0.8719 | 0.7283 |
| | *Hokkaido 2* | 0.8722 | 0.7198 | 0.9354 | 0.8260 | 0.9003 | 0.7906 | 0.8963 | 0.7537 | 0.9003 | 0.7788 | **0.9009** | **0.7738** |
| | *Telescope* | 0.9539 | 0.8563 | 0.9337 | 0.7977 | 0.8223 | 0.6452 | 0.9052 | 0.7706 | 0.9127 | 0.7507 | 0.9056 | 0.7641 |
| | *Trains* | 0.8765 | 0.7160 | 0.9187 | 0.7752 | 0.9345 | 0.7870 | 0.9242 | 0.8072 | 0.9196 | 0.7870 | **0.9147** | **0.7745** |
| Eye Defect | *Biscayne* | 0.8342 | 0.6327 | 0.8660 | 0.6936 | 0.8244 | 0.6814 | 0.8747 | 0.7221 | 0.8616 | 0.7057 | 0.8522 | 0.6871 |
| | *Flowers* | 0.4737 | 0.3621 | 0.5336 | 0.4246 | 0.6149 | 0.4745 | 0.5326 | 0.4322 | 0.6292 | 0.4995 | 0.5568 | 0.4386 |
| | *Hokkaido 2* | 0.7213 | 0.5779 | 0.7643 | 0.6160 | 0.7275 | 0.5779 | 0.7695 | 0.6324 | 0.7392 | 0.5652 | 0.7444 | 0.5939 |
| | *Telescope* | 0.9326 | 0.8206 | 0.9194 | 0.7847 | 0.8270 | 0.6529 | 0.8943 | 0.7509 | 0.9001 | 0.7607 | 0.8947 | 0.7539 |
| | *Trains* | 0.8055 | 0.6570 | 0.8444 | 0.7057 | 0.8586 | 0.7179 | 0.8714 | 0.7618 | 0.8409 | 0.6936 | 0.8442 | 0.7072 |
| Glasses | *Biscayne* | 0.8558 | 0.6970 | 0.8796 | 0.7330 | 0.8249 | 0.6729 | 0.8936 | 0.7615 | 0.8487 | 0.6970 | 0.8605 | 0.7123 |
| | *Flowers* | 0.7652 | 0.5935 | 0.8813 | 0.7360 | 0.9156 | 0.8072 | 0.9153 | 0.7855 | 0.9191 | 0.8191 | **0.8793** | **0.7483** |
| | *Hokkaido 2* | 0.8383 | 0.6927 | 0.8973 | 0.7524 | 0.8462 | 0.6927 | 0.8262 | 0.6727 | 0.8471 | 0.7046 | 0.8510 | 0.7030 |
| | *Telescope* | 0.9653 | 0.8850 | 0.9486 | 0.8496 | 0.8573 | 0.7080 | 0.9253 | 0.8107 | 0.9354 | 0.8142 | **0.9264** | **0.8135** |
| | *Trains* | 0.8677 | 0.7042 | 0.9134 | 0.7515 | 0.9327 | 0.7989 | 0.9250 | 0.8072 | 0.9090 | 0.7515 | 0.9096 | 0.7627 |
| Average | *Biscayne* | 0.9218 | 0.7870 | 0.9429 | 0.8225 | 0.9069 | 0.7634 | 0.9569 | 0.8512 | 0.9306 | 0.7870 | **0.9318** | **0.8022** |
| | *Flowers* | 0.7354 | 0.5631 | 0.8469 | 0.7038 | 0.9179 | 0.7859 | 0.9054 | 0.7643 | 0.9276 | 0.8094 | 0.8666 | 0.7253 |
| | *Hokkaido 2* | 0.8875 | 0.7397 | 0.9262 | 0.8107 | 0.8752 | 0.7515 | 0.8720 | 0.7322 | 0.8840 | 0.7634 | 0.8890 | 0.7595 |
| | *Telescope* | 0.9561 | 0.8713 | 0.9351 | 0.8129 | 0.8298 | 0.6608 | 0.9100 | 0.7859 | 0.9158 | 0.7661 | 0.9094 | 0.7794 |
| | *Trains* | 0.8669 | 0.7080 | 0.9144 | 0.7552 | 0.9311 | 0.8024 | 0.9233 | 0.8048 | 0.9100 | 0.7552 | 0.9091 | 0.7651 |

*The best results of mean values in each row i.e. according to each ref. image types are highlighted in bold.*

## V. CONCLUSION

This paper deals with the evaluation of the subjective quality tests of omnidirectional images compressed by emerging codecs. It analyzes subjective results from the perspective of different types of users such as dioptric glasses use, VR experience, or other eye defects. It offers the outcomes of the analysis and opens new questions for further research on the subjective quality assessment of omnidirectional images and the overall development of QoE in VR.

For example, one of the many results mentioned is the finding of people using dioptric glasses. Users using glasses showed a high degree of correlation with objective metrics similar to people who had VR experience, for example. Thus, this suggests that a visual impairment requiring dioptric glasses may not be a major obstacle to perceiving the high quality of 360° visual content, which is crucial for the user to achieve the desired immersive experience.

By clarifying the influence of user characteristics on subjective assessments, we aim to contribute valuable insights to the evolving field of omnidirectional content, paving the way for enhanced user experiences and advancements in VR. Besides the mentioned insights and the presented results, the paper does not aim to establish general conclusions but presents subjective assessments for specific groups of people.

Further work progress could be focused on investigating the subjective perception of the quality of omnidirectional content for other groups of people such as according to their age, gender, health impairments, etc. With a sizeable number of respondents, there is great potential to find new insights about users - for medical applications or VR multimedia content production.

## REFERENCES


[1] E. Upenik and T. Ebrahimi, "Saliency driven perceptual quality metric for omnidirectional visual content," in *2019 IEEE International Conference on Image Processing (ICIP)*, 2019, pp. 4335–4339.

[2] E. Upenik, M. Řeřábek, and T. Ebrahimi, "Testbed for subjective evaluation of omnidirectional visual content," in *2016 Picture Coding Symposium (PCS)*, 2016.

[3] H.-J. Zepernick, M. Elwardy, Y. Hu, and V. Sundstedt, "On the Number of Participants Needed for Subjective Quality Assessment of 360° Videos," in *2019 13th International Conference on Signal Processing and Communication Systems (ICSPCS)*, 2019, pp. 1–10.

[4] Y. Hu, M. Elwardy, and H.-J. Zepernick, "On the Effect of Standing and Seated Viewing of 360° Videos on Subjective Quality Assessment: A Pilot Study," *Computers*, vol. 10, no. 6, 2021.

[5] H. T. Tran, N. P. Ngoc, C. T. Pham, Y. J. Jung, and T. C. Thang, "A subjective study on QoE of 360 video for VR communication," in *2017 IEEE 19th international workshop on multimedia signal processing (MMSP)*. IEEE, 2017, pp. 1–6.

[6] F. Jabar, J. Ascenso, and M. P. Queluz, "Field-Of-View Effect on The Perceived Quality of Omnidirectional Images," in *2020 IEEE International Conference on Multimedia Expo Workshops (ICMEW)*, 2020, pp. 1–6.

[7] M. Elwardy, H.-J. Zepernick, T. M. C. Chu, and Y. Hu, "On the Consistency of 360 Video Quality Assessment in Repeated Subjective Tests: A Pilot Study," *EAI Endorsed Transactions on Industrial Networks and Intelligent Systems*, vol. 11, no. 1, p. 22, 2024.

[8] M. Simka and et al., "Omnidirectional Image Quality Assessment Database (OMNIQAD): Description and Examples," in *33rd International Conference Radioelektronika*, 2023.

[9] K. Rao and et al., "JPEG XL," *JPEG Series*, pp. 257–261, 2021.

[10] "High efficiency coding and media delivery in heterogeneous environments — part 12: Image file format," Coding of audio, picture, multimedia and hypermedia information, ISO/IEC 23008-12:2017, 2017.

[11] N. Barman and M. G. Martini, "An Evaluation of the Next-Generation Image Coding Standard AVIF," in *2020 Twelfth International Conference on Quality of Multimedia Experience (QoMEX)*, 2020.

[12] ITU-T, "P.919: Subjective test methodologies for 360º video on head-mounted displays," ITU, Recommendation, Series P.900-P.999 - Audiovisual quality in multimedia services, 10/2020, pp. 1-38.

[13] ITU-R, "BT.500-14: Methodologies for the subjective assessment of the quality of television images," ITU, Geneva, Recommendation, Series BT - Broadcasting service (television), 10/2019, pp. 1-102.

[14] ITU-T, "P.800.2: Mean opinion score interpretation and reporting," ITU, Geneva, Recommendation, Series P.800 - Methods for objective and subjective assessment of speech and video quality, 07/2016, pp. 1-18.
[15] A. Horé and D. Ziou, "Image quality metrics: PSNR vs. SSIM," in *20th International Conference on Pattern Recognition*, 2010, pp. 2366–2369.
[16] Z. Wang, E. Simoncelli, and A. Bovik, "Multiscale structural similarity for image quality assessment," in *The Thrity-Seventh Asilomar Conference on Signals, Systems  Computers*, vol. 2, 2003, pp. 1398–1402.
[17] L. Zhang, L. Zhang, X. Mou, and D. Zhang, "FSIM: A feature similarity index for image quality assessment," *IEEE Transactions on Image Processing*, vol. 20, no. 8, pp. 2378–2386, 2011.
[18] H. Sheikh and A. Bovik, "Image information and visual quality," *IEEE Transactions on Image Processing*, vol. 15, no. 2, pp. 430–444, 2006.
[19] Y. Fang, L. Huang, J. Yan, X. Liu, and Y. Liu, "Perceptual quality assessment of omnidirectional images," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 36, no. 1, 2022, pp. 580–588.
[20] M. Zhou and et al., "Attentional Feature Fusion for End-to-End Blind Image Quality Assessment," *IEEE Transactions on Broadcasting*, vol. 69, no. 1, pp. 144–152, 2023.

# General Solution of a Planar Linear Discrete System with a Weak Single Delay in the Case of Single Zero Eigenvalue of the Matrix of Nondelayed Terms

1st Marie Hartmanová
*Department of Mathematics*
*Brno University of Technology*
*Faculty of Electrical Engineering and Communication*
*and Central European Institute of Technology*
Brno, Czech Republic
marie.hartmanova@vut.cz

2nd Josef Diblík
*Department of Mathematics*
*Brno University of Technology*
*Faculty of Electrical Engineering and Communication*
*and Central European Institute of Technology*
Brno, Czech Republic
diblik@vut.cz

***Abstract*—Considered are linear planar discrete systems with a single delay. It is shown that their solutions can be expressed by explicit formulas given that the matrix of the linear non-delayed terms has two real eigenvalues with exactly one of them equaling zero and that the coefficients of the involved matrices depending on time satisfy certain conditions known for weakly delayed discrete systems with constant coefficients. Formulas derived can be useful in processing digital signals.**

***Index Terms*—discrete equation, planar system, single delay, initial problem, single zero eigenvalue, explicit solution**

## I. Introduction and the Problem Description

Deriving explicit formulas for solutions of various discrete equations and systems is important for their applications in digital signal processing as documented by [4]–[9]. The paper considers the following planar systems of discrete equations with a single delay

$$z(k+1) = \alpha(k)Fz(k) + \beta(k)Gz(k-r), \quad k = 0, 1, \dots \quad (1)$$

where $k$ is the independent variable, $r$ is a positive fixed integer, called a delay, $z = (z_1, z_2)^T : \{-r, -r+1, \dots\} \to \mathbb{R}^2$ is an unknown vector function, $T$ stands for the transpose of a vector, $\alpha, \beta : \{0, 1, \dots\} \to \mathbb{R}$ are scalar functions such that $\alpha(k)\beta(k) \neq 0$, $k = 0, 1, \dots$, $F = \{f_{mn}\}_{m,n=1}^2$ and $G = \{d_{mn}\}_{m,n=1}^2$ are constant $2 \times 2$ matrices. Below we assume that $G \neq \Theta$, where $\Theta$ is the zero $2 \times 2$ matrix. In the paper, explicit formulas for solutions to (1) will be found given that matrices $F$ and $G$ satisfy the assumptions for so called weakly delayed systems. Moreover, it is assumed that one of the eigenvalues of the matrix $F$ equals zero, while the second one is nonzero. This research continues the investigations initiated in [2], where a particular case of the system (1) is considered for the choice

$$\alpha(k) = \beta(k) = 1, \quad k = 0, 1, \dots . \quad (2)$$

In the present paper, a solution to system (1) is defined as a function $z : \{-r, -r+1, \dots\} \to \mathbb{R}^2$ satisfying (1) for all $k = 0, 1, \dots$. We work with solutions to system (1) that are determined by an initial problem defined as follows: for a given vector function $\varphi : \{-r, -r+1, \dots, 0\} \to \mathbb{R}^2$, referred to as an initial one, find a solution $z = z(k)$ to system (1) such that

$$z(k) = \varphi(k), \quad k = -r, -r+1, \dots, 0. \quad (3)$$

Since system (1) is linear, it is easy to show that the initial problem (3) defines a unique solution of this system. The purpose of the present paper is to derive an explicit formula solving the initial problem (3), that is, since the initial problem is arbitrary, to derive a general solution of the system (1).

## II. Weakly Delayed Systems

In [3] there are defined weakly delayed linear discrete systems. A short description of this notion is the following. Consider system (1) under assumption (2), that is, consider the system

$$z(k+1) = Fz(k) + Gz(k-r), \quad k = 0, 1, \dots \quad (4)$$

not assuming that $F$ has a zero eigenvalue. Then, system (4) is called weakly delayed if

$$\det(F - \lambda I) = \det(F - \lambda I + \lambda^{-r} G) \quad (5)$$

for each

$$\lambda \in \mathcal{C} \setminus \{0\}, \quad (6)$$

where $I$ is a $2 \times 2$ unit matrix. From equation (5) it is easy to derive a coefficient criterion saying (see Theorem 1.3 in [3])

Each of the authors has been supported by the project of specific university research FEKT-S-23-8179 (Faculty of Electrical Engineering and Communication, Brno University of Technology) for doctoral students and their supervisors and by the project of the Czech Science Foundation 23-06476S.

that the system (4) is weakly delayed if and only if the entries of matrices $F$ and $G$ satisfy

$$g_{11} + g_{22} = 0, \ \begin{vmatrix} g_{11} & g_{12} \\ g_{21} & g_{22} \end{vmatrix} = 0 \tag{7}$$

and

$$\begin{vmatrix} f_{11} & f_{12} \\ g_{21} & g_{22} \end{vmatrix} + \begin{vmatrix} g_{11} & g_{12} \\ f_{21} & f_{22} \end{vmatrix} = 0. \tag{8}$$

In short, conditions (7), (8) can be written as

$$\operatorname{tr} G = \det G = 0, \ \det(F + G) = \det F. \tag{9}$$

The meaning of condition (5) is the following. If it holds, then a quasi-characteristic equation of the system (4)

$$\det(F - \lambda I + \lambda^{-r} G) = 0 \tag{10}$$

has the same roots as the characteristic equation

$$\det(F - \lambda I) = 0 \tag{11}$$

of the system without delay

$$z(k+1) = Fz(k) \ \ k = 0, 1, \dots . \tag{12}$$

As equation (11) has only two roots, the general solution to system (12) is a linear combination of two linearly independent solutions of this system. If conditions (7), (8) hold, or (5) is valid, equation (10) has only two roots as well (instead of the expected $2(r+1)$ roots since (10) is a polynomial equation of order $2(r+1)$). Therefore, one can expect that the general solution of the system (4) will be a linear combination of two linearly independent solutions as well rather than a linear combination of $2(r+1)$ linearly independent solutions for all sufficiently large values $k$. It is shown in [1] that this is true within the following meaning: the general solution is a linear combination of two linearly independent solutions for all sufficiently large values $k$, but these two linearly independent solutions are not solutions of the linear system (12). Two linearly independent solutions of a linear system other than (12) must be used instead with such linear non-delayed systems being constructed in [1].

Consider a transformation

$$z(k) = Ty(k), \ \ k = -r, -r+1, \dots, \tag{13}$$

where $T$ is a regular $2 \times 2$ constant matrix. Then, the system (4) is transformed into a system

$$y(k+1) = F^* y(k) + G^* y(k-r), \ \ k = 0, 1, \dots, \tag{14}$$

where

$$F^* = T^{-1} F T, \ \ G^* = T^{-1} G T \tag{15}$$

and $T^{-1}$ is the inverse matrix of $T$. The initial problem (3) for the system (4) is transformed into the initial problem for the system (14), $y(k) = \omega(k) := T^{-1}\varphi(k)$, $k = -r, -r+1, \dots, 0$. An important and basic property of weakly delayed systems is their invariance with respect to any linear regular transformation. Namely, Lemma 1.2. in [3] states that, if the system (4) is a system with weak delay, then its arbitrary linear nonsingular transformation again leads to a system with a weak delay.

## III. "Weakly Delayed" System (4) in the Case of a Single Zero Eigenvalue of the Matrix $F$

The assumption (6) substantially restricts the class of the systems considered because the case of the matrix $F$ having one or two zero eigenvalues is excluded and the theory developed in [1], [3] does not work. The restriction $\lambda \neq 0$ has been removed in [2], where the system (4) is considered, in the case of the matrix $F$ having a single zero eigenvalue. Because the proofs and definitions in [3] cannot be used, in [2], it is assumed "ad hoc" that conditions (7) and (8) hold. The property, described by the above mentioned Lemma 1.2. in [3] is discussed and proved by a different method. If, for the entries of matrices $F$ and $G$, conditions (7) and (8) hold, then, for the entries of matrices $F^*$ and $G^*$ in the system (14) derived from the system (4) by the regular transformation (13), these conditions hold as well provided that the entries of matrices $F$ and $G$ are replaced by the entries of matrices $F^*$ and $G^*$ defined by formulas (15). Under the above condition, the general solution of the auxiliary system (14) as well as the general solution of the initial system (4) can be found in [2] where their dependence on the initial data is analyzed.

## IV. System (1) in the Case of a Single Zero Eigenvalue of the Matrix $F$

Let us consider the system (4). Assume that the matrices $F$ and $G$ satisfy coefficient conditions (7), (8). The following lemma says that if these matrices are replaced by matrices $\alpha(k)F$ and $\beta(k)G$, $k = 0, 1, \dots$, the relevant conditions remain true (note that $\det F = 0$ since one eigenvalue equals zero).

*Lemma 1:* Let conditions (7) and (8) hold. Then, for all $k = 0, 1, \dots,$

$$\operatorname{tr}(\beta(k)G) = \det(\beta(k)G) = 0, \tag{16}$$

and

$$\det((\alpha(k)F) + (\beta(k)G)) = 0. \tag{17}$$

*Proof:* The equations (16) are obvious since

$$\operatorname{tr}(\beta(k)G) = \beta(k)g_{11} + \beta(k)g_{22} = \beta(k)(g_{11} + g_{22}) = 0$$

and

$$\det(\beta(k)G) = \begin{vmatrix} \beta(k)g_{11} & \beta(k)g_{12} \\ \beta(k)g_{21} & \beta(k)g_{22} \end{vmatrix} = \beta^2(k) \begin{vmatrix} g_{11} & g_{12} \\ g_{21} & g_{22} \end{vmatrix} = 0.$$

Let us prove condition (17). We get

$$\begin{aligned} &\det(\alpha(k)F + \beta(k)G) \\ &\quad = \begin{vmatrix} \alpha(k)f_{11} + \beta(k)g_{11} & \alpha(k)f_{12} + \beta(k)g_{12} \\ \alpha(k)f_{21} + \beta(k)g_{21} & \alpha(k)f_{22} + \beta(k)g_{22} \end{vmatrix} \\ &\quad = \alpha^2(k) \begin{vmatrix} f_{11} & f_{12} \\ f_{21} & f_{22} \end{vmatrix} + \beta^2(k) \begin{vmatrix} g_{11} & g_{12} \\ g_{21} & g_{22} \end{vmatrix} \\ &\qquad + \alpha(k)\beta(k) \left( \begin{vmatrix} f_{11} & f_{12} \\ g_{21} & g_{22} \end{vmatrix} + \begin{vmatrix} g_{11} & g_{12} \\ f_{21} & f_{22} \end{vmatrix} \right) = 0. \end{aligned}$$

∎

Transforming system (1) by the regular transformation (13) leads to the system

$$y(k+1) = \alpha(k)F^*y(k) + \beta(k)G^*y(k-r), \;\; k = 0, 1, \ldots,$$

where matrices $F^*$ and $G^*$ are defined by formula (15). By the following lemma, relations (16) and (17) remain invariant under the regular transformation (13).

*Lemma 2:* Let conditions (7) and (8) hold. Then, for all $k = 0, 1, \ldots,$

$$\operatorname{tr}(\beta(k)G^*) = \det(\beta(k)G^*) = 0, \tag{18}$$

and

$$\det(\alpha(k)F^* + \beta(k)G^*) = 0. \tag{19}$$

*Proof:* The first equation in (18) can be proved with the aid of the first formula of the set of formulas labeled as (6) in [2]. It says that $g^*_{11} + g^*_{22} = 0$ if $g_{11} + g_{22} = 0$. In such a case

$$\operatorname{tr}(\beta(k)G^*) = \beta(k)(g^*_{11} + g^*_{22}) = 0.$$

The second equation in (18) is obvious since

$$\begin{aligned}\det(\beta(k)G^*) &= \beta^2(k)\det G^* = \beta^2(k)\det(T^{-1}GT)\\ &= \beta^2(k)\det T^{-1}\det G\det T = 0.\end{aligned} \tag{20}$$

To prove equation (19), we have

$$\begin{aligned}&\det(\alpha(k)F^* + \beta(k)G^*) =\\ &\quad\begin{vmatrix}\alpha(k)f^*_{11} + \beta(k)g^*_{11} & \alpha(k)f^*_{12} + \beta(k)g^*_{12}\\ \alpha(k)f^*_{21} + \beta(k)g^*_{21} & \alpha(k)f^*_{22} + \beta(k)g^*_{22}\end{vmatrix}\\ &\quad= \alpha^2(k)J_1 + \beta^2(k)J_2 + \alpha(k)\beta(k)J_3,\end{aligned} \tag{21}$$

where

$$J_1 = \det F^* = \begin{vmatrix}f^*_{11} & f^*_{12}\\ f^*_{21} & f^*_{22}\end{vmatrix}, \;\; J_2 = \det G^* = \begin{vmatrix}g^*_{11} & g^*_{12}\\ g^*_{21} & g^*_{22}\end{vmatrix},$$

and

$$J_3 = \begin{vmatrix}f^*_{11} & f^*_{12}\\ g^*_{21} & g^*_{22}\end{vmatrix} + \begin{vmatrix}g^*_{11} & g^*_{12}\\ f^*_{21} & f^*_{22}\end{vmatrix}.$$

Because the matrix $F$ has one zero eigenvalue, the same holds for the matrix $F^*$. Therefore, $J_1 = 0$. Moreover, we have $J_2 = 0$ as a consequence of (20). To prove the equation $J_3 = 0$ we use the last formula of the set of formulas labeled as (6) in [2] stating that $\det(F^* + G^*) = 0$. Since it is easy to verify (e.g., by setting $\alpha(k) = \beta(k) = 1$ in (21)) that

$$\det(F^* + G^*) = J_1 + J_2 + J_3$$

and, as we get above, $J_1 = J_2 = 0$, we derive

$$\det(F^* + G^*) = J_3 = 0.$$

Then

$$\det(\alpha(k)F^* + \beta(k)G^*) = 0,$$

equation (19) holds, and the lemma is proved. ∎

## V. Equivalent System to System (1) with the Jordan Matrix of Non-Delayed Terms

Assume that a tranformation (13) with an appropriate regular matrix $T := T_F$ transforms the system (1) into a system

$$y(k+1) = \alpha(k)Jy(k) + \beta(k)G^0y(k-r), \;\; k = 0, 1, \ldots, \tag{22}$$

where

$$J = \begin{pmatrix}\lambda & 0\\ 0 & 0\end{pmatrix}, \;\; G^0 = \{g^0_{ij}\}_{i,j=1,2} = T_F^{-1}GT_F, \tag{23}$$

i.e., $J$ is the Jordan form of the matrix $F$ and $G^0$ is the result of transformation of the matrix $G$. We show that the system can be solved easily if conditions (7) and (8) hold. Rewriting (22) in the scalar form as

$$\begin{aligned}y_1(k+1) &= \alpha(k)\lambda y_1(k) + \beta(k)(g^0_{11}y_1(k-r) + g^0_{12}y_2(k-r)),\\ y_2(k+1) &= \beta(k)(g^0_{21}y_1(k-r) + g^0_{22}y_2(k-r)),\end{aligned}$$

where $k = 0, 1, \ldots,$ we apply Lemma 2. From formulas (18) we deduce $\operatorname{tr}(\beta(k)G^0) = \det(\beta(k)G^0) = 0$, that is,

$$g^0_{11} + g^0_{22} = 0 \;\;\text{and}\;\; g^0_{11}g^0_{22} - g^0_{12}g^0_{21} = 0. \tag{24}$$

Formula (19) yields $\det(\alpha(k)J + \beta(k)G^0) = 0$, that is

$$\alpha(k)\beta(k)\lambda g^0_{22} - \beta^2(k)(g^0_{11}g^0_{22} - g^0_{12}g^0_{21}) = 0. \tag{25}$$

Formulas (24), (25) imply

$$g^0_{11} = g^0_{22} = g^0_{12}g^0_{21} = 0$$

and the following two possibilities must be considered. Either

$$g^0_{11} = g^0_{22} = g^0_{21} = 0 \;\;\text{and}\;\; g^0_{12} \neq 0 \tag{26}$$

or

$$g^0_{11} = g^0_{22} = g^0_{12} = 0 \;\;\text{and}\;\; g^0_{21} \neq 0. \tag{27}$$

Consequently, we will consider two systems. The first one

$$y_1(k+1) = \alpha(k)\lambda y_1(k) + \beta(k)g^0_{12}y_2(k-r), \tag{28}$$

$$y_2(k+1) = 0, \tag{29}$$

where $k = 0, 1, \ldots,$ corresponds to the case (26) while the second one

$$y_1(k+1) = \alpha(k)\lambda y_1(k), \tag{30}$$

$$y_2(k+1) = \beta(k)g^0_{21}y_1(k-r), \tag{31}$$

where $k = 0, 1, \ldots,$ corresponds to (27). The initial problem for both systems (28), (29) and (30), (31) is transformed from the initial problem (3) for the system (1):

$$y(k) = \omega^0(k) := T_F^{-1}\varphi(k), \;\; k = -r, -r+1, \ldots, 0. \tag{32}$$

## VI. Auxiliary Formula for Solutions of Discrete Equations

In the following two parts VII and VIII, we will find general solutions $y(k) = (y_1(k), y_2(k))^T$, $k = -r, -r+1, \dots$ of the system (28), (29) and the system (30), (31). We will apply a formula for solutions of scalar discrete nondelayed equations given in [4], formula (1.2.4). Consider an initial problem

$$w(k+1) = a(k)w(k) + g(k) = 0, \;\; w(k_0) = w_0. \tag{33}$$

where $k = k_0, k_0+1, \dots$ and $a, g\colon \{k_0, k_0+1, \dots\} \to \mathbb{R}$. Its solution is

$$w(k) = \left(\prod_{i=k_0}^{k-1} a(i)\right) w_0 + \sum_{\ell=k_0}^{k-1} \left(\prod_{i=\ell+1}^{k-1} a(i)\right) g(\ell). \tag{34}$$

## VII. General solution to system (28), (29)

To derive formulas for the general solution to system (28), (29) we proceed as follows. If $k = -r, -r+1, \dots, 0$, the solution is given by the initial data, that is, by formula (32),

$$y_1(k) = \omega_1^0(k), \tag{35}$$
$$y_2(k) = \omega_2^0(k). \tag{36}$$

Now, for $k = 0, 1, \dots r$, consider equation (28) where $y_2(k-r)$ is replaced by formula (36). Then,

$$y_1(k+1) = \alpha(k)\lambda y_1(k) + \beta(k) g_{12}^0 \omega_2^0(k-r).$$

By (34) with $k_0 = 0$, $y_1(0) = \omega_1^0(0)$, its solution is

$$y_1(k) = \left(\prod_{i=0}^{k-1} \alpha(i)\right) \lambda^k \omega_1^0(0) + \lambda^{k-1} g_{12}^0 \sum_{\ell=0}^{k-1} \lambda^{-\ell} \left(\prod_{i=\ell+1}^{k-1} \alpha(i)\right) \beta(\ell) \omega_2^0(\ell - r), \tag{37}$$

where $k = 1, 2, \dots, r+1$. From equation (29) we have

$$y_2(k) = 0, \;\; \forall k = 1, 2, \dots . \tag{38}$$

Finally, for $k = r+1, r+2, \dots$ consider equation (28) where $y_2(k-r)$ is replaced by formula (38),

$$y_1(k+1) = \alpha(k)\lambda y_1(k).$$

The initial value $y_1(r+1)$ is derived from formula (37), i.e.,

$$y_1(r+1) = \left(\prod_{i=0}^{r} \alpha(i)\right) \lambda^{r+1} \omega_1^0(0) + \lambda^r g_{12}^0 \sum_{\ell=0}^{r} \lambda^{-\ell} \left(\prod_{i=\ell+1}^{r} \alpha(i)\right) \beta(\ell) \omega_2^0(\ell - r).$$

By formula (34) with $k_0 = r+1$, its solution is

$$\begin{aligned} y_1(k) &= \left(\prod_{i=r+1}^{k-1} \alpha(i)\right) \lambda^{k-r-1} y_1(r+1) \\ &= \lambda^k \left(\prod_{i=0}^{k-1} \alpha(i)\right) \omega_1^0(0) \\ &\quad + \lambda^{k-1} g_{12}^0 \sum_{\ell=0}^{r} \lambda^{-\ell} \left(\prod_{i=\ell+1}^{k-1} \alpha(i)\right) \beta(\ell) \omega_2^0(\ell - r), \end{aligned} \tag{39}$$

where $k = r+2, r+3, \dots$.

## VIII. General Solution to System (30), (31)

Now we derive formulas for the general solution to system (30), (31). If $k = -r, -r+1, \dots, 0$, the solution is given by the initial data, that is, by (32) or (35) and (36).

Consider equation (30), where $k = 0, 1, \dots$. By formula (34), where $k_0 = 0$ and $y_1(0) = \omega_1^0(0)$, we have

$$y_1(k) = \left(\prod_{i=0}^{k-1} \alpha(i)\right) \lambda^k \omega_1^0(0), \tag{40}$$

where $k = 1, 2, \dots$. For $k = 0, 1, \dots, r$, equation (31) can be written as $y_2(k+1) = \beta(k) g_{21}^0 \omega_1^0(k-r)$ and, obviously,

$$y_2(k) = \beta(k-1) g_{21}^0 \omega_1^0(k-r-1) \tag{41}$$

where $k = 1, 2, \dots, r+1$.

Consider equation (31) for $k = r+1, r+2, \dots$. Using formula (40), this equation becomes

$$\begin{aligned} y_2(k+1) &= \beta(k) g_{21}^0 y_1(k-r) \\ &= \beta(k) g_{21}^0 \left(\prod_{i=0}^{k-r-1} \alpha(i)\right) \lambda^{k-r} \omega_1^0(0), \end{aligned}$$

having a solution

$$y_2(k) = \beta(k-1) g_{21}^0 \left(\prod_{i=0}^{k-r-2} \alpha(i)\right) \lambda^{k-r-1} \omega_1^0(0), \tag{42}$$

where $k = r+2, r+3, \dots$.

## IX. Formulation of Results

In parts VII and VIII, formulas are found describing the general solution of system (22). We complete them in the following theorem giving the general solution of system (1).

*Theorem 1:* Let the system (1) with matrices $F$ and $G$ satisfying conditions (9) be transformed by a transformation

$$z(k) = T_F y(k), \;\; k = -r, -r+1, \dots, \tag{43}$$

into the system (22) with matrices $J$ and $G^0$ given by formulas (23), where the matrix $J$ is the Jordan form of $F$. Then, the solution of initial problem (1), (3) is given by formula (43), where $y(k) = (y_1(k), y_2(k))^T$, $k = -r, -r+1, \dots$, and

*i*) If $k \in \{-r, -r+1, \dots, 0\}$, then

$$y(k) = \omega^0(k) := T_F^{-1} \varphi(k).$$

*ii*) If $k \in \{1, 2, \dots, r+1\}$ and $g_{12}^0 \neq 0$, then

$$y_1(k) = \left(\prod_{i=0}^{k-1} \alpha(i)\right) \lambda^k \omega_1^0(0) + \lambda^{k-1} g_{12}^0 \sum_{\ell=0}^{k-1} \lambda^{-\ell} \left(\prod_{i=\ell+1}^{k-1} \alpha(i)\right) \beta(\ell) \omega_2^0(\ell - r),$$

$$y_2(k) = 0.$$

$iii)$ If $k \in \{r+2, r+3, \dots\}$ and $g_{12}^0 \neq 0$, then

$$y_1(k) = \lambda^k \left( \prod_{i=0}^{k-1} \alpha(i) \right) \omega_1^0(0)$$
$$+ \lambda^{k-1} g_{12}^0 \sum_{\ell=0}^{r} \lambda^{-\ell} \left( \prod_{i=\ell+1}^{k-1} \alpha(i) \right) \beta(\ell) \omega_2^0(\ell - r),$$
$$y_2(k) = 0.$$

$iv)$ If $k \in \{1, 2, \dots, r+1\}$ and $g_{21}^0 \neq 0$, then

$$y_1(k) = \left( \prod_{i=0}^{k-1} \alpha(i) \right) \lambda^k \omega_1^0(0),$$
$$y_2(k) = \beta(k-1) g_{21}^0 \omega_1^0(k-r-1).$$

$v)$ If $k \in \{r+2, r+3, \dots\}$ and $g_{21}^0 \neq 0$, then

$$y_1(k) = \left( \prod_{i=0}^{k-1} \alpha(i) \right) \lambda^k \omega_1^0(0),$$
$$y_2(k) = \beta(k-1) g_{21}^0 \left( \prod_{i=0}^{k-r-2} \alpha(i) \right) \lambda^{k-r-1} \omega_1^0(0).$$

*Proof:* $i)$ is a consequence of (3), (13), and (32), $ii)$ is implied by (37) and (38), while $iii)$ by (39) and (38), $iv)$ is given by (40) and (41). The last point $v)$ is implied by (40) and (42). ∎

## X. Example

Let the system (1) be reduced to

$$z(k+1) = (k+1)Fz(k) + (k+1)Gz(k-1), \qquad (44)$$

where $r = 1$, $\alpha(k) = \beta(k) = k+1$, $k = 0, 1, \dots$ and matrices

$$F = \begin{pmatrix} 3 & 6 \\ 0 & 0 \end{pmatrix}, \quad G = \begin{pmatrix} -2 & 4 \\ -1 & 2 \end{pmatrix}$$

satisfy conditions (9). Transformation (43) with a regular matrix

$$T = T_F = \begin{pmatrix} 1 & -2 \\ 0 & 1 \end{pmatrix}, \; T_F^{-1} = \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix}$$

transforms the system (44) into a system

$$y(k+1) = (k+1)Jy(k) + (k+1)G^0 y(k-1),$$

where $k = 0, 1, \dots,$

$$J = \begin{pmatrix} 3 & 0 \\ 0 & 0 \end{pmatrix}, \; G^0 = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix},$$

$\lambda = 3$, $g_{21}^0 = 1 \neq 0$ and the initial data, for $k = -1, 0$, are

$$y(k) = \begin{pmatrix} \omega_1^0(k) \\ \omega_2^0(k) \end{pmatrix} = T_F^{-1} \varphi(k) = \begin{pmatrix} \varphi_1^0(k) + 2\varphi_2^0(k) \\ \varphi_2^0(k) \end{pmatrix}.$$

We apply Theorem 1. Part $i)$ repeats the initial data. By $iv)$ we get, for $k = 1, 2$,

$$y_1(k) = \left( \prod_{i=0}^{k-1} \alpha(i) \right) \lambda^k \omega_1^0(0) = k3^k(\varphi_1^0(0) + 2\varphi_2^0(0)),$$

$$y_2(k) = \beta(k-1) g_{21}^0 \omega_1^0(k-r-1)$$
$$= k(\varphi_1^0(k-2) + 2\varphi_2^0(k-2)),$$

and by $v)$, for $k = 3, 4, \dots,$

$$y_1(k) = \left( \prod_{i=0}^{k-1} \alpha(i) \right) \lambda^k \omega_1^0(0) = k!3^k(\varphi_1^0(0) + 2\varphi_2^0(0)),$$

$$y_2(k) = \beta(k-1) g_{21}^0 \left( \prod_{i=0}^{k-r-2} \alpha(i) \right) \lambda^{k-r-1} \omega_1^0(0)$$
$$= k(k-2)!3^{k-2}(\varphi_1^0(0) + 2\varphi_2^0(0)).$$

The general solution of the system (44) is given by the formula

$$z(k) = T_F y(k) = \begin{pmatrix} y_1(k) - 2y_2(k) \\ y_2(k) \end{pmatrix}, \quad k = -1, 0, \dots .$$

## XI. Concluding Remarks

The paper presents an analysis of planar discrete systems with a single delay and variable coefficients (satisfying certain restrictions) assuming that the matrix of linear nondelayed terms has a single zero eigenvalue. General solutions of such systems are constructed. The previously known results are not applicable because these investigations assume nonzero eigenvalues of the matrix of linear nondelayed terms. Assuming that the coefficients of matrices satisfy conditions similar to those of weakly delayed systems with nonzero eigenvalues, it is proved, by a different technique that such conditions are invariant with respect to any nonsingular transformation of the system. Since discrete systems play an important role in signal processing, every new class of discrete systems that is solvable by explicit formulas can be helpful because these formulas make it possible to perform not only a quantitative analysis but also a qualitative one, such as determining the behaviour of solutions for all very large values of time.

## References


[1] J. Diblík, H. Halfarová, and J. Šafařík, Two-parameters formulas for general solution to planar weakly delayed linear discrete systems with multiple delays, equivalent non-delayed systems, and conditional stability," Appl. Math. Comput. vol. 459, Art. ID 128270, pp. 1-14, 2023.

[2] J. Diblík, and M. Hartmanová, "Solutions of Delayed Linear Discrete Two-dimensional Systems with Constant Coefficients in the Case of Single Zero Eigenvalue of the Matrix of Nondelayed Terms," AIP Conference Proceedings, International Conference of Numerical Analysis and Applied Mathematics 2023, 11–17 September 2023, Heraklion, Greece, 4 pp., accepted paper.

[3] J. Diblík, D. Khusainov, and Z. Šmarda, "Construction of the general solution of planar linear discrete systems with constant coefficients and weak delay," Adv. Difference Equ. vol. 2009, Art. ID 784935, pp. 1-18, 2009.

[4] S. N. Elaydi, "An Introduction to Difference Equations," Third Edition, Springer, 2005.

[5] M. Sami Fadali, and A. Visioli, "Digital Control Engineering, Analysis and Design," Third Edition, Academic Press in an imprint of Elsevier, Elsevier, 2019.

[6] V. Lakshmikantham, and D. Trigiante, "Theory of Difference Equations," Second Edition, Marcel Dekker, Inc., 2002.

[7] A. V. Oppenheim, and Ronald W. Schafer, "Discrete-Time Signal Processing," Third Edition, Pearson, 2009.

[8] John G. Proakis, and Dimitris G. Manolakis, "Digital Signal Processing, Principles, Algorithms, and Aplications," Fourth Edition, Pearson, 2006.

[9] M. A. Radin, "Difference Equations for Scientists and Engineering: Interdisciplinary Difference Equations," World Scientific, 2019.

# Steady-state Thermal Analysis of Fault-tolerant PMSM During the Open-phase Mode of Operation

1st Vitaliy Sizonenko
*Department of Power Electrical and Electronic Engineering*
*Brno University of Technology*
Brno, Czech Republic
vitaliy.sizonenko@vut.cz

***Abstract*—In this paper, the initial outline of lumped-parameter thermal network and steady-state thermal finite element analysis for the fault-tolerant permanent magnet synchronous machine is demonstrated. The calculations show that the faulty operational mode with two open phases leads to substantial overheating of the machine when the need for the constant or overload torque is present. This challenge can be overcome by utilizing the machine's restricted operating time, with implementing transient thermal finite element analysis. Alternatively, the machine has the option to operate continuously at a reduced torque.**

***Index Terms*—Fault-tolerant electrical machine, permanent magnet synchronous machine (PMSM), finite element analysis (FEA), lumped-parameter thermal network (LPTN), thermal calculations.**

## Nomenclature

| | |
|---|---|
| $\alpha$ | Convection heat transfer coefficient $[\mathrm{W} \cdot \mathrm{m}^{-2} \cdot \mathrm{K}^{-1}]$ |
| $\alpha_{\mathrm{c}}$ | Conduction heat transfer coefficient $[\mathrm{W} \cdot \mathrm{m}^{-2} \cdot \mathrm{K}^{-1}]$ |
| $\dot{q}$ | Heat flux density $[\mathrm{W} \cdot \mathrm{m}^{-2}]$ |
| $\lambda$ | Thermal conductivity $[\mathrm{W} \cdot \mathrm{m}^{-1} \cdot \mathrm{K}^{-1}]$ |
| $b$ | Width [m] |
| $d$ | Diameter [m] |
| $h$ | Height [m] |
| $k_{\mathrm{Cu}}$ | Copper space factor [-] |
| $l_{\mathrm{s}}$ | Stack length [m] |
| $R$ | Thermal resistance $[\mathrm{K} \cdot \mathrm{W}^{-1}]$ |
| $S$ | Area $[\mathrm{m}^2]$ |
| $T$ | Temperature [K] |

## I. Introduction

The advancement of fault-tolerant electrical machines is greatly driven by the current reliability requirements in safety-critical electrical systems such as electric vehicles, aerospace technologies, and wind power generation systems. One of the distinctive features of such machines is the potential presence of asymmetric electromagnetic and thermal loads, defined by the requirement of operating during a fault. In particular, especially asymmetric thermal models have not yet been frequently encountered in the literature, providing thereby a space in this area for research and development. Accordingly, this paper attempts to approach this problem, proposing thermal calculations of a 5-phase outer-rotor fault-tolerant permanent magnet synchronous machine (PMSM) for the operation with open phases, and laying the foundation for the development of a general thermal analytical model of such machines.

## II. PMSM for Thermal Calculations

Machine rated data for the given problem are provided in Table I, example of its modes of operation is given in Table II. There, together with open phases fault, the overload requirement is presented, which is frequently applicable to fault-tolerant electrical machines. Thereby, the most dangerous for the machine overload mode with two open phases will be considered in the present paper. Such a situation can occur, e.g. in a propulsion motor of a plane during its take off.

Results of electromagnetic calculations, necessary as input for thermal calculations, are provided in Table III. The machine in faulty mode of operation requires separation of losses by different tooth-coils, to prepare precise thermal model.

TABLE I
Motor to Design

| **Parameter** | **Value** |
|---|---|
| Number of phases, m | 5 |
| Limit for rated line-to-line voltage $U_n$, V | 400 |
| Rated power $P_n$, kW | 30 |
| Rated speed $n$, rpm | 2000 |
| Rated torque $T$, Nm | 144 |

TABLE II
Modes of Operation and Faults

| **Mode of operation** | **Power, kW** | **Torque, Nm** | **Speed, rpm** | **Fault presence** |
|---|---|---|---|---|
| Rated mode | 30 | 144 | 2000 | healthy |
| Overload mode | 59 | 212 | 2653 | open phases C & E |

TABLE III
Electromagnetic Loads of Fault-Tolerant PMSM.

| **Mode of operation** | $\hat{U}_{n,max}$, [V] | $\hat{I}_{s,max}$, [A] | $\Delta P_{Cu}$, [W] | $\Delta P_{Fe}$, [W] | $\Delta P_{PM}$, [W] |
|---|---|---|---|---|---|
| Rated mode | 325 | 105 | 500 | 450 | 180 |
| Overload mode | 530 | 393 | 3300 | 700 | 780 |

## III. Thermal Modeling of PMSM

### A. Heat Convection

Heat is dissipated from the machine through convection. The formula for the calculation of $\Delta P_{\text{conv}}$ from a fin with an equivalent length, taking into account the convection from its tip, is given in [1], page 175. In the thermal finite element analysis (FEA) model, the convection through the fins at the stator yoke will be taken into account by empirical coefficient.

Additionally, the heat transfer through convection appears in the air-gap of the machine, in the radial direction. The present machine has a cooling circuit without axial flow in the air-gap. A comparative study of heat transfer with and without axial flow in the air-gap is given in [2].

### B. Heat Conduction

At first, a sketch of Lumped-Parameters Thermal Network (LPTN) will be created for PMSM working with open phases. An example of an asymmetrical LPTN for PMSM is provided in [3]. Then, this sketch must be verified and refined with the help of thermal Finite Element Analysis (FEA) model, and finally with the measurements on the machine sample.

Thermal resistance in the x-direction of a heat conductor shown in Fig. 1a can be calculated as follows:

$$R = \frac{b}{\lambda S} = \frac{1}{\alpha_{\text{c}} S} \tag{1}$$

According to [4], the one-dimensional thermal conductor can be described with the help of the LPTN shown in Fig. 1b, where the thermal resistances $R_0$ and $R_1$ are calculated as follows:

$$R_0 = \frac{b}{2\lambda S} = \frac{R}{2}, \tag{2}$$

$$R_1 = -\frac{b}{6\lambda S} = -\frac{R}{6}. \tag{3}$$

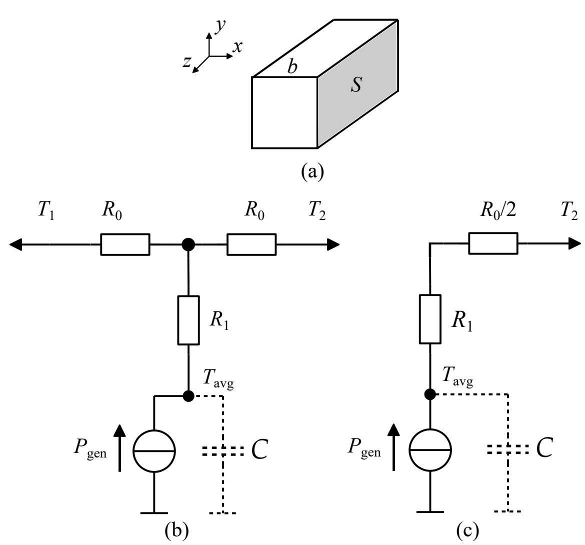



Fig. 1. One-dimensional LPTN of a thermal conductor.

The explanation and derivation of the given formulas is provided in [5], pages 57 - 60.

Applicable to electrical machines, one-dimensional LPTN corresponds to the assumption, when the heat generated by a copper conductor in the narrow stator slot is transferred only through the slots' sides, but not through the air-gap and the stator iron yoke. If the heat in the copper conductor is generated evenly, none of its sides is thermally insulated, and $T_1 = T_2$, the LPTN can be simplified to the one shown in Fig. 1c, where heat transfer is considered through the same conductor. There, iron losses and heat transfer in the axial direction are neglected.

Series connected resistances in Fig. 1c can be replaced with a single resistance $R_{01}$:

$$R_{01} = \frac{R_0}{2} + R_1 = \frac{R}{4} - \frac{R}{6} = \frac{R}{12}, \tag{4}$$

Accordingly, for the present PMSM with outer rotor, the first approximation of the heat flow generated by the copper conductor is drawn in Fig. 2a, its LPTN is shown in Fig. 2b. If the widths of the copper $b_{\text{Cu}}$ at the top and bottom of the slot are not equal, their average width $b_{\text{Cu,avg}}$ can be used for the calculation of copper thermal resistance $R_{01,\text{Cu}}$:

$$R_{01,\text{Cu}} = \frac{1}{12\lambda_{\text{Cu}}} \cdot \frac{b_{\text{Cu,avg}}}{h_{\text{Cu}} l_{\text{s}}}, \tag{5}$$

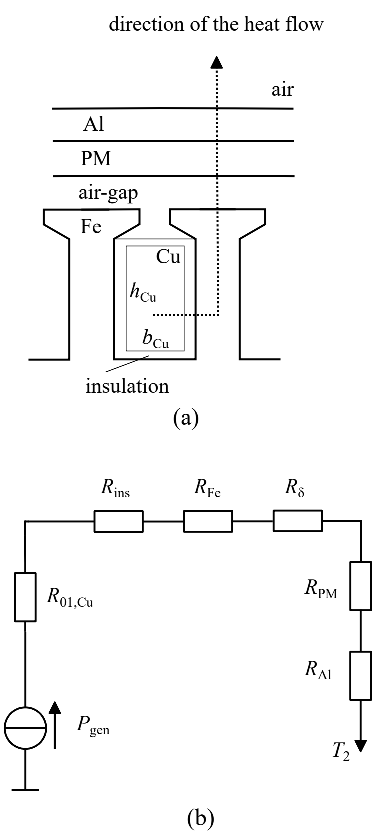



Fig. 2. Copper conductor generating heat. (a) General view. (b) One-dimensional LPTN.

### C. Equivalent Thermal Conductivity

In Fig. 2 an important assumption is that the slot area consists of only a single copper conductor and a layer of insulation around it. However, in electrical machines, the slot area usually contains the composition of conductors with very good thermal conductivity (TC), conductor insulation with very low TC, and impregnation with very low TC. Heat conduction in the slot area in such a case cannot be accurately described and must be approximated. Reference [6] provides analytical equations for calculating the equivalent thermal conductivity of a winding. Example calculations using these equations and their comparison with the FEA results are given in [7].

Thus, according to [6] and [7], thermal conductivity of the stator slot winding $\lambda_{\mathrm{w}}$ containing copper, insulation, and impregnation can be approximated in x-y plane as follows:

$$\lambda_{\mathrm{w}} = \lambda_{\mathrm{imp}} \cdot \frac{\lambda_{\mathrm{wire}}(1+k_{\mathrm{Cu}}) + \lambda_{\mathrm{imp}}(1-k_{\mathrm{Cu}})}{\lambda_{\mathrm{wire}}(1-k_{\mathrm{Cu}}) + \lambda_{\mathrm{imp}}(1+k_{\mathrm{Cu}})}, \tag{6}$$

where the thermal conductivity of the wire $\lambda_{\mathrm{wire}}$ is:

$$\lambda_{\mathrm{wire}} = \lambda_{\mathrm{ins}} \cdot \frac{\lambda_{\mathrm{Cu}}(1+(d_{\mathrm{Cu}}/d_{\mathrm{wire}})^2) + \lambda_{\mathrm{ins}}(1-(d_{\mathrm{Cu}}/d_{\mathrm{wire}})^2)}{\lambda_{\mathrm{Cu}}(1-(d_{\mathrm{Cu}}/d_{\mathrm{wire}})^2) + \lambda_{\mathrm{ins}}(1+(d_{\mathrm{Cu}}/d_{\mathrm{wire}})^2)}. \tag{7}$$

Then, the LPTN can be simplified to the one shown in Fig. 3, with thermal conductivity of the winding in the slot $R_{01,\mathrm{w}}$ calculated as:

$$R_{01,\mathrm{w}} = \frac{1}{12\lambda_{\mathrm{w}}} \cdot \frac{b_{\mathrm{w,avg}}}{h_{\mathrm{w}} l_{\mathrm{s}}}, \tag{8}$$

where $b_{\mathrm{w,avg}}$ is the average width of the winding in the slot, $h_{\mathrm{w}}$ is the height of the winding in the slot.

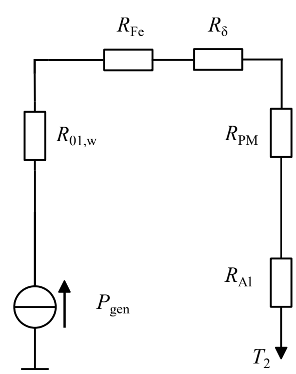



Fig. 3. Stator winding generating heat.

## IV. Thermal FEA

In FEA model there are contact faces between the iron core and the copper winding. The thermal flow through these contacts is defined by the thermal contact conductance of the insulation $\alpha_{\mathrm{c,ins}}$. From Equation (1), it can be approximated as follows:

$$\alpha_{\mathrm{c,ins}} = \frac{\lambda_{\mathrm{ins}}}{b_{\mathrm{ins}}} \approx \frac{0.2}{0.4 \cdot 10^{-3}} \approx 500\ \frac{\mathrm{W}}{\mathrm{m^2K}}, \tag{9}$$

where $\lambda_{\mathrm{ins}} = 0.2\ \frac{\mathrm{W}}{\mathrm{m \cdot K}}$ can be taken, e.g., from [8], page 536. However, in reality, the insulation contact is not ideal; there exist small air voids between the insulation and the surrounding metals, contributing to the total thermal resistance of the contact. Therefore, the total thermal contact conductance $\alpha_{\mathrm{c,tot}}$ between copper and iron is somehow lower:

$$R_{\mathrm{tot}} = \frac{1}{\alpha_{\mathrm{c,tot}} \cdot S} = \frac{b_{\mathrm{ins}}}{\lambda_{\mathrm{ins}} \cdot S} + \frac{1}{\alpha_{\mathrm{c,air}} \cdot S}, \tag{10}$$

$$\alpha_{\mathrm{c,tot}} = \frac{\alpha_{\mathrm{c,air}} \cdot \lambda_{\mathrm{ins}}}{\alpha_{\mathrm{c,air}} \cdot b_{\mathrm{ins}} + \lambda_{\mathrm{ins}}}. \tag{11}$$

For $\alpha_{\mathrm{c,air}} = 1200\frac{\mathrm{W}}{\mathrm{m^2K}}$ this equation gives:

$$\alpha_{\mathrm{c,tot}} = \frac{1200 \cdot 0.2}{1200 \cdot 0.4 \cdot 10^{-3} + 0.2} \approx 350\ \frac{\mathrm{W}}{\mathrm{m^2K}}. \tag{12}$$

Then, the properties of the materials used in the thermal FEA model are summarized in Table IV. There, the thermal conductivity $\lambda$ of the end winding is defined separately, due to its bending in this area. Consequently, individual cylindrical coordinate systems must be applied for each end-coil individually, with the Y-coordinate pointing along the end-coil curvature. Finally, the summary of the empirical coefficients for the thermal FEA model is given in Table V.

TABLE IV
FEA Materials Properties

| Parameter | Value |
|---|---|
| $\lambda$ of iron in XY directions [W/(m·K)] | 35 |
| $\lambda$ of iron in Z direction [W/(m · K)] | 2 |
| $\lambda$ of active winding in XY directions [W/(m · K)] | 1 |
| $\lambda$ of active winding in Z direction [W/(m · K)] | 150 |
| $\lambda$ of end winding in XZ directions [W/(m · K)] | 1 |
| $\lambda$ of end winding in Y direction [W/(m · K)] | 150 |

TABLE V
Empirical Data for Thermal FEA

| Parameter | Value |
|---|---|
| Total $\alpha_{\mathrm{c,tot}}$ [W/(m$^2$ · K)] between copper and iron | 350 |
| Convection from the stator yoke $\alpha_{\mathrm{yoke}}$ [W/(m$^2$ · K)] | 350 |
| Ambient temperature by the stator yoke [°C] | 40 |
| Convection from the stator tooth tips $\alpha_{\mathrm{tt}}$ [W/(m$^2$ · K)] | 200 |
| Ambient temperature by the air-gap [°C] | 100 |

It is important to note, that the empirical data provided in Table V for the thermal FEA model are considered as the first iteration of the thermal analysis. The convection heat transfer coefficients and the ambient temperatures are selected in such a way as to provide an acceptable steady-state temperature in the stator winding during the operation of the motor with the rated torque in healthy mode, see Fig. 4. Selected notations of dimensions for the iron part of the tooth-coil, necessary for further calculations, are given in Fig. 5.

The thermal FEA of the rotor is not considered in the current stage of the calculations, and is taken into account by the boundary conditions of the thermal FEA model of the stator. These boundary conditions will be corrected after measurements of the temperature in the machine sample with real cooling ambience.

Then, distribution of the steady-state temperature in the stator of the motor working with two open phases under the overload torque is provided in Table VI and in Fig. 6. There, acc. to Fig. 6, on the inner surface of the iron yoke the temperature gradient within a single tooth-coil span is up to $\Delta T \approx 281 - 176 \approx 105$ °C.

TABLE VI
DISTRIBUTION OF THE STEADY-STATE TEMPERATURE IN THE WINDING OF THE MOTOR, WORKING WITH TWO OPEN PHASES UNDER THE OVERLOAD TORQUE

| **Parameter** | **Phase** | | | | |
|---|---|---|---|---|---|
| | ***A*** | ***B*** | ***C*** | ***D*** | ***E*** |
| Minimum temperature [°C] | 340 | 337 | 191 | 358 | 191 |
| Maximum temperature [°C] | 647 | 634 | 243 | 553 | 242 |
| Average temperature [°C] | 541 | 532 | 210 | 482 | 210 |

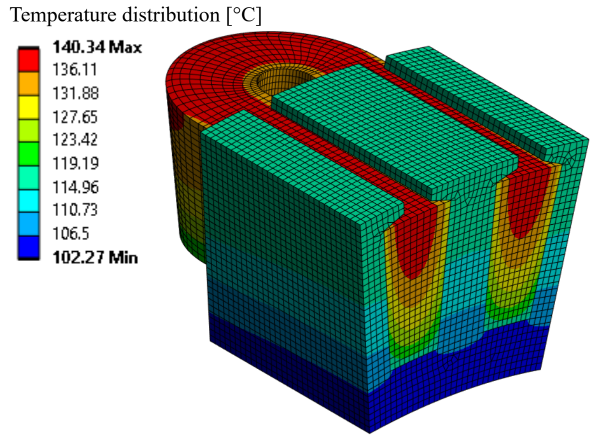



Fig. 4. Distribution of the steady-state temperature in the tooth-coil, healthy mode of operation of the motor with the rated torque.

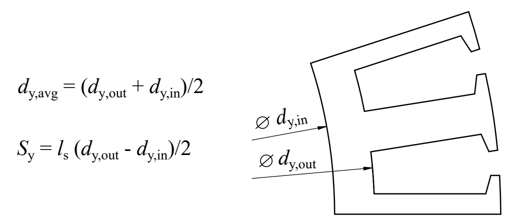



Fig. 5. Iron part of the tooth-coil.

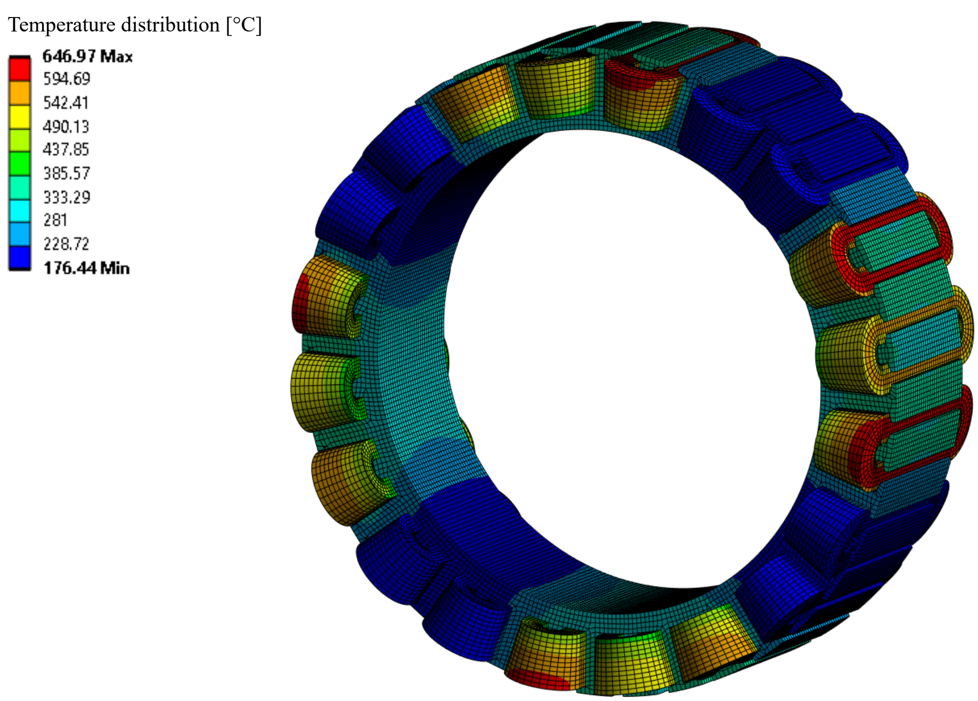



Fig. 6. Distribution of the steady-state temperature in the stator of the motor, working with two open phases under the overload torque.

Such result can be explained by the high heat flux density $\dot{q} = 115000$ W/m$^2$ in this region, calculated by FEA. This value can be used to check the temperature gradient with the formula:

$$\Delta\ T = S_{\mathrm{y}} \cdot \dot{q} \cdot R = S_{\mathrm{y}} \cdot \dot{q} \cdot \frac{\pi d_{\mathrm{y,avg}} \cdot 1/20}{\lambda S_{\mathrm{y}}} = \dot{q} \cdot \frac{\pi d_{\mathrm{y,avg}}}{\lambda \cdot 20}, \quad (13)$$

which provides the same temperature gradient $\Delta\ T \ \approx\ 105$ °C along the arc in the middle of the stator yoke with the diameter $d_{\mathrm{y,avg}} \ = \ 0.204$ m, and iron thermal conductivity in XY direction $\lambda = 35$ W/(m·K).

Finally, from the temperature distribution in the winding acc. to Table VI, it is clear that the given PMSM is not designed for long-term faulty operation with overload torque. Therefore, the working time of this machine during a fault with overload torque must be limited, which requires precise transient thermal FEA. Alternatively, the designed machine can work continuously under open phases with reduced torque.

## V. Conclusion

In the present paper, the first sketch of LPTN and steady-state thermal FEA for the fault-tolerant PMSM have been proposed. Calculations show that the faulty mode of operation with two open phases causes significant overheating of the machine if the requirement of constant or overload torque applies. This challenge can be overcome with the help of a limited working time of the machine, which requires precise transient thermal FEA. Alternatively, the machine can work continuously with open phases only under reduced load torque.

Then, further steps of the machine thermal FEA imply transient temperature analysis, verification and refinement of the thermal FEA model with the help of measurements on the machine sample, and finalization of the LPTN. These measures will allow us to develop a general analytical solution of thermal analysis for the machines of this kind.

## Acknowledgment

This research work has been carried out in the Centre for Research and Utilization of Renewable Energy (CVVOZE). Authors gratefully acknowledge financial support from the Ministry of Education, Youth and Sports under institutional support and BUT specific research programme (project No. FEKT-S-23-8430). Mr. Vitaliy Sizonenko acknowledges Mr. Ondřej Vítek and Mr. Marek Toman as his research supervisors for their valuable comments and help while writing this paper.

## References

[1] Y. A. Çengel and A. J. Ghajar, *Heat and mass transfer: fundamentals and applications*, fifth edition, New York, McGraw-Hill Education, 2015, ISBN 978-0-07-339818-1.

[2] M. Fénot, Y. Bertin, E. Dorignac and G. Lalizel, "A review of heat transfer between concentric rotating cylinders with or without axial flow", *International Journal of Thermal Sciences*, Volume 50, Issue 7, July 2011, pages 1138-1155, doi:10.1016/j.ijthermalsci.2011.02.013.

[3] F. Graffeo *et al.*, "Lumped-Parameters Thermal Network of PM Synchronous Machines for Automotive Brake-by-Wire Systems", in *Energies*, vol. 14, no. 18, 2021, doi.org/10.3390/en14185652.

[4] R. Wrobel and P. H. Mellor, "A General Cuboidal Element for Three-Dimensional Thermal Modelling," in *IEEE Transactions on Magnetics*, vol. 46, no. 8, pp. 3197-3200, Aug. 2010, doi: 10.1109/TMAG.2010.2043928.

[5] M. Toman, "Calculation of Optimum Magnetic Flux of Induction Motor to Minimize Losses with Consideration of Temperature Influences," Ph.D. dissertation, Brno University of Technology, Brno, Czech Republic, 2022.

[6] H. Liu, S. Ayat, R. Wrobel and C. Zhang, "Comparative study of thermal properties of electrical windings impregnated with alternative varnish materials", The Journal of Engineering, 2019, 17, ISSN 2051-3305, doi:10.1049/joe.2018.8198

[7] M. Toman, "Analysis of Equivalent Thermal Conductivity of Winding using FEM-based Model", in *Proceedings I of the 27th student EEICT 2021*, General papers 1, Brno, Brno University of Technology, 2021, pp. 544-548, ISBN: 978-80-214-5942-7.

[8] J. Pyrhönen, T. Jokinen, and V. Hrabovcova, *Design of Rotating Electrical Machines*, 2nd ed. New York, NY, USA: Wiley, 2014.

# Analysis of Heat Transfer in the Air Gap of Electrical Machines Using Empirical Equations and Computational Fluid Dynamics

Jakub Súkeník
*Department of Power Electrical and Electronic Engineering*
*Brno University of Technology, FEEC*
Brno, Czech Republic
xsuken00@vut.cz

Marek Toman
*Department of Power Electrical and Electronic Engineering*
*Brno University of Technology, FEEC*
Brno, Czech Republic
marek.toman@vut.cz

***Abstract*—This paper deals with the thermal analysis of the air gap using empirical equations and computational fluid dynamics (CFD). First, empirical equations are introduced to calculate the air gap heat transfer coefficients. The thermal air gap model is shown. Next, numerical modelling is discussed, namely the shear stress transport k-omega model. Afterwards, parametric air gap models are presented. The air gap heat transfer coefficients calculated from empirical equations are compared with the air gap heat transfer coefficients from CFD. Next, a comparison between the thermal resistances calculated using the empirical equations and between thermal resistances of CFD is shown. Finally, velocity air gap profiles with air gap temperature contours are presented.**

***Index Terms*—air gap, rotating electrical machine, computational fluid dynamics simulation, shear stress transport k-omega model, thermal model, Taylor-Couette flow**

## I. Introduction

In industrialised nations, motors and generators consume more than half of all electrical energy [1], [2]. Heat generation, resulting from losses, negatively impacts the life expectancy and performance of machines.

In general, electrical machine designers have focused on improving electromagnetic properties, while placing less emphasis on the thermal phenomena of electrical machines. However, the thermal aspects of electrical machines have become the centre of attention in recent years due to the increasing demand for electrical machines with higher power density. A limiting factor of electrical machines is usually determined by the insulation material, most of which degrade above a temperature of $130 - 180\,°\mathrm{C}$ [3], [4].

The next consideration is the temperature of the permanent magnets (PMs) in the rotor, if present. High temperatures can partially or completely demagnetise PMs in the rotor of the electrical machine [5]. In terms of heat transfer from the rotor to the stator of the electrical machine, the air gap is important. Complex physical phenomena take place in the air gap.

For thermal analysis of the air gap, empirical equations or computational fluid dynamics (CFD) can be used. Empirical relations have been obtained from numerous measurements. Their advantage lies in the speed of calculation and in the relative simplicity of their use. The disadvantage of these equations lies in their validity for a certain range and type of electrical machine. On the other hand, the use of CFD is highly effective in modelling complex geometries and providing precise temperature, pressure, and velocity distributions. However, this computational technique requires substantial computational resources, which could potentially affect processing speed [5], [6]. The thermal analysis of the air gap may or may not consider axial flow. In the case presented in this paper, the axial flow in the air gap of the electrical machine is not considered. The thermal analysis of the air gap is performed using empirical equations and CFD, and the results of both methods are compared.

## II. Empirical Equations for Calculating Air Gap Heat Transfer Coefficient

The heat transfer coefficient in the air gap $\alpha_\delta$ is an important parameter that significantly influences the rotor temperature. In the following, a procedure is proposed to calculate the heat transfer coefficient in the air gap $\alpha_\delta$, which is based on [7], further used and extended on [4].

The air gap is essentially an annulus formed by two concentric cylinders with rotation of the inner cylinder [4], [7]. The calculation is based on the theory of the flow between two concentric rotating cylinders, which is generally referred to as the Taylor-Couette flow or the Taylor vortex flow. It differs from the flow between two straight, parallel, mutually moving walls by presence of toroidal vortices, which significantly affect the heat transfer in the air gap. Taylor vortices

This research work has been carried out in the Centre for Research and Utilization of Renewable Energy (CVVOZE). Authors gratefully acknowledge financial support from the Ministry of Education, Youth and Sports under institutional support and BUT specific research programme (project No. FEKT-S-23-8430). Authors acknowledges Mr. Radoslav Cipín as my research supervisor for his valuable comments and help while writing this paper.

are quantified by the Taylor number Ta, which represents the ratio of viscous forces to centrifugal forces in the form

$$\mathrm{Ta} = \frac{\rho^2 \omega^2 r_\delta \delta^3}{\mu^2}, \tag{1}$$

where $\rho$ is density of the fluid, $\omega$ is the angular velocity of the rotor, $\mu$ is the dynamic viscosity of the fluid, $r_\delta$ is average radius of the air gap, and $\delta$ is the radial air gap length. The modified Taylor number $\mathrm{Ta_m}$ is determined from the Taylor number Ta in the form of

$$\mathrm{Ta_m} = \frac{\mathrm{Ta}}{F_\mathrm{g}}, \tag{2}$$

where $F_\mathrm{g}$ is the air gap curvature correction factor, which takes into account the relationship between the air gap thickness $\delta$ and the average air gap radius $r_\delta$, and is defined as

$$F_\mathrm{g} = \frac{\pi^4 \left[\frac{2r_\delta - 2.304\delta}{2r_\delta - \delta}\right]}{1697 \left[\frac{2r_\delta - 2.304\delta}{2r_\delta - \delta}^2\right] \left[1 - \frac{\delta}{2r_\delta}\right]}. \tag{3}$$

In most cases, the length of the air gap is small compared to the average radius of the air gap, which leads to $F_\mathrm{g} \to 1$ and $\mathrm{Ta_m} \approx \mathrm{Ta}$. In heat transfer theory, the Nusselt number Nu refers to the increase in the intensity of convection heat transfer compared to conduction heat transfer [8]. According to [4], [7], the Nusselt number Nu in the air gap is calculated as follows

$$\begin{aligned} \mathrm{Nu} &= 2 && \text{for} \quad \mathrm{Ta_m} \leq 1700, \\ \mathrm{Nu} &= 0.128\mathrm{Ta_m^{0.367}} && \text{for} \quad 1700 < \mathrm{Ta_m} < 10^4, \\ \mathrm{Nu} &= 0.409\mathrm{Ta_m^{0.241}} && \text{for} \quad 10^4 < \mathrm{Ta_m} < 10^7. \end{aligned} \tag{4}$$

The limiting case $\mathrm{Nu} = 2$ occurs at low rotor circumferential speeds, when the heat transfer through the air gap is provided only by the conduction and the air gap has the highest possible thermal resistance. Empirical equation (4) was obtained based on many measurements [7]. It is assumed that (4) is also valid for Ta greater than $10^7$. The heat transfer coefficient from the rotor surface to the air gap and from the stator surface to the air gap $\alpha_\delta$ is defined as

$$\alpha_\delta = \frac{\mathrm{Nu}\lambda}{\delta}, \tag{5}$$

where $\lambda$ is the thermal conductivity of the fluid in the air gap.

Fig. 1 shows a dependence of the Nusselt number on Taylor number for zero axial flow. Extrapolation indicated that the transition from laminar to turbulent flow occurs at $\mathrm{Ta_m}$ number of approximately 1700. The measurements in Fig. 1 are approximated by three regions. Empirical correlations in (4) were derived based on measurements; therefore, they may not match exactly with different types of air gap, for example, the air gap with rotor slots.

The thermal resistance from the rotor surface to the air gap and from the stator surface to the air gap $R_\delta$ according to [4], is calculated

$$R_\delta = \frac{1}{\alpha_\delta S}, \tag{6}$$

where $S$ is outer surface of the rotor or inner surface of the stator.

Fig. 2 shows the thermal network of air gap, where $\vartheta_\mathrm{rotor}$ is the outer surface temperature of the rotor, $\vartheta_\mathrm{stator}$ is the inner surface temperature of the stator, and $\vartheta_\delta$ is the average temperature of the air gap. This simple

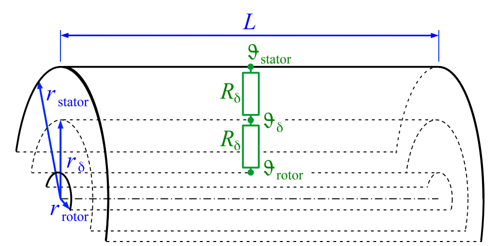



Fig. 2. Thermal network of air gap.

thermal model will be used in order to compare overall air gap thermal resistivity.

## III. Numerical Modelling

There are a number of numerical models for flow simulations, which are presented, for example, in [9]. In this paper, only one numerical flow model is discussed.

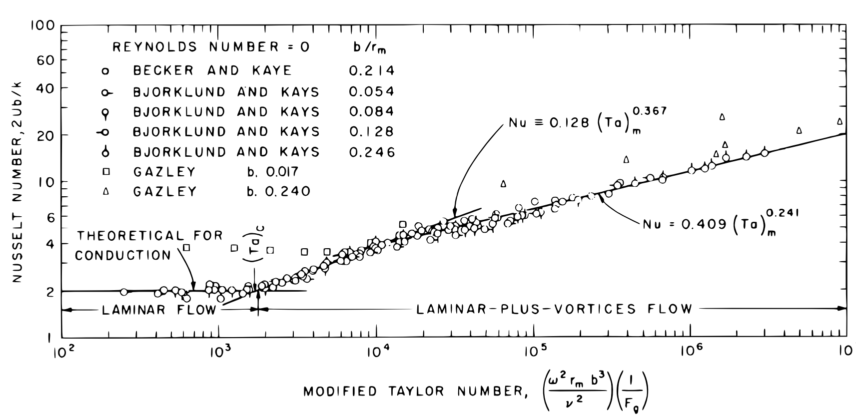



Fig. 1. Nusselt number vs Taylor number for zero axial flow, modified from [7].

### A. Shear Stress Transport $k-\omega$ Model

The Shear Stress Transport (SST) $k-\omega$ model is a widely used two-equation turbulence model in CFD. It combines the advantages of both the $k-\epsilon$ model [10], which performs well in the free stream region, and the standard $k-\omega$ model [9], which is more accurate near the wall boundaries.

The SST $k-\omega$ model, therefore, provides a comprehensive and accurate representation of turbulent flows in a wide range of flow conditions, making it a versatile tool for the analysis and prediction of complex fluid dynamics problems [9], [11], [12].

SST $k-\omega$ was introduced in 1992 by Menter [13]. The SST $k-\omega$ model is described by two main transport equations. The first transport equation is for the kinetic energy of the turbulence $k$, according to [12], in the form of

$$\frac{\partial}{\partial t}(\rho k)+\frac{\partial}{\partial x_i}(\rho k u_i)=\frac{\partial}{\partial x_j}\left(\Gamma_k\frac{\partial k}{\partial x_j}\right)+\tilde{G}_k-Y_k+S_k, \quad (7)$$

where $\tilde{G}_k$ represents the generation of turbulence kinetic energy due to mean velocity gradients, $\Gamma_k$ is the effective diffusivity of $k$, $Y_k$ represent the dissipation of $k$ due to turbulence, and $S_k$ are user-defined source terms. The second transport equation is for the specific dissipation rate $\omega$, and this equation is in the form of

$$\frac{\partial}{\partial t}(\rho\omega)+\frac{\partial}{\partial x_i}(\rho\omega u_i)=\frac{\partial}{\partial x_j}\left(\Gamma_\omega\frac{\partial \omega}{\partial x_j}\right)+G_\omega - Y_\omega + D_\omega + S_\omega, \quad (8)$$

where $G_\omega$ represents the generation of $\omega$, $\Gamma_\omega$ is the effective diffusivity of $\omega$, $Y_\omega$ represent the dissipation of $\omega$ due to turbulence, $D_\omega$ represents the cross-diffusion term, and $S_\omega$ are user-defined source terms. The procedure to calculate the individual terms in (7) and (8) is listed in [9], [11], [12]. To calculate the individual terms in (7) and (8), coefficients must be established. The coefficient values along with their labels are shown in Table I.

TABLE I
SST $k-\omega$ MODEL CONSTANTS

| Label | Value | Label | Value |
|---|---|---|---|
| $\alpha_\infty^*$ | 1.000 | $R_k$ | 6.000 |
| $\alpha_\infty$ | 0.520 | $R_\omega$ | 2.950 |
| $\alpha_0$ | $0.11\bar{1}$ | $\zeta^*$ | 1.500 |
| $\beta_\infty^*$ | 0.090 | $M_{t0}$ | 0.250 |
| $\beta_i$ | 0.072 | $\sigma_k$ | 2.000 |
| $R_\beta$ | 8.000 | $\sigma_\omega$ | 2.000 |

To illustrate the different flow areas and the advantages of the SST $k-\omega$ model described above, a simple model is proposed in Fig. 3. The inlet is located on the left side, while the outlet is situated on the right side. The fixed slip wall boundary condition is close to the bottom of the inlet. The slip wall boundary condition prevents creating the boundary layer to its surface; thus, viscous effects are neglected. On the other side, close to the bottom of the outlet, non-slip wall boundary condition is used. The non-slip wall is a wall where

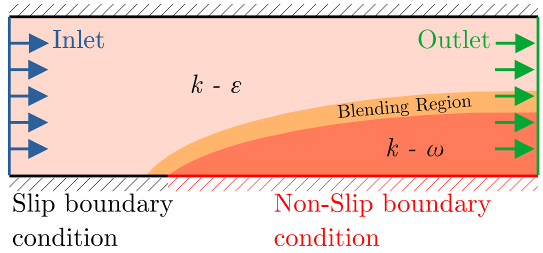



Fig. 3. Fluid flow model with different types of boundary conditions, modified from [14].

the fluid velocity relative to the wall velocity is zero [8]. In Fig. 3 there is also a mixing region. Transition occurs between the $k-\omega$ and $k-\epsilon$ models. The SST $k-\omega$ model can easily alternate between $k-\omega$ and $k-\epsilon$ using the blending function $F_1$, which is located in (7) and (8). If $F_1 = 1$, then the SST $k-\omega$ model uses the $k-\omega$ model, and if $F_1 = 0$, the $k-\epsilon$ model is used [9], [14]. The mixing function is shown in Fig. 4. The same model as in Fig. 3 is considered,

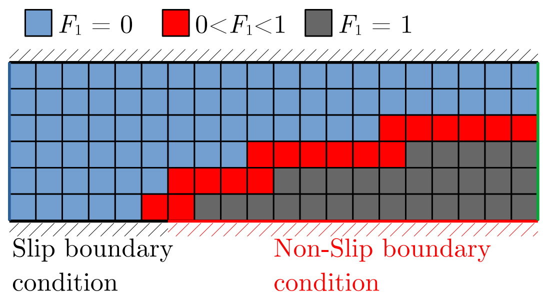



Fig. 4. Use of mixing function on fluid flow model with different types of boundary conditions, modified from [14].

but this model is divided into cells that form a mesh. Each cell in Fig. 4 contains its own value of the mixing function. The SST $k-\omega$ model uses another mixing function. This function is called the viscosity limiter $F_2$. The $F_2$ depends on the distance to the closest wall, such as $F_1$ [9], [14].

### B. Parametric air gap models

To simulate complex phenomena in the air gap, parametric air gap models are proposed. First, the geometric model must be prepared. The approach described above is largely based on [15]. The geometry of the parametric model has been created on the basis of the typical dimensions of the air gap [4]. The width of the air gap lies between 0.3 mm and 1.5 mm. Next, the mesh is proposed to divide the models into small, uniform cells. Fig. 9 (a) depicts the 3D air gap mesh, and Fig. 9 (b) shows a 2D air gap mesh. Like the air gap model geometries, the meshes in Fig. 9 are fully parameterised. In the second to last step, CFD simulations were created. First, the meshes were exported to Ansys Fluent. One of the most important parts of the initialisation process was the assignment of boundary conditions. A temperature of 85 °C was assigned to the stator wall, and a rotational condition was added to the rotor wall and a temperature of 75 °C was set to create a temperature gradient from the rotor to the stator of the electrical machine. For the 2D model, an axisymmetric

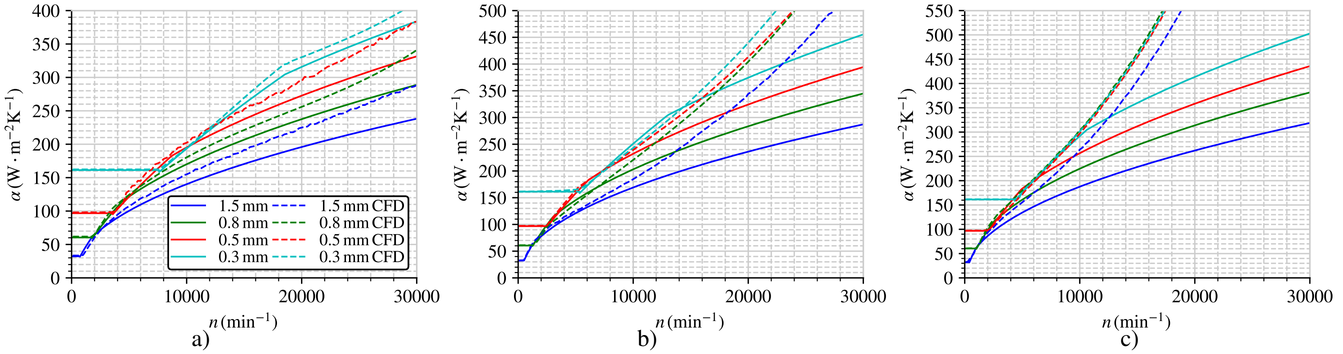


Fig. 5. Comparison of heat transfer coefficient in the air gap between empirical equations from [4], [7] and CFD models at rotational speed and different air gap widths, (a) profiles for 20 mm rotor radius, (b) profiles for 40 mm rotor radius, (c) profiles for 60 mm rotor radius.

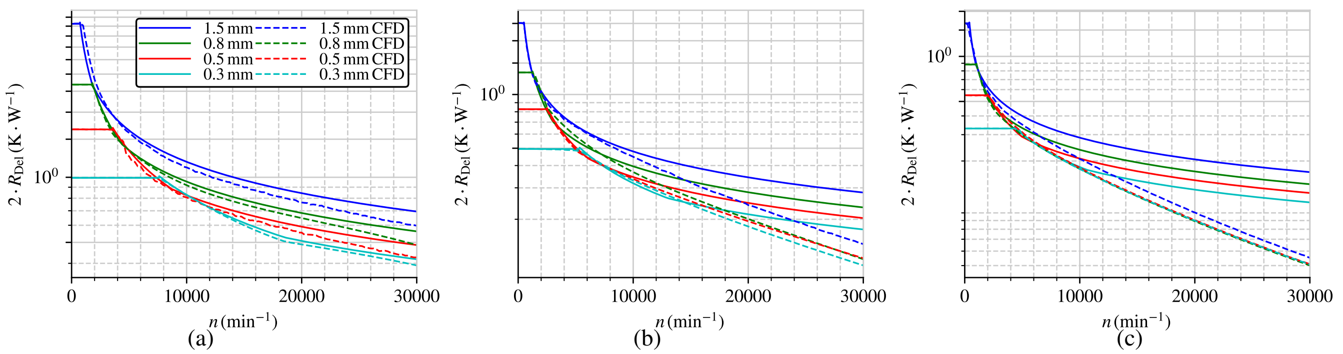


Fig. 6. Comparison of thermal resistances calculated using the empirical equations and CFD air gap fluid flow models, considering various air gap widths and rotation speeds, (a) profiles for the 20 mm rotor radius, (b) profiles for the 40 mm rotor radius, (c) profiles for the 60 mm rotor radius.

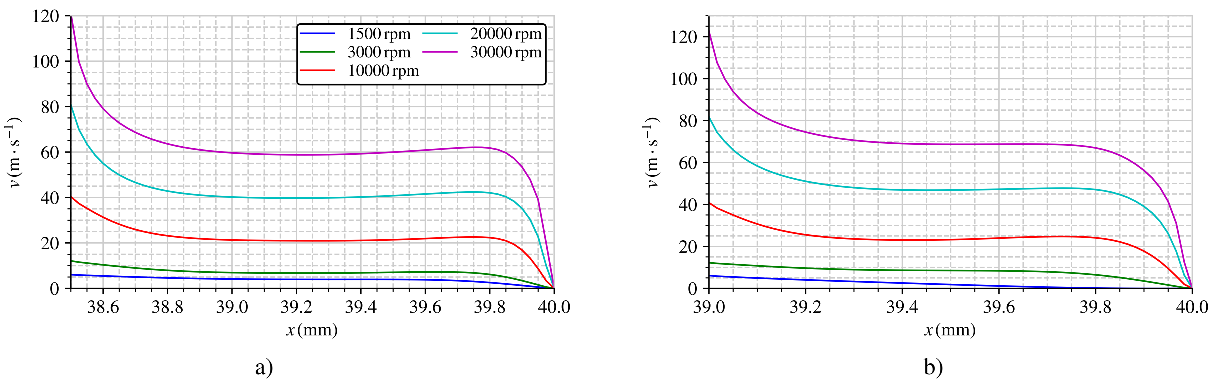


Fig. 7. Velocity profiles on line in the air gap for different rotation speeds, (a) profile for models with 1.5 mm air gap, (b) profile for models with 1.0 mm air gap.

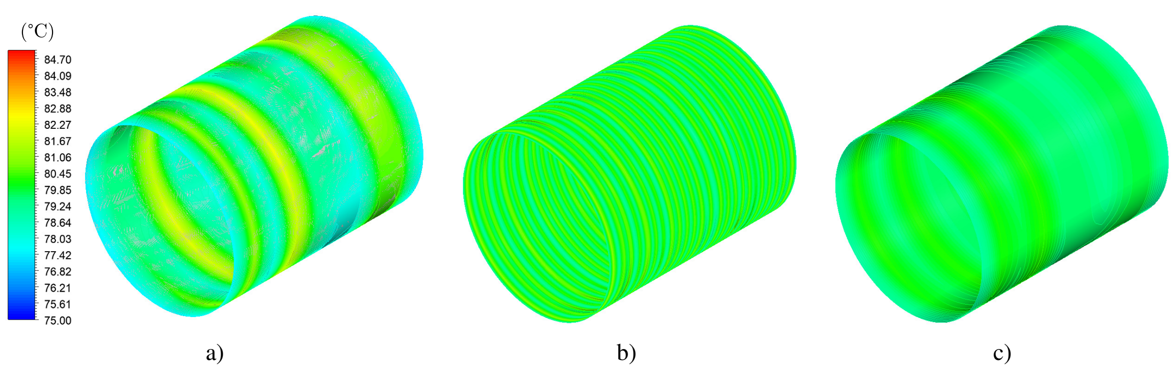
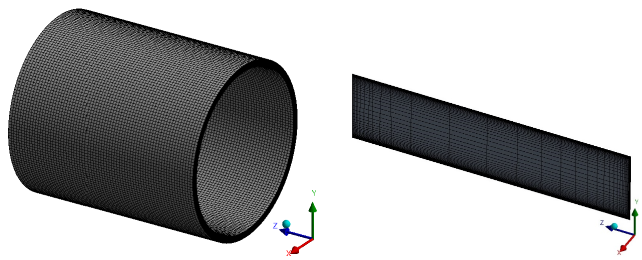


Fig. 8. Temperature contours on middle surface of 1.5 mm air gap, (a) temperature profile for 1500 rpm, (b) temperature profile for 10000 rpm, (c) temperature profile for 30000 rpm.



(a) (b)

Fig. 9. Meshes for CFD air gap models, (a) 3D mesh, (b) 2D mesh.

flow model with swirl was used. The axisymmetric flow model with swirl also works with the tangential direction of the fluid flow. In the last step, the Ansys Fluent results are exported to Ansys Results, and the results are evaluated.

## IV. Results

Fig. 5 shows the comparison of heat transfer coefficients in the air gap between empirical equations from the literature and CFD models at rotational speed and different air gap widths. It was observed that with an increase in velocity of the inner cylinder, the agreement between the CFD and the experimental results becomes poor. This is suspected to be as a result of not capturing the three-dimensional structures present within turbulent Taylor-Couette flows. As the air gap widened, the results were slightly more precise. As the rotor radius increases, the heat transfer coefficient increases, and the deviation between the empirical equations and the CFD fluid flow simulations at higher inner cylinder speeds also increases.

Comparison of thermal resistances calculated using empirical equations and CFD air gap fluid flow models, considering various air gap widths and rotation speeds, is shown in Fig. 6. The y-axis is displayed in logarithmic coordinates in Fig. 6. The differences among the thermal resistances calculated using the empirical equations and the CFD air gap fluid flow models are manifested at increasing speed and radius of the rotor.

The velocity profiles on line in the air gap for different rotation speeds are depicted in Fig. 7. Large velocity gradients can be seen near the walls, which is in good agreement with the theoretical assumptions stated in, e.g. [1].

Results from CFD air gap fluid flow models in Fig. 5 and in Fig. 6 are simulated as 2D axisymmetric flow models with swirl. One 2D simulation took about 460 seconds to solve for a total of 1440 simulations; thus, the solution time is approximately 170 hours. It took around 45 minutes to solve one 3D air gap CFD model.

Fig. 8 shows temperature contours on the middle surface of a 1.5 mm air gap at different rotor speeds. With increasing speed of the rotor, the temperature contours are more uniform and the temperature contours are getting closer to the average temperature in the air gap.

## V. Conclusion

This paper describes the thermal analysis of the air gap using empirical equations and computational fluid dynamics (CFD). Introduces empirical equations for calculating air gap heat transfer coefficients and presents the thermal air gap model. The paper then discusses numerical modelling, specifically the shear stress transport k-omega model, and presents parametric air gap models. Finally, the paper concludes with a comparison of air gap heat transfer coefficients and thermal resistances calculated from both empirical equations and CFD, along with the presentation of air gap velocity profiles with air gap temperature contours.

## References


[1] M. L. Hosain and R. B. Fdhila, "Air-gap heat transfer in rotating electrical machines: A parametric study," *Energy Procedia*, vol. 142, pp. 4176–4181, 2017, proceedings of the 9th International Conference on Applied Energy. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S187661021736085X

[2] G. Bramerdorfer, J. A. Tapia, J. J. Pyrhonen, and A. Cavagnino, "Modern electrical machine design optimization," *IEEE Transactions on Industrial Electronics*, vol. 65, no. 10, pp. 7672–7684, 2018. [Online]. Available: https://ieeexplore.ieee.org/document/8280552/

[3] K. Bersch, P. H. Connor, C. N. Eastwick, M. Galea, and R. Rolston, "Cfd optimisation of the thermal design for a vented electrical machine," in *2017 IEEE Workshop on Electrical Machines Design, Control and Diagnosis (WEMDCD)*. Nottingham, UK: IEEE, 2017, pp. 39–44. [Online]. Available: http://ieeexplore.ieee.org/document/7947721/

[4] J. Pyrhönen, T. Jokinen, and V. Hrabovcová, *Design of rotating electrical machines*, 2nd ed. Chichester: Wiley, 2014.

[5] D. A. Howey, P. R. N. Childs, and A. S. Holmes, "Air-gap convection in rotating electrical machines," *IEEE Transactions on Industrial Electronics*, vol. 59, no. 3, pp. 1367–1375, 2012. [Online]. Available: http://ieeexplore.ieee.org/document/5671481/

[6] D. Liang, Z. Q. Zhu, Y. Zhang, J. Feng, S. Guo, Y. Li, J. Wu, and A. Zhao, "A hybrid lumped-parameter and two-dimensional analytical thermal model for electrical machines," *IEEE Transactions on Industry Applications*, vol. 57, no. 1, pp. 246–258, 2021. [Online]. Available: https://ieeexplore.ieee.org/document/9219248/

[7] K. M. Becker and J. Kaye, "Measurements of diabatic flow in an annulus with an inner rotating cylinder," *Journal of Heat Transfer*, vol. 84, no. 2, pp. 97–104, 1962-05-01. [Online]. Available: https://asmedigitalcollection.asme.org/heattransfer/article/84/2/97/413579/Measurements-of-Diabatic-Flow-in-an-Annulus-With

[8] Y. A. Cengel and J. M. Cimbala, *Fluid mechanics*, third edition ed. New York: McGraw Hill, [2014].

[9] D. C. Wilcox, *Turbulence Modeling for CFD*, 3rd ed. San Diego, California: DCW Industries, Inc., November, 2006.

[10] B. Amini, H. M. Zoka, and H. Khaleghi, "A comparative assessment of variant k-epsilon models for engine flow applications," in *2011 International Conference on Computational and Information Sciences*. Chengdu, Sichuan, China: IEEE, 2011, pp. 1091–1094. [Online]. Available: http://ieeexplore.ieee.org/document/6086395/

[11] D. C. Wilcox, "Reassessment of the scale-determining equation for advanced turbulence models," *AIAA Journal*, vol. 26, no. 11, pp. 1299–1310, 1988. [Online]. Available: https://arc.aiaa.org/doi/10.2514/3.10041

[12] F. R. Menter, "Two-equation eddy-viscosity turbulence models for engineering applications," *AIAA Journal*, vol. 32, no. 8, pp. 1598–1605, 1994. [Online]. Available: https://arc.aiaa.org/doi/10.2514/3.12149

[13] F. R. Menter, "Improved two-equation k - omega turbulence models for aerodynamic flows," *NASA*, vol. 1992, no. Quick Release, pp. 1–8, October 1992. [Online]. Available: https://ntrs.nasa.gov/api/citations/19930013620/downloads/19930013620.pdf

[14] "K-omega turbulence models," 2023. [Online]. Available: https://www.simscale.com/docs/simulation-setup/global-settings/k-omega-sst/

[15] J. Súkeník, "Automated python-controlled thermal analysis of rotor of line-start permanent magnet synchronous machine," in *Proceedings II of the 29th Conference STUDENT EEICT 2023*. Brno: Brno University of Technology, Faculty of Electrical Engineering and Communication, 2023, pp. 105–108. [Online]. Available: https://www.eeict.cz/eeict_download/archiv/sborniky/EEICT_2023_sbornik_2_v2.pdf

# ŠKODA Kariéra


## Stáže a závěrečné práce

Studuješ vysokou školu a zároveň chceš získat pracovní zkušenosti? Zkus odbornou stáž ve ŠKODA AUTO. Pod vedením našich odborníků se zapojíš do inovativních projektů – třeba těch v oblasti elektromobility. Získané znalosti pak můžeš zúročit i ve své závěrečné práci.

## Doktorandský program

Aplikuj výsledky svého výzkumu v reálném prostředí, využívej nejnovější technologie a dej své disertační práci nový rozměr. Poskytneme ti flexibilní pracovní dobu a možnost spolupracovat s profesionály ve svém oboru.

## Trainee program

Dokončil jsi studia a přemýšlíš kam dál? Trainee program pro absolventy ti pomůže najít obor, který tě bude opravdu bavit a naplňovat. V průběhu jednoho roku budeš na plný úvazek pracovat v mezinárodním týmu, poznáš různá oddělení a vycestuješ na zahraniční stáž.

## Volné pozice

Vyber si ze široké nabídky volných pozic, staň se součástí nejprogresivnější české firmy a tvoř s námi budoucnost automobilového průmyslu.

# Intra-Pulse Modulation Recognition by Using the Ambiguity Function

Jan Virgala
*Dept. of CTEWR*
*University of Defence*
Brno, Czech Republic
ORCID 0000-0003-4867-5501

Marie Richterova
*Dept. of CTEWR*
*University of Defence*
Brno, Czech Republic
ORCID 0000-0003-1277-8955

***Abstract*—This paper presents the design and implementation of a signal classifier with intra-pulse modulation based on an ambiguity function. Linear frequency modulation, non-linear frequency modulation and intra-pulse modulation with Barker codes with different noise levels are used to experiment with the designed classifier. The Matlab programming environment was used for the conceptual design and implementation of the signal classifier with intra-pulse modulation.**

***Keywords—intra-pulse modulation, radar signals, ambiguity function, recognition***

## I. Introduction

It is very difficult to detect and capture signals with intra-pulse modulation in radio monitoring because these signals can be affected by various types of noise propagating in the radio channel. Intra-pulse modulations are used in radar technology. A very common case in monitoring is that signals with intra-pulse modulation are often partially degraded by noise and thus can be incorrectly identified. For this reason, new methods are being explored to identify and classify signals with intra-pulse modulation so that these methods are suitable for automatic classification and allow users to unequivocally identify these types of signals.

In the last few years, the world has developed methods for classification of intra-pulse modulations based on deep convolutional neural networks [1], [2], [3], [4], [5], the published classification methods are suitable for the off-line classification of intra-pulse modulations. The main goal of the authors is to find a suitable method for classification of intra-pulse modulations using an ambiguous function, which could be used for the on-line classification during radio monitoring in a real battlefield, where the immediate classification result and the subsequent countermeasures to eliminate the enemy are crucial.

Radar is a technology that emits electromagnetic waves capable of detecting objects. It can be used to determine the position, distance and speed of objects. Most of the radar consists of a transmitter and receiver. It works on the principle of emitting electromagnetic waves that bounce off objects from the back of the radar.

Based on the connectivity of the transmitted signal, there are two types of radar [6], [7]:

- CW radar - transmits a signal with constant energy, usually over short distances. Two separate antennas are required for continuous transmission and reception. This type of radar is usually used in police radars, motion detectors or parking sensors for cars.

- Pulse radar - this radar uses antenna mode switching between transmitting and receiving. The antenna first sends out a signal and then waits for the corresponding reflected signal. This method of operation is mainly used for measuring longer distances or for meteorological purposes.

Linear Frequency Modulation (LFM), Non-Linear Frequency Modulation (NLFM) and Barker code intra-pulse modulation are the most commonly used types of intra-pulse modulation for aerospace applications.

This paper describes the implementation of ambiguity function for the classification of LFM, NLFM and intra-pulse modulation with Barker codes.

The outline of this paper is as follows. The section II introduces the necessary information about intra-pulse modulations. The section III contains basic information about the ambiguity function and its characteristic shapes for the modulation techniques under investigation. The section IV deals with the design of an intra-pulse modulation classifier and the presentation of experimental results. Finally, in the section V, all the results obtained are evaluated and intentions for further research and implementation for radio monitoring are presented.

## II. Intra-Pulse Modulations

This chapter will describe the LFM, NLFM and Barker code intra-pulse modulations that are used in the design and implementation of an intra-pulse modulated signal classifier based on the ambiguity function. The authors selected the intra-pulse modulation classification method based on the analysis of real segments of intra-pulse signals captured during radio monitoring and on publications [6], [7] that analyze the characteristics of intra-pulse signals.

### *A. Linear Frequency Modulation*

A very important part of radar system design is the design of the time waveform of the transmitted signal. The intent is usually to maximize target detectability (the ability to intercept weak and distant targets) while maintaining the highest resolution at range (the ability to distinguish between two nearby targets). As shown in [6], improving target detectability can be achieved by increasing the total transmitted energy and more accurate distance resolution thereafter as the instantaneous bandwidth of the transmitted signal increases. Radars using a single constant frequency pulse with a rectangular complex envelope can only increase their transmitted energy from a signal perspective by extending the duration of the pulse. However, this will also reduce the maximum instantaneous bandwidth of the transmitted signal and thus inevitably reduce the distance resolution. This phenomenon is based on the physical uncertainty principle, which relates the maximum achievable resolution of two frequencies $\delta f$ in the signal spectrum to its duration $T$ as follows [7]:

$$\delta f \approx \frac{1}{T}. \tag{1}$$

A possible solution to separate these two design parameters is to introduce modulation of the transmitted signal. Here, the LFM, which is very common in radar systems, will be considered, the time waveform of which is also referred to as chirp. The instantaneous frequency of the chirp signal is scaled over the desired bandwidth $B$ over the chirp duration $Tc$ with a constant rate of change or chirp steepness $\alpha$:

$$df/dt = \alpha = konst. \tag{2}$$

By using LFM, it is possible to increase the bandwidth of a radar pulse of a given duration while preserving its transmitted energy [6], thus circumventing the limitations of equation (1) that apply to the unmodulated signal. As mentioned earlier, the use of the LFM signal also allows to determine not only the speed but also the distance of the target, as compared to a conventional CW radar.

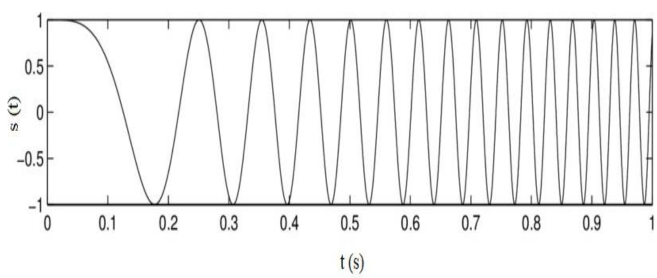



Fig. 1. LFM "upchirp" waveform

Fig. 1 shows the time waveform of the LFM "up-chirp" signal, where the increasing frequency of the signal over time can be seen. This signal starts with a lower frequency and gradually increases linearly over the duration of the pulse. This type of modulation is often used in modern radar systems because of its ability to achieve higher resolution and bandwidth. The advantage of LFM signals is their ability to effectively detect and identify targets in space while reducing the effects of interference and reflections from obstacles.

*B. Non-Linear Frequency Modulation*

The use of the NLFM brings several advantages over LFM. One of these advantages is that NLFM does not require the use of weighting windows in the frequency domain to suppress sideband frequencies. This is due to the fact that FM waveform modulation is designed to provide the desired spectral shape with the appropriate level of side lobes.

The NLFM uses a variable frequency deviation function, which causes the relationship between frequency and signal time to become nonlinear. This effectively changes the rate of phase change of the LFM waveform, with less time spent at the edges of the frequency band. This process leads to suppression of the sideband components of the signal. The instantaneous phase of the signal is also an important characteristic of the NLFM [8].

Non-linear frequency modulation (NLFM) is an alternative to LFM. Its basic idea is a non-linear change in frequency. It allows us to shape the power spectral density of the pulse so that the autocorrelation function has much smaller side lobes compared to LFM. As a result, no additional filtering is required and the maximum Signal to Noise Ratio (SNR) is maintained. This gives us a 1 to 2 dB advantage over LFM with equivalent sidelobe filtering. Also, compared to LFM, we can shape the spectrum differently and spend more time on the frequencies where it is needed. However, the disadvantage is the complexity of designing and creating accurate pulses and processing them [8], [10]. An example of NLFM signal is shown in Fig. 2.

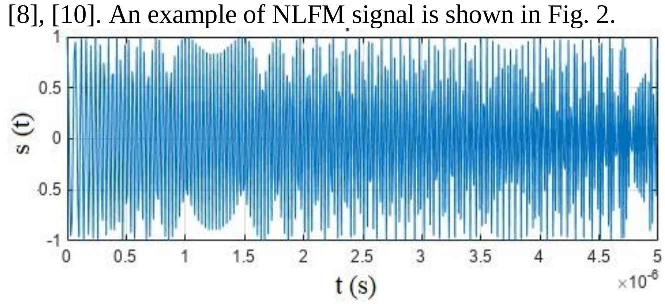



Fig. 2. NLFM waveform

*C. Intra-Pulse Modulation with Barker Codes*

The Barker codes were introduced in 1953 by Ronald Hugh Barker. They are classified as binary phase codes, which means that the phase takes on two states (0 and π) during the pulse. The Barker codes are specific by suppressing the side lobes in the autocorrelation function. In total, 7 different Barker codes are known, a list of them is given in [8].

We can also encounter so-called composite Barker codes (a Barker code inside a Barker code). We can illustrate this with the example of $Nc$=16 (a 4th order Barker code inside a 4th order Barker code) see Fig. 3. Although there is more side-lobe suppression in compound Barker codes than in normal codes, this suppression is not directly proportional. The composite code $Nc$=169 is often used. That is, a combination of 13th-order Barker codes that is also the longest possible binary sequence from a single concatenation.

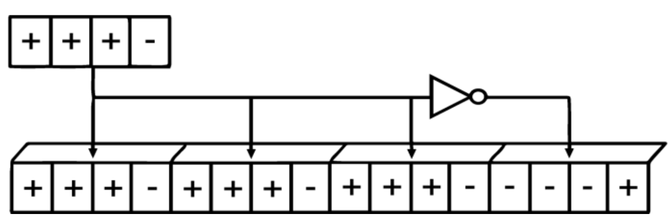



Fig. 3. Composite Barker code Nc=16. Source [8]

The Barker codes are the most commonly used binary code due to the fact that in the uncertainty function at zero Doppler shift, the minor lobes are no larger than *1/Nc*, where Nc is the order of the Barker code and 1 is the relative size of the major lobe. Sometimes they are even called "perfect codes". However, they have their drawbacks. In addition to the limited sequence length, there is also a high sensitivity to Doppler shift, as can be seen from the uncertainty function in Fig. 4. It is also not classified as an LPI (Low Probability of Intercept) signal, since it can be easily picked up by a receiver that uses frequency doubling. In this technique, the received signal is multiplied by itself and the result processed by an envelope detector.

The Barker code cannot be determined directly from the time waveform (Fig. 4). It may be confused with a normal BPSK signal. On the other hand, the autocorrelation function of the Barker code is very characteristic and we can clearly define from it that it is a Barker code, as can be seen in Fig. 5.

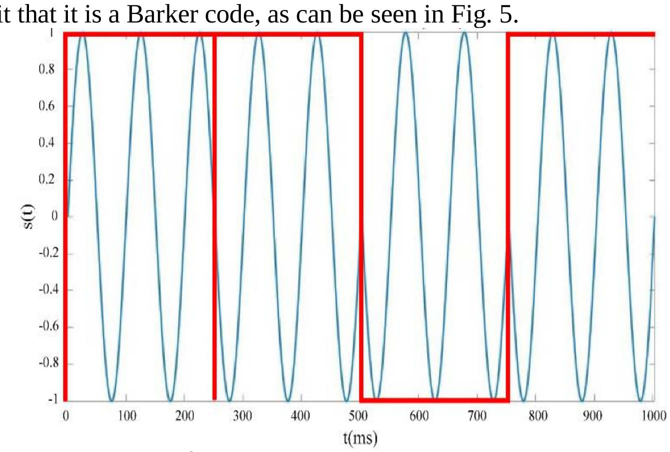



Fig. 4. Waveform of 4th order Barker code

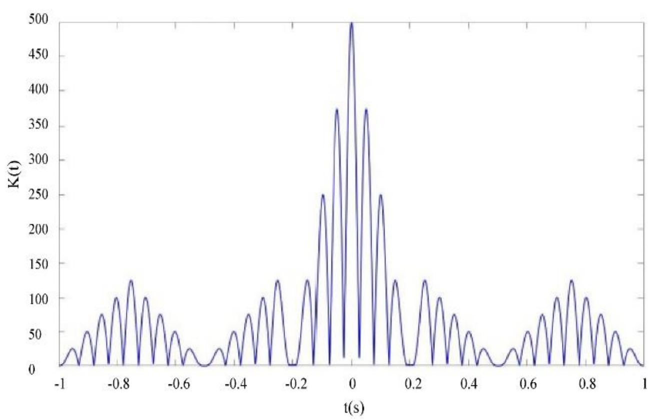



Fig. 5. Autocorrelation function of the 4th order Barker code

The autocorrelation function of the Barker code has an unmistakable waveform and is more suitable for identification than the time-frequency waveform of the code, which may be shape-interchangeable with another kind of signal, provided the useful signal is partially hidden in the noise.

## III. Ambiguity Function

The ambiguity function is very important in time-frequency signal processing because it is closely related to the Wigner-Ville distribution via the two-dimensional Fourier transform. This relationship is very important for the formulation of other time-frequency distributions [9].

The ambiguity function is a two-dimensional function of the propagation delay of the pulse radar and sonar signal processing. The ambiguity function represents the distortion of the returned pulse due to the filter adapted to the receiver when the signal returns from a moving target. The ambiguity function is defined by the characteristics of the pulse and the filter. The monostatic ambiguity function is derived from the matched filter, the multistatic ambiguity function is derived from the corresponding optimal multistatic detector. The nature of this detection algorithm depends on whether the target motions are correlated or not. If they are, the optimal detector performs a phase-coherent summation of the received signals, resulting in very high target localization accuracy.

The motivation for performing Frequency Time Domain Analysis is that the classical Fourier transform cannot determine the temporal localization of the individual frequency components and is therefore only suitable for periodic signals. For aperiodic signals, it is appropriate to use the so-called time-frequency transforms. The ambiguity function has a specific and non-interchangeable form for each intra-pulse signal. Based on this fact, the uncertainty function can be used in the classification process of signals with intra-pulse modulation. The following Figs. 6 to 9 show the ambiguity functions obtained by experiments with real intra-pulse signals of LFM, NLFM and intra-pulse signals with Costas code.

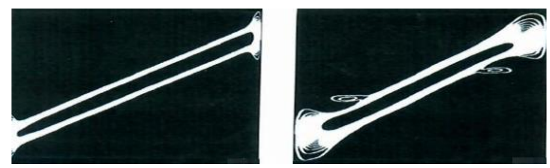


Fig. 6. LFM with SNR = 15 dB and LFM with SNR = 7 dB

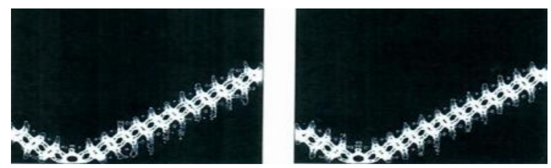


Fig. 7. NLFM with SNR = 15 dB and NLFM with SNR = 7 dB

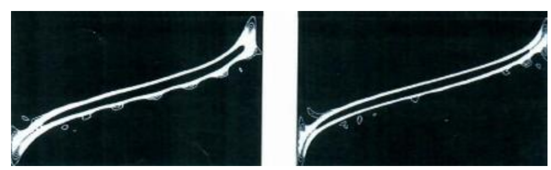


Fig. 8. NLFM with SNR = 7 dB and NLFM with SNR = 10 dB

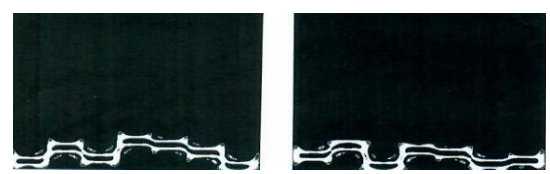


Fig. 9. Intra-pulse signal with Costas code with SNR = 10 dB

Experiments have demonstrated that each intra-pulse signal has an unambiguous and unambiguous shape of the uncertainty function. Each ambiguity function retains its characteristic shape even for different SNRs. This property as one of the basic features can be used to design an intrapulse modulation classifier.

In order to use the ambiguity function to classify intra-pulse signals, it was necessary to perform a large number of experiments with both deterministic intra-pulse signals and real segments of intra-pulse signals. In the case of deterministic intrapulse signals, the SNR level was set from 5 to 10 dB and the change in the shape of the ambiguity function was monitored. The experiments showed that the basic shape of the ambiguity function was preserved, but depending on the SNR level, different energy levels occurred in the obtained pattern. Images of the normalized and centered ambiguity functions of the LFM, NLFM, and Barker-coded intra-pulse signals were stored in a pattern database.

## IV. Design of Intra-Pulse Signal Classifier

The experiments with ambiguity functions for deterministic and real segments of intra-pulse signals have shown that it is also necessary to take into account the characteristics of the radio channel that affect the quality of the signal captured by the receiving antenna during radio monitoring. If intra-pulse signals are to be detected in real time, it is necessary to develop methods that are capable of instantaneously detecting the intrapulse signal of interest with a specified probability. One possible way is to use the features of the ambiguity function to classify the intra-pulse signals. LFM, NLFM, and intra-pulse Baker code signal were chosen for the experiments, since these mentioned signal types are very often present in the real frequency spectrum in aeronautical applications.

The Matlab software environment version 2022b was used for the design and implementation of the intra-pulse modulation classifier. Standard built-in functions and original algorithms designed and implemented by the authors were used. The *comm.RayleighChannel* function was used to simulate the radio channel with Reyleigh fading. The radio channel was simulated for six paths to obtain results similar to real signals captured during radio monitoring. The experiments were performed on an ASUS Zenbook 14 (UM3402) laptop equipped with an AMD Ryzen 7 5800X processor.

In the design of the intra-pulse signal classifier, a database of ambiguity function images for the selected deterministic intra-pulse signals with Additive White Gaussian Noise (AWGN) and SNR of 5, 7 and 10 dB was firstly created. The SNR interval was chosen to simulate a real radio channel with multipath propagation and additive noise when the receiver is still able to capture and process the signal (for simultaneous devices, the minimum threshold is 5 dB).

Next, the ambiguity function images of ideal intra-pulse signals without noise were stored in the database. Subsequently, a set of deterministic intra-pulse signals with AWGN and Rayleigh fading were added. The database was supplemented with ambiguity function images of real LFM, NLFM, and Barker-coded real intra-pulse signals. The database contains a total of 98 uncertainty function patterns, of which 28 are for LFM, 34 for NLFM and 36 for intra-pulse signals with Barker codes.

The second step was the design and implementation of an intra-pulse signal classifier. The block diagram of the classifier is shown in Fig. 10.

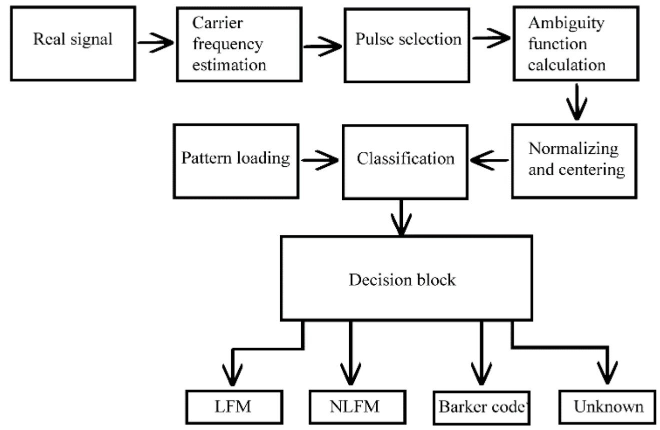



Fig. 10. Block diagram of the intra-impulse signal classifier

The real signal is passed to the input of the classifier and an estimation of the carrier frequency is performed. One pulse is selected from the group of pulses and the ambiguity function is calculated. The normalization and centering process of the ambiguity function is followed by the classification process. The ambiguity function patterns from the database are successively input to the classifier and the probability of matching the real signal with the pattern is tested.

The decision block assigns the classified real signal to four outputs. Thus, the real signal is classified as LFM, NLFM or intra-pulse signal with Barker code. If the classifier has not found a pattern match with the database, then the real signal is classified as unknown.

The classifier was tested with deterministic intra-pulse signals with AWGN, the experimental results are presented in Table I.

TABLE I. EXPERIMENTS WITH DETERMINISTIC SIGNALS

| **Signal** | **Intra-Pulse Signals with AWGN** | | | | |
|---|---|---|---|---|---|
| | ***SNR (dB)*** | ***Correct*** | ***Incorrect*** | ***Unknown signal*** | ***Total success rate (%)*** |
| LFM | 5 | 635 | 274 | 91 | 64% |
| | 7 | 826 | 152 | 22 | 83% |
| | 10 | 990 | 10 | 0 | 99% |
| NLFM | 5 | 621 | 243 | 136 | 62% |
| | 7 | 795 | 135 | 70 | 80% |
| | 10 | 966 | 34 | 0 | 97% |
| Barker code | 5 | 321 | 262 | 293 | 32% |
| | 7 | 611 | 111 | 75 | 61% |
| | 10 | 680 | 9 | 15 | 68% |

The classifier for LFM had the highest success rate, ranging from 99 to 64%. NLFM has a slightly lower classification success rate overall, but the decline in correct classification success follows the same trend as for LFM. The Barker code represents intra-pulse signals with phase change. The correct classification results range from 68% to 32%. Experimental results for signals with AWGN showed that for an SNR ratio of 10 dB, the overall correct classification success rate is higher than for an SNR ratio of 5 dB. The classifier was tested with deterministic intra-pulse signals with AWGN and Rayleigh fading, the experimental results are presented in Table II.

TABLE II. EXPERIMENTS WITH DETERMINISTIC INTRA-PULSE SIGNALS

| **Signal** | **Intra-Pulse Signals with AWGN and Rayleigh fading** | | | | |
|---|---|---|---|---|---|
| | ***SNR (dB)*** | ***Correct*** | ***Incorrect*** | ***Unknown signal*** | ***Total success rate (%)*** |
| LFM | 5 | 556 | 363 | 81 | 56% |
| | 7 | 640 | 351 | 9 | 65% |
| | 10 | 789 | 211 | 0 | 79% |
| NLFM | 5 | 521 | 358 | 136 | 52% |
| | 7 | 635 | 351 | 70 | 64% |
| | 10 | 940 | 34 | 14 | 94% |
| Barker code | 5 | 421 | 262 | 202 | 42% |
| | 7 | 608 | 101 | 65 | 61% |
| | 10 | 830 | 5 | 7 | 83% |

The experimental results for intra-pulse signals with AWGN and Rayleigh fading show more classification fails than for signals with AWGN. The only exception is the Barker code, which has a slightly increased identification success rate compared to the first experiment. For LFM, an increase in classification failures can be observed, which is probably due to the fact that deterministic LFM without noise has no symptoms that start to appear when the intra-pulse signal passes through a radio channel with multipath propagation. The NLFM retains the dominant symptoms, so the experiment shows only a slightly reduced correct classification success rate.

After performing experiments with deterministic signals, it was necessary to perform experiments with segments captured during real radio monitoring. The results of the experiments are presented in Table III. The classification of the unknown real intra-pulse signal was performed in 2.4 seconds for a database with 98 patterns of uncertainty functions.

The authors' proposed and implemented function for intra-pulse signal classification using the ambiguity function database is shown in Fig. 11.

```
function [Korel] = klasifikace(W)
 %Classification: The function calculates the correlation coefficient between
 %computed matrix of the signal under investigation and the matrices in the database.
% The input is the matrix of the signal under study and
%output is the row vector of the calculated correlations
   matice1=W; % Assignment of input to a variable
   load("database.mat") %Loading a cell structure with a database
   for n = 1:length(WW) %Compare all matrices in the database
      matice2=cell2mat(WW(1,n)); %Conversion of cell structure to a matrix
      r=corrcoef(matice1,matice2); %Comparison of matrices between each other
      Korel(1,n)=r(2,1); %Write to the result vector
   end
```

Fig. 11. The algorithm for classification using the ambiguity function pattern database.

TABLE III. EXPERIMENTS WITH REAL INTRA-PULSE SIGNALS

| **Signal** | **Real Intra-Pulse Signals** | | | |
|---|---|---|---|---|
| | ***Correct*** | ***Incorrect*** | ***Unknown signal*** | ***Total success rate (%)*** |
| LFM | 238 | 25 | 61 | 73% |
| NLFM | 211 | 59 | 89 | 59% |
| Barker code | 152 | 92 | 29 | 56% |

The experimental results for real segments of intra-pulse signals show that the success rate of correct identification ranges from 56 to 73%. For LFM, NLFM signals, it can be observed that the number of unknown signals is larger than the number of incorrectly classified signals. This is due to the fact that these are new signals with different number of samples in the segment, so the overall correlation between these intra-pulse signals and the patterns of intra-pulse signals in the database is lower. The number of unknown and incorrectly classified signals could probably be reduced by adding a pattern to the database. In contrast, the intra-pulse signals based on phase change (Barker codes) show a higher number of incorrectly classified segments compared to the experimental results in Table I and Table II.

## V. CONCLUSION

The experiments demonstrated that the proposed and implemented intra-pulse signal classifier is able to identify real segments of LFM, NLFM signals and real Barker code signals. The success rate of correct classification ranges from 56% for the intra-pulse Barker code signal to 73% for the LFM signal. The success rate of classification of real segments of intra-pulse signals will always depend on a properly constructed database of the ambiguity function patterns of the investigated modulation types.

An application (*.mlapp) named IPCAF (Intra-Pulse Classifier Ambiguity Function) was created in the Matlab version 2022b programming environment for experiments with deterministic and real intra-pulse signals.

The results presented in Tables I, II and III show that the designed and implemented intra-pulse modulation classifier is able to identify signals with more than 50% probability for real signal segments. Experiments with real signal segments have shown that it is appropriate to consider a combination of several frequency-time analysis methods and methods that use static characteristics and correlation.

Experiments confirmed that additional patterns can be added to the ambiguity function pattern database, as the classifier response time of approximately 2.5 seconds is adequate for on-line monitoring of intra-pulse signals. The authors develop the knowledge gained so that the classifier can be modified for additional intra-pulse signals. The design and implementation of combined parallel intra-pulse signal classifier is the subject of further research.

## ACKNOWLEDGMENT

This paper was created with a support of the project AIROPS of University of Defence, Brno - Research of Conduct of Airspace Operations. I would like to thank my supervisor, Associate Professor Marie Richterova, for her valuable advice and support during my PhD studies.

## REFERENCES

[1] W. Si, C. Wan and Z. Deng. "Intra-Pulse Modulation Recognition of Dual-Component Radar Signals Based on Deep Convolutional Neural Network." IEEE Communications Letters, vol. 25, no. 10, october 2021.

[2] F. C. Akyon, Y. K. Alp, G. Gok and O. Arikany. "Classification of Intra-Pulse Modulation of Radar Signals by Feature Fusion Based Convolutional Neural Networks." In 26th European Signal Processing Conference (EUSIPCO), 2018.

[3] Z. Qu, X. Mao and Z. Deng. "Radar Signal Intra-Pulse Modulation Recognition Based on Convolutional Neural Network." IEEE Access, vol. 6, pp. 2169-3536, 2018.

[4] Z. Qu, C. Hou, C. Hou and W. Wang. "Radar Signal Intra-Pulse Modulation Recognition Based on Convolutional Neural Network and Deep Q-Learning Network." IEEE Access, vol. 8, pp. 49125 - 49136, 2020.

[5] S. Wei, Q. Qu, H. Su, M. Wang, J. Shi and X. Hao. "Intra-pulse modulation radar signal recognition based on CLDN network." IET Radar, Sonar & Navigation, vol. 14 Iss. 6, pp. 803-810, 2020.

[6] M. A. Richards, Fundamentals of radar signal processing. New York: McGraw-Hill, 2005.

[7] G. Brooke, "Understanding millimetre wave FMCW radars." In 1st International Conference on Sensing Technology, 2005.

[8] L. Nadav and E. Mozeson, Radar Signals, Wiley-IEEE Press, 2004.

[9] S. Qian and D. Chen, Joint Time-Frequency Analysis: Methods and Applications, Chap. 6, Prentice Hall, N.J., 1996.

[10] D. A. Hague and J. R. Buck. "The generalized sinusoidal frequency-modulated waveform for active sonar." IEEE Journal of Oceanic Engineering 42.1 (2016): 109-123.

# Modelling of Magnetic Films: A Scientific Perspectives

Denis Misiurev, Vladimír Holcman
*Department of Physics*
*Brno University of Technology*
Brno Czech Republic
xmisiu00@vutbr.cz, holcman@vut.cz

***Abstract* — Modeling magnetic thin films represents a dynamic intersection of scientific inquiry and technological progress, at the forefront of materials science exploration. Researchers use various computational methods, such as Monte Carlo simulations and molecular dynamics, to understand magnetism and thin-film growth on different surfaces. Recent advancements in multiscale modeling and machine learning have improved predictive abilities, leading to a better understanding of thin-film dynamics over different spatial and temporal scales. This interdisciplinary approach, coupled with advanced experimental techniques like in situ microscopy, promises significant advancements in magnetic materials. These advancements have wide-ranging implications in areas such as magnetic data storage, spintronics, and magnetic sensors. The integration of computational modeling and experimental validation marks a new era of scientific rigor, providing deep insights into the real-time behavior of magnetic films and enhancing the accuracy of predictive models. As researchers navigate unexplored territory, the field of magnetic thin-film modeling holds great promise for unlocking new possibilities in materials science and engineering. Through a combination of theoretical exploration and empirical validation, magnetic thin-film modeling is poised to drive innovation and revolutionize various industries in the future.**

***Keywords — Magnetic films; substrates; first–principles calculation; density functional theory; electronic structure; molecular dynamics simulations; interface phenomena; Monte Carlo simulation.***

## I. Introduction

Modeling techniques have significantly advanced our understanding of magnetic phenomena by offering crucial insights into the underlying physical mechanisms, interactions, and responses. These methodologies encompass a diverse range of approaches, from atomistic simulations to continuum methods, each offering distinct advantages and perspectives. As we explore the scientific dimensions of modeling magnetic films on various substrates, it becomes evident that a thorough examination of these techniques is necessary to decipher the complex nature of magnetic behavior.

The main objective of this review is to present a comprehensive overview of the current state–of–the–art modeling techniques utilized in the study of magnetic films on diverse substrates. This exploration will be structured as follows: First, we will delve into the foundational concepts underlying magnetic interactions and phenomena observed in thin films on diverse substrates. Subsequently, we will embark on a journey through various modeling techniques, starting with atomistic simulations aimed at capturing fundamental atomic–scale interactions.

Internal Grant Agency (FEKT–S–23–8228)

Throughout our discussion of these modeling techniques, we will emphasize the unique insights they provide into different aspects of magnetic film–substrate systems, such as surface effects, domain wall dynamics, and magnetic anisotropy. Moreover, we will underscore the significance of the synergy between experimental observations and theoretical predictions, illustrating how modeling enhances our understanding by elucidating complex phenomena that may not be readily discernible through experiments alone.

In the extensive realm of scientific exploration, modeling techniques serve as potent tools enabling us to peer into the intricate mechanisms governing the behavior of magnetic films on various substrates. Through a multidisciplinary approach drawing upon physics, materials science, and computational methodologies, our aim is to contribute to the evolving discourse surrounding magnetic materials and their applications. By unraveling the scientific perspectives offered by diverse modeling techniques, we embark on a journey toward harnessing the full potential of magnetic films for technological innovation.

## II. Quantum Mechanical Modeling

Quantum mechanical modeling is a powerful tool for studying the electronic structure and properties of magnetic films on substrates. It offers detailed insights into electron behavior, enabling calculations of energy levels, wave functions, and electronic interactions [1].

These models, grounded in fundamental physics principles, provide unparalleled accuracy in understanding electron behavior, including spin orientations and magnetic interactions [2]. By delving into quantum states and electron distributions, these models shed light on phenomena like magnetism switching and domain wall movement. They also facilitate the design of novel materials and configurations, driving advancements in magnetic-based technologies [2].

However, the computational demands and time constraints of quantum mechanical calculations pose challenges, often requiring approximations and abstractions [3]. Additionally, accommodating complex real-world scenarios, such as multicomponent materials and intricate geometries, demands advanced theoretical and computational approaches. While quantum mechanical models yield precise numerical outcomes, their interpretation may require a deep understanding of quantum physics principles [3].

In the realm of quantum mechanics, researchers explore quantum molecular dynamics to understand complex molecular systems. This approach, based on the Born–Oppenheimer approximation, separates variables into nuclei and electrons. Car–Parrinello molecular dynamics (CPMD) simultaneously computes electronic and atomic subsystems using function minimization techniques. Nuclei trajectories are determined by their coordinates {Ri|i = 1,…,Nn}, while electronic degrees of freedom are described by quantum-mechanical wave functions {Ψj|j = 1,…,Ne} [4].

$$M_i \ddot{R}_i = -\frac{\partial E}{\partial R_i} = F_i \quad (1),$$

$$\mu \ddot{\psi}_j(r,t) = -\hat{H}\psi_j(r,t) + \sum_k \wedge_{jk}\, \psi_k(r,t) \qquad (2)$$

In the given context, the symbol $F_i$ represents the resultant force exerted on the atom. E and Ĥ represents the energy functional and the Kohn–Sham Hamiltonian. μ denotes the fictitious electron mass. $\wedge_{jk}$ stands for indefinite Lagrange multipliers [4].

The selection of a fictitious electron mass serves the purpose of expediting the convergence of the algorithm, a crucial aspect in computational simulations. Concurrently, in the Car–Parrinello framework, Lagrange multipliers play a pivotal role in generating additional forces that enforce the orthonormality of the wave functions [4].

In the Car–Parrinello theory, the total energy of a nanosystem is intricately linked to the coefficients in the expansion of the electron wave function over a given basis. By minimizing specific coefficients, the system undergoes a process of cooling and stabilization, ultimately reaching equilibrium [5]. This equilibrium is achieved through the iterative solution of classical Newtonian motion equations for a set of particles until stability is attained. However, it's essential to note that while this method provides a means to simulate the system's evolution, it does not directly represent the genuine dynamics of the nanosystem; rather, it leads to an equilibrium state with stable energy for a multi-electron particle system [5].

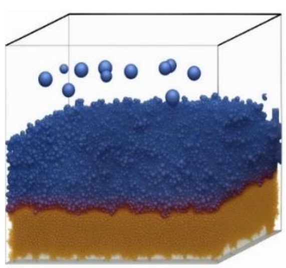


Fig. 1. Simulated $TiO_2$ thin film growth [6].

The Car–Parrinello molecular dynamics method has gained recognition for its capability to accurately replicate the properties of semiconductor and dielectric materials. Nonetheless, challenges arise when dealing with metallic systems in close proximity to the band gap, where even minor alterations in total energy can induce significant fluctuations in electron density. Moreover, the applicability of the Car–Parrinello approach is constrained by the complexity of solving the equations and the associated computational expenses, limiting its use to smaller nanosystems [7].

An alternative to Car–Parrinello molecular dynamics is the Born–Oppenheimer molecular dynamics (BOMD), founded on the Born–Oppenheimer approximation. This method involves the segregation of nuclear and electronic subsystem descriptions. While the motion of the nuclei is governed by classical mechanics equations, the energy and forces acting on them are computed by solving the Schrödinger equation for electronic wave functions at each time step. This approach proves advantageous in managing the complexity of metallic systems and circumventing computational challenges associated with larger nanosystems [8].

$$M_i \ddot{R}_i = -\nabla_i [\min_{\psi_1,\dots,\psi_{N_e}} E(R_1,\dots,R_{N_n};\, \psi_1,\dots,\psi_{N_e})] \ (3)$$

In this context, the total energy (E) is minimized by optimizing electron wave functions while keeping nuclear coordinates fixed. This often involves solving Kohn–Sham equations or using multidimensional optimization algorithms. Born–Oppenheimer molecular dynamics (BOMD) segregates nuclear and electronic subsystems, accurately describing dielectrics and metals, but both BOMD and Car–Parrinello molecular dynamics (CPMD) are computationally expensive, limiting their use to smaller nanosystems [9].

Efforts to overcome this limitation focus on combining classical molecular modeling with electronic structure calculations. One approach is employing embedded atom potentials like the Embedded Atom Model (EAM) and Modified Embedded Atom Method (MEAM). Derived from electron density functional theory, EAM describes nanosystem behavior through equations, reducing computational complexity and broadening applicability to larger systems.

$$M_i \ddot{R}_i = -\frac{\partial E}{\partial R_i} = F_i, \qquad (4)$$

$$U_i = F(\rho_i) + \sum_{J;1\cdot\neq j} \phi(|R_i - R_j|) \quad (5)$$

$$\rho_i = \sum_{J;1\cdot\neq j} \Psi(|R_i - R_j|) \quad (6)$$

The Born–Oppenheimer molecular dynamics (BOMD) method is renowned for its accuracy in describing materials like dielectrics and metals. However, both BOMD and Car–Parrinello molecular dynamics (CPMD) incur significant computational costs, limiting their practical use to small nanosystems. This high accuracy of quantum mechanics methods is intricately tied to the complexity of the algorithms and approaches employed, presenting a notable limitation.

To mitigate the challenges posed by the complexity of quantum mechanics models, various assumptions and modifications have been introduced. These aim to eliminate limitations and reduce the complexity and dimensionality of solved problems by constraining the degrees of freedom of the studied objects. Key approximations include the Born–Oppenheimer approximation, which separates variables into massive nuclei and fast electrons, and the Hartree–Fock approximation, which simplifies Schrödinger equations by correlating one-electron functions instead of multi-electron wave functions [10].

A significant aspect of quantum mechanics methods involves the theory of electron density functional. Thise addresses increasing complexity and dimensionality as the number of observed particles rises. In this approach, the energy state of a nanosystem, involving nuclei and electrons interacting in an external field, is described as a functional dependent on particle density. Various developments and descriptions of exchange-correlation energy, such as Thomas–Fermi theory, Hohenberg–Kohn theorems, Kohn–Sham formalism, local density approximation, generalized and meta-generalized approximations, and hybridization of the generalized gradient method, underscore the relevance and ongoing advancement of electronic density functional theory.

## III. Monte Carlo Simulations

While quantum mechanical modeling offers crucial insights into the electronic structure, classical modeling approaches are frequently employed to investigate the dynamics and thermal fluctuations in magnetic films on substrates. These approaches enable the simulation of large–scale systems and longer time scales, proving particularly valuable for studying systems characterized by randomness, uncertainty, and nonlinearity, spanning from quantum mechanics and statistical physics to complex biological processes.

Monte Carlo simulation employs random sampling techniques to examine the thermodynamic properties and thermal fluctuations of magnetic films on substrates [11]. By accounting for temperature, exchange interaction, and anisotropy effects, Monte Carlo simulations can explore phase transitions, domain formation, and magnetization dynamics within the system. They are also adept at elucidating the behavior of large and intricate systems that defy traditional analytical methods, providing statistical insights into temperature–dependent properties of film–substrate systems [11].

In Monte Carlo simulation, the deposition process is calculated by tracking the movement and interaction of individual atoms or molecules [12]. These simulations consider parameters such as incident particle flux, energy, and surface interactions to determine the probability of deposition at various sites on the substrate. Through numerous interactions in these stochastic processes [12], Monte Carlo simulations furnish statistical details about the resulting film growth, including its microstructure, morphology, and surface roughness [12]. Consequently, researchers can employ Monte Carlo simulations to investigate the influence of deposition conditions such as temperature, pressure, and induced flux on thin film growth. These simulations facilitate the exploration of various deposition parameters and their effects on film properties [13]. Additionally, Monte Carlo simulations offer insight into processes like nucleation, island growth, coalescence, and other phenomena occurring during thin film deposition.

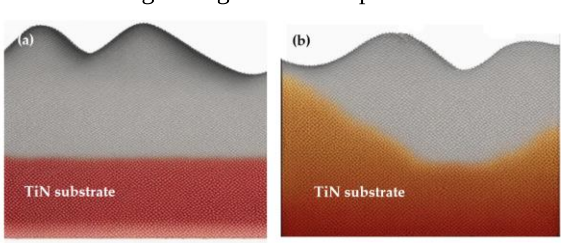



Fig. 2. Simulated formation of thin films on TiN substrate [6].

The Monte Carlo Metropolis algorithm provides a natural approach for simulating temperature effects in scenarios where dynamic considerations are unnecessary, owing to its rapid convergence to equilibrium and relatively straightforward implementation [14].

In the context of a classical spin system, the Monte Carlo Metropolis algorithm unfolds as follows: Initially, a stochastic selection is made of a specific spin index *i*, followed by a random alteration of its initial spin direction $S_i$ to a novel trial configuration $'S_i'$, referred to as a "trial move" [14]. Subsequently, the change in energy $\Delta E = E(S'_i) - E(S_{0i})$ between the pre–existing and updated states is evaluated, and the trial move is accepted with a probability determined by a specified criterion [14].

$$P = \exp\left(-\frac{\Delta E}{K_b T}\right) \qquad (7)$$

The determination of acceptance in the Monte Carlo algorithm involves comparing the calculated probability with a randomly generated number from a uniform distribution between 0 and 1 (Eq. 7) [14]. Probabilities exceeding 1, indicating energy reduction, are accepted without conditions [14]. This iterative process continues until a total of N trial moves have been executed, where N represents the count of spins within the entire system. Each set of N trial moves constitutes a single iteration of the Monte Carlo algorithm [14].

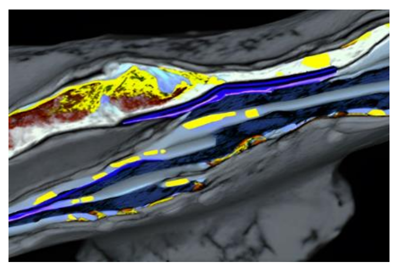


Fig. 3. Simulated magnetic thin film [6].

The nature of the trial move holds significance concerning two fundamental requirements for any Monte Carlo algorithm: ergodicity and reversibility [14]. Ergodicity necessitates that all potential states of the system are accessible, while reversibility mandates that the transition probability between two states remains invariant. Equation (7) demonstrates evident reversibility as $P(Si \rightarrow S'i) = P(Si \rightarrow S'i)$, where the probability of a spin change depends solely on the initial and final energy [15].

Ergodicity can be achieved by relocating the selected spin to a random position on the unit sphere [15]. However, this approach is problematic at low temperatures, where significant deviations of spins from the collinear direction are improbable due to the strength of the exchange interaction [16]. Consequently, at low temperatures, a sequence of trial moves on the unit sphere often leads to the rejection of the majority of moves [16]. Ideally, an acceptance rate of approximately 50% is desired, as excessively high or low rates require a significantly larger number of Monte Carlo steps to reach a state representative of true thermal equilibrium [16].

By integrating Monte Carlo simulations with experimental data and other modeling techniques, researchers can enhance their comprehension of thin film growth mechanisms [17]. Optimization of the positional processes and prediction of resulting film properties enhance the integration of simulation and experimentation, enabling more efficient and targeted development of films for various applications [17]. While Monte Carlo simulation and dynamic modeling shed light on dynamic behavior and thermal phenomena, these modeling approaches are essential for elucidating the magnetic properties, transport behavior, and response of magnetic films on substrates, thereby facilitating the design and optimization of magnetic devices.

## IV. DENSITY FUNCTIONAL THEORY

At the core of density–functional theory lie the Hohenberg–Kohn and Kohn–Sham theorems. These foundational principles of DFT readily lend themselves to extensive adaptation beyond their initial formulations, addressing a broad spectrum of physical scenarios.

The Landau–Lifshitz–Gilbert calculations for thin films involve several key steps. Initially, geometry optimization is conducted to determine the optimized atomic positions and lattice parameters of the thin film structure [18]. This process entails minimizing the total energy of the system by iteratively adjusting the atomic positions until coherence is achieved [19]. Total energy calculations typically utilize appropriate exchange–correlation functionals, such as the generalized gradient approximation (GGA) or hybrid functionals [19]. Subsequently, electronic structure calculations are performed. Once the optimized geometry is obtained, the electronic structure of the thin film is computed [19]. This step involves solving the Kohn–Sham equations (as per the Hohenberg–Kohn theorem), which describe the behavior of electrons in the system [20]. By self–consistently solving these equations (Eq. 8), information about the energy levels, band structure, and density of states of the film can be extracted [20].

$$\widehat{H}_e = \sum_i\left(-\frac{\hbar^2}{2m}\frac{\partial^2}{\partial {r_i}^2} - \sum_I \frac{Z_I e^2}{|r_i - R_j|}\right) + \sum_{i<j}\frac{e^2}{|r_i - r_j|} \qquad (8)$$

In this expression, $H_e$ represents the electronic Hamiltonian. The first term corresponds to the kinetic energy of electrons, where $\hbar^2/2m\ (\partial^2/\partial r_i^2)$ denotes the kinetic energy operator. The second term represents the potential energy arising from the interaction between the i–th electron and a nucleus situated at $R_j$, expressed as $(Z_k e^2)/|r_i – R_j|$. The third term accounts for the potential energy arising from the electron–electron interaction, denoted as $e^2/|r_i – r_j|$ for all distinct pairs (i,j) of electrons.

In Density Functional Theory (DFT), the electron density $\rho(r)$ serves as the fundamental variable, contrasting with the many–body wave function utilized in traditional quantum mechanics [21]. The electron density proves considerably more tractable than the many–body wave function as it depends solely on three spatial coordinate variables, regardless of the number of electrons in the system [21]. This simplification is justified by the Hohenberg–Kohn theorem, which establishes a one–to–one correspondence between the external potential and the ground–state electron density, thus rendering the electron density a sufficient and convenient descriptor for the quantum system [21].

Consequently, if the ground–state electron density is known, the external potential is uniquely determined. Moreover, physical properties associated with the ground–state wave function can, in principle, be unambiguously derived from the electron density [22]. Specifically, the kinetic and electron–electron interaction energies of the ground state can be expressed as universal functionals of the electron density, denoted as $E_{\text{kin}}[\rho]$ and $E_{\text{ee}}[\rho]$ respectively [22]. The term "universal" indicates that the functional forms are independent of the specific external potential [22].

The theorem also introduces a variational principle. Defining the energy functional:

$$E_v[\rho] = E_{kin}[\rho] + \int \rho(r)v(r)dr + E_{ee}[\rho] \quad (9)$$

for some external potential $V(r)$, the functional satisfies the inequality:

$$E_v[\rho] \geq E_v[\rho_0] = E_0, \quad (10)$$

where $\rho_0$ ($E_0$) is the ground–state electron density (energy) under the potential $V(r)$, respectively. Hence, the ground–state electron density can be obtained by seeking the electron density that minimizes the energy functional $Ev[\rho]$ [23].

Although the variational principle in Eq. 9,10 appears straightforward, a significant challenge lies in the fact that the exact forms of the functionals, $E_{\text{kin}}[\rho]$ and $E_{\text{ee}}[\rho]$, are unknown [23]. To tackle this issue, Kohn and Sham introduced the concept of "orbitals" to approximate the kinetic energy functional $E_{\text{kin}}[\rho]$. This innovative approach has paved the way for performing DFT calculations with sufficient accuracy for practical applications [23].

Hence, the majority of modern DFT implementations opt for the Kohn–Sham scheme [23]. We will explore this scheme in greater detail below. As for the direct variational approach, known as orbital–free DFT, which is less accurate compared to the Kohn–Sham approach but offers the advantage of faster computations, ongoing research endeavours are directed towards developing accurate kinetic energy functionals [23]. One strategy involves aiming to replicate the Kohn–Sham kinetic energy as precisely as possible [23].

The Kohn–Sham scheme introduces an auxiliary non–interacting system tailored to produce the same electron density as that of the interacting system [23]. This non–interacting system is characterized by the single–particle Schrödinger equation, commonly known as the Kohn–Sham equation:

$$\left[-\frac{\hbar^2}{2m}\frac{\partial^2}{\partial {r_i}^2} + v_{ef}(r)\right] \phi_i(r) = \varepsilon_i \phi_i(r) \qquad (11)$$

where $v_{eff}(r)$ is an effective potential, $\phi_i(r)$ is the Kohn–Sham state, and $\varepsilon_i$ is the Kohn–Sham energy eigenvalue. The electron density is given by:

$$\rho(r) = \sum_{i=1}|\phi_i(r)|^2 \qquad (12)$$

Following the acquisition of results from DFT calculations, post–processing analysis techniques are employed to interpret and analyze the data. This may encompass visualizing the charge difference distribution, plotting the band structure, calculating surface energies, or examining the electronic density of states [26,27].

## V. Experimental Validation and Incorporation of Dynamics Effects

Incorporating dynamic effects into thin film modelling entails considering the time–dependent behaviour and evolutions of magnetic properties and other pertinent phenomena within the films. These effects are integrated by conducting time–dependent simulations, where the thin film's behaviour is examined over various time intervals. This process entails solving relevant equations of motion or governing equations utilizing methods that account for time evolution. For instance, in magnetic thin films, the Landau–Lifshitz–Gilbert equation can be solved to capture dynamic magnetization behaviours, including precession, relaxation, and damping effects.

Dynamic effects can be investigated by subjecting the thin film to external excitations or perturbations, accomplished through the application of time–varying magnetic fields, temperature changes, or mechanical strain. Analysing the thin film's response to these excitations facilitates understanding its dynamic behaviour. Techniques such as linear response theory, Floquet theory, or time–dependent perturbation theory are commonly employed to characterize the system's dynamic response.

Hegedus and Kugler (2005) used quantum mechanical simulation to model the dynamic interaction of selenium molecules, particularly for simulating the growth of amorphous selenium thin films. Despite limitations, it provided the most realistic simulation [26].

Turowski et al. (2015) utilized quantum mechanical simulation to study $Al_2O_3$ thin films, optimizing the scattering process. Their virtual coater concepts showed remarkable performance, suggesting potential for studying various layer materials and coating processes beyond $Al_2O_3$ thin films [27].

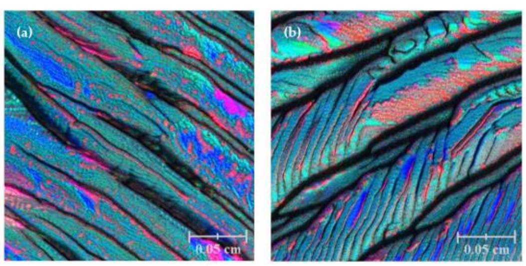



Fig. 4. Surface topology on different locations (a,b) of chromium oxide film [6].

Spin waves and magnons, which are collective excitations in magnetic materials, play a significant role in their dynamic behaviour. Incorporating spin wave or magnon dynamics into thin film models allows for the study of wave propagation, dispersion, and interaction phenomena. Techniques such as spin wave theory, micromagnetic simulations, or dynamic matrix methods can be utilized to integrate these dynamic effects.

Namakian et al. (2021) combined Monte Carlo and molecular dynamic simulations to study Cu thin film growth on TiN substrates. They used sequential molecular dynamic time-stepped force bias Monte Carlo simulations to simulate the deposition process and compared kinetic Monte Carlo simulations, revealing behaviors during the initial growth [28].

Prudnikov (2016) employed Monte Carlo simulations to investigate giant magnetoresistance effects in magnetic multilayer structures. They developed a methodology to determine magnetoresistance using Monte Carlo simulations and studied the temperature dependence of magnetoresistance for different configurations of ferromagnetic films [29].

Candia (2001) used Monte Carlo simulations to investigate the growth of ferroelectric materials, particularly the growth of magnetic films with ferromagnetic interaction between nearest spins [30].

Ultrafast laser techniques and time–resolved measurements have enabled the study of extremely rapid dynamic processes in thin films. Through the use of femtosecond laser pulses and time–resolving detection methods, researchers can explore ultrafast magnetization dynamics, spin dynamics, and relaxation processes in thin films. These techniques provide valuable insights into processes such as magnetization switching, demagnetization, or ultrafast phase transitions.

The integration of dynamic effects into thin film modelling enhances our comprehension of their time–dependent behaviour, transient phenomena, and stability. This approach offers significant insights into processes such as magnetization dynamics, spin wave propagation, ultrafast phenomena, and growth kinetics. Understanding these dynamic effects is paramount for the design and optimization of thin film–based devices, including magnetic memories, spintronic devices, and sensors.

## VI. Scaling up and Integration with Complex Systems

Scaling up complex systems in the modeling of thin films involves extending the study from individual films to multiple interacting films or film–substrate systems. This approach enables researchers to explore collective behavior, emergent properties, and interactions between different film layers or between films and substrates. Scaling up complex systems in the modeling of thin films necessitates the use of advanced techniques and modern methodologies to capture the intricate relationships and phenomena that arise in such systems.

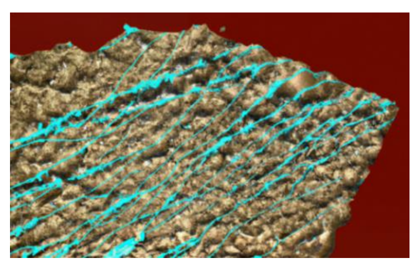


Fig. 5. Simulated surface of a chromium oxide film [6].

Many applications involving thin films require the deposition of multiple layers to create complex structures and functional devices. Modeling multilayer systems entails considering the interactions and interfaces between different film layers. This involves incorporating interlayer coupling, such as exchange interactions or interfacial roughness effects, into the models. Techniques like multilayer simulations, interplay coupling calculations, or effective medium approximation can be employed to study the properties of multilayer thin films.

In a study by Beck and Ederer (2020), DFT and DFPT were employed to investigate interfacial properties in $CaVO_3$ thin films. The impact of the polar $CaVO_3$ interface on physical properties was simulated. Results suggest that comparing experimental and computational outcomes requires careful consideration of corresponding boundary conditions [31].

Kaviani and Aschauer (2022) utilized density functional theory to study oxygen vacancies in $SrMnO_3$ grown on $SrTiO_3$ substrates. Their findings emphasize the significant influence of surface and interface effects on the stability and electronic structure of oxygen vacancies [32].

Thin films are often deposited on substrates, and the interaction between the film and the substrate can significantly influence their properties. Modeling the film–substrate interface involves considering factors such as lattice mismatch, strain effects, surface roughness, and interfacial bonding. Techniques such as finite element modeling, lattice mismatch calculation, or atomic simulation can be used to study the effects of film–substrate interface interactions on the structural, mechanical, and electronic properties of thin films.

In another study, Unal et al. (2007) conducted a density functional theory investigation of the initial bilayer growth of Ag on NiAl thin films. Analysis of supported Ag films on NiAl with an ideal structure revealed that bilayer growth mode is facilitated by a quantum side effect [33].

Atomic modelling techniques, such as molecular dynamics simulation or Monte Carlo simulations, offer detailed insight into atomic–scale phenomena but face limitations in simulating large systems or long–time scales. Bridging the gap between atomistic and continuum scales involves employing atomistic–to–continuum coupling techniques. These methods aim to connect atomic–scale models with continuum models, such as finite element analysis (FEA) or continuum mechanics, to simulate thin film behaviour over large length and time scales. Multiscale modelling approaches integrate models at different length and time scales to capture the hierarchical nature of thin film systems, facilitating the study of phenomena occurring across multiple scales, such as defect formation, phase transition, or interface interaction. Techniques like concurrent coupling, adaptive mesh refinement, or bridging scale techniques ensure consistency and accuracy across different scales.

Integrating first–principles calculations with other modelling techniques provides a comprehensive understanding of the interplay between electronic structure and macroscopic behaviour. This integration involves coupling first–principles calculations with continuum models, embedding density functional theory (DFT) calculations within larger simulations, or using DFT–derived parameters for other models. Validation and refinement of models through comparison with experimental measurements are crucial. Data simulation techniques, statistical analysis, and optimization algorithms help achieve better alignment between models and experimental data. Additionally, experimental data inform model development by providing input parameters or boundary conditions.

Hybrid models combine different modelling techniques within a single framework to capture specific aspects of thin film behaviour. For instance, combining micromagnetic simulation with finite element analysis enables the study of magnetization dynamics in the presence of mechanical stress or external fields. Hybrid models can also integrate different physics, such as combining electromagnetic simulation with thermal modelling to study the thermal response of thin films. These methods often involve coupling or exchanging data between different simulations or models.

Investigating multiple modelling techniques requires careful consideration of interfaces, data transfer, and parameter calibration. Techniques like data coupling, model calibration, or sensitivity analysis ensure the consistency and accuracy of the integrated modelling framework. Integrating multiple modelling techniques in thin film modelling enhances predictive capabilities, broadens the scope of integration, and provides a more comprehensive understanding of film behaviour and properties. It allows researchers to study phenomena at different length and time scales, capturing complex structures and making more accurate predictions for material design, device optimization, and technological advancement.

## Acknowledgment

This work was supported by the Internal Grant Agency of Brno University of Technology, grant No. FEKT–S–23–8228.

## References

[1] J. L. Bretonnet, “Basics of the density functional theory,” *AIMS Mater. Sci.*, vol. 4, no. 6, pp. 1372–1405, 2017, doi: 10.3934/MATERSCI.2017.6.1372.

[2] T. Zhou, P. Zheng, S. C. Pandey, R. Sundararaman, and D. Gall, “The electrical resistivity of rough thin films: A model based on electron reflection at discrete step edges,” *J. Appl. Phys.*, vol. 123, no. 15, pp. 37–39, 2018, doi: 10.1063/1.5020577.

[3] N. Sitapure and J. S. Il Kwon, “Neural network-based model predictive control for thin-film chemical deposition of quantum dots using data from a multiscale simulation,” *Chem. Eng. Res. Des.*, vol. 183, pp. 595–607, 2022, doi: 10.1016/j.cherd.2022.05.041.

[4] A. Fedotov *et al.*, “Theoretical basis of quantum-mechanical modeling of functional nanostructures,” *Symmetry (Basel).*, vol. 13, no. 5, pp. 1–21, 2021, doi: 10.3390/sym13050883.

[5] C. Altafini and F. Ticozzi, “Modeling and control of quantum systems: An introduction,” *IEEE Trans. Automat. Contr.*, vol. 57, no. 8, pp. 1898–1917, 2012, doi: 10.1109/TAC.2012.2195830.

[6] D. Misiurev, “Modeling of Magnetic Films : A Scientific Perspective,” 2024.

[7] Q. Cui, “Perspective: Quantum mechanical methods in biochemistry and biophysics,” *J. Chem. Phys.*, vol. 145, no. 14, 2016, doi: 10.1063/1.4964410.

[8] O. Acevedo and W. L. Jorgensen, “Quantum and molecular mechanical Monte Carlo techniques for modeling condensed-phase reactions,” *Wiley Interdiscip. Rev. Comput. Mol. Sci.*, vol. 4, no. 5, pp. 422–435, 2014, doi: 10.1002/wcms.1180.

[9] W. Peng, Y. Hu, D. Li, and Q. He, “Full-Scale Simulation of the Fluid–Particle Interaction Under Magnetic Field Based on IIM–IBM–LBM Coupling Method,” *Front. Mater.*, vol. 9, no. August, pp. 1–11, 2022, doi: 10.3389/fmats.2022.932854.

[10] M. Turowski, M. Jupé, H. Ehlers, T. Melzig, A. Pflug, and D. Ristau, “Simulation in thin film technology,” *Opt. Syst. Des. 2015 Adv. Opt. Thin Film. V*, vol. 9627, p. 962707, 2015, doi: 10.1117/12.2191693.

[11] M. A. Hassan, M. B. Elfiky, Y. Nukman, and R. Mahmoodian, “Monte Carlo Simulation Model for Magnetron Sputtering Deposition,” *Adv. Mater. Res.*, vol. 1105, no. May, pp. 69–73, 2015, doi: 10.4028/www.scientific.net/amr.1105.69.

[12] V. V. Prudnikov, P. V. Prudnikov, and D. E. Romanovskiy, “Monte Carlo calculations of the magnetoresistance in magnetic multilayer structures with giant magnetoresistance effects,” *J. Phys. D. Appl.*

*Phys.*, vol. 49, no. 23, 2016, doi: 10.1088/0022-3727/49/23/235002.
[13] R. Goyal, S. Lamba, and S. Annapoorni, "Modelling of strain induced magnetic anisotropy in Au additive FePt thin films," *Prog. Nat. Sci. Mater. Int.*, vol. 29, no. 5, pp. 517–524, 2019, doi: 10.1016/j.pnsc.2019.09.001.
[14] K. A. Fichthorn and W. H. Weinberg, "Theoretical foundations of dynamical Monte Carlo simulations," *J. Chem. Phys.*, vol. 95, no. 2, pp. 1090–1096, 1991, doi: 10.1063/1.461138.
[15] J. A. Purton, A. M. Elena, and G. Teobaldi, "Kinetic Monte Carlo modeling of oxide thin film growth," *J. Chem. Phys.*, vol. 156, no. 21, 2022, doi: 10.1063/5.0089043.
[16] N. Bäckström, J. Löfgren, and V. Rydé, "Study of Magnetic Nanostructures using Micromagnetic Simulations and Monte Carlo Methods," 2014.
[17] J. D. Agudelo-Giraldo, E. Restrepo-Parra, and J. Restrepo, "Monte Carlo simulation of roughness effect on magnetic and magnetotransport behavior of La2/3Ca1/3MnO3/La 1/3Ca2/3MnO3 bilayers," *Phys. B Condens. Matter*, vol. 434, no. 1, pp. 149–154, 2014, doi: 10.1016/j.physb.2013.11.014.
[18] T. Sakhraoui and M. Said, "Effect of the interfacial O and Mg vacancies on electronic structure and transport properties of the FeRh/MgO/FeRh (0 0 1) magnetic tunnel junction: DFT calculations," *J. Magn. Magn. Mater.*, vol. 444, pp. 394–400, 2017, doi: 10.1016/j.jmmm.2017.08.060.
[19] M. Tas, E. Şaşıoğlu, C. Friedrich, and I. Galanakis, "A first-principles DFT+GW study of spin-filter and spin-gapless semiconducting Heusler compounds," *J. Magn. Magn. Mater.*, vol. 441, pp. 333–338, 2017, doi: 10.1016/j.jmmm.2017.05.062.
[20] T. Toffoli, "Programmable matter methods," *Futur. Gener. Comput. Syst.*, vol. 16, no. 2, pp. 187–201, 1999, doi: 10.1016/S0167-739X(99)00046-1.
[21] T. Van Mourik, M. Bühl, and M. P. Gaigeot, "Density functional theory across chemistry, physics and biology," *Philos. Trans. R. Soc. A Math. Phys. Eng. Sci.*, vol. 372, no. 2011, 2014, doi: 10.1098/rsta.2012.0488.
[22] H. Jiang and H. Y. Sun, "Density-functional theory," *Quantum Chem. Age Mach. Learn.*, pp. 27–65, 2022, doi: 10.1016/B978-0-323-90049-2.00002-0.
[23] M. Orio, D. A. Pantazis, and F. Neese, "Density functional theory," *Photosynth. Res.*, vol. 102, no. 2, pp. 443–453, 2009, doi: 10.1007/s11120-009-9404-8.
[24] D. Misiurev, P. Kaspar, D. Sobola, N. Papež, S. H. Fawaeer, and V. Holcman, "Exploring the Piezoelectric Properties of Bismuth Ferrite Thin Films Using Piezoelectric Force Microscopy: A Case Study," *Materials (Basel).*, vol. 16, no. 8, 2023, doi: 10.3390/ma16083203.
[25] D. Misiurev, P. Kaspar, and V. Holcman, "Brief Theoretical Overview of Bi-Fe-O Based Thin Films," *Materials (Basel).*, vol. 15, no. 24, 2022, doi: 10.3390/ma15248719.
[26] J. Hegedüs and S. Kugler, "Growth of amorphous selenium thin films: Classical versus quantum mechanical molecular dynamics simulation," *J. Phys. Condens. Matter*, vol. 17, no. 41, pp. 6459–6468, 2005, doi: 10.1088/0953-8984/17/41/016.
[27] M. Turowski, M. Jupé, T. Melzig, A. Pflug, and D. Ristau, " Multiple scale modeling of Al 2 O 3 thin film growth in an ion beam sputtering process ," *Opt. Syst. Des. 2015 Adv. Opt. Thin Film. V*, vol. 9627, p. 96271M, 2015, doi: 10.1117/12.2191049.
[28] R. Namakian, B. R. Novak, X. Zhang, W. J. Meng, and D. Moldovan, "A combined molecular dynamics/Monte Carlo simulation of Cu thin film growth on TiN substrates: Illustration of growth mechanisms and comparison with experiments," *Appl. Surf. Sci.*, vol. 570, no. August, p. 151013, 2021, doi: 10.1016/j.apsusc.2021.151013.
[29] V. V. Prudnikov, P. V. Prudnikov, and D. E. Romanovskiy, "Monte Carlo simulation of magnetic multilayered structures with giant magnetoresistance effects," *J. Phys. Conf. Ser.*, vol. 681, no. 1, 2016, doi: 10.1088/1742-6596/681/1/012016.
[30] J. Candia and E. V. Albano, "Monte Carlo simulation of the irreversible growth of magnetic thin films," *J. Appl. Phys.*, vol. 90, no. 10, pp. 5395–5401, 2001, doi: 10.1063/1.1412841.
[31] S. Beck and C. Ederer, "Tailoring interfacial properties in CaVO3 thin films and heterostructures with SrTiO3 and LaAlO3: A DFT+DMFT study," *Phys. Rev. Mater.*, vol. 4, no. 12, pp. 1–13, 2020, doi: 10.1103/PhysRevMaterials.4.125002.
[32] M. Kaviani and U. Aschauer, "Surface and interface effects in oxygen-deficient SrMnO3thin films grown on SrTiO3," *Phys. Chem. Chem. Phys.*, vol. 24, no. 6, pp. 3951–3957, 2022, doi: 10.1039/d1cp04998d.
[33] B. Unal *et al.*, "Scanning tunneling microscopy and density functional theory study of initial bilayer growth of Ag films on NiAl(110)," *Phys. Rev. B - Condens. Matter Mater. Phys.*, vol. 76, no. 19, 2007, doi: 10.1103/PhysRevB.76.195410.

# A Humanoid Robot on the Basis of Modules Controlled Through a Serial Half-duplex UART Bus


Sebastian Matoušek
*Gymnázium Matyáše Lercha*
Brno, Czech Republic
seb.matousek@gmail.com

Supervised by prof. Ing. Roman Maršálek, Ph.D.
*Department of Radio Electronics*
*Brno University of Technology*
Brno, Czech Republic
marsaler@vut.cz



***Abstract*—Just a few decades ago the concept of humanoid robots assisting us at work would be something better suited for a movie plot than as a business plan. However, nowadays tests of such robots are already being performed, with big companies exploring the possibilities of such robots assisting in warehouses, robotic assistants roaming around airports directing lost passengers and more. In the future, such robots could very well be employed in many more areas of work – thanks to their construction being similar to humans, they integrate well into our everyday environment and make cooperation more intuitive. This thesis is the result of an attempt at building a small-scale humanoid robot. It expands upon the thesis *"Bipedal Walking Robot"* [1] from last year, which described the process of building a two-legged robot capable of walking, at that time still lacking arms and a head, however. This thesis focuses on the expansion of the robot, mainly the addition of the upper limbs and a head missing in the previous project, thus finally creating a humanoid robot. Furthermore, a custom ecosystem of modules on the basis of the ESP32-C3FH4 microcontroller is discussed; communicating through a serial half-duplex UART bus, these modules were used to construct the upper body, smoothen the movement of the robot, and collect additional data about the robots movement. Such modules can be further used in other projects requiring the usage of numerous interconnected, yet independent, electronics parts.**



***Index Terms*—Arnold Biped Robot, humanoid robot, bipedal walking robot, intelligent servomotor, ESP32, ecosystem of modules**


## I. Introduction

The servomotors on the two legged robot described in last years thesis were store-bought could not be reprogrammed – this meant that no functions could be added programatically, or even slightly edited. Due to the very abrupt start and end of the motion of these servomotors together with the inability to reprogram this movement and the necessity of smaller servomotors for building the upper limbs of the robot, it was decided that custom electronics for the robots servomotors would be made. Throughout, this remake of the electronics developed into the design of a whole ecosystem consisting of different modules, all with shared basics – the same microcontroller, all controlled through half-duplex UART, and even with parts of the program being the same. Going forward, the robot evolved from a single central computer controlling individual motors to more of an interconnected network with individual modules all working together to deliver the same result – to walk.

## II. Project Goals

The focus of this project was on two main subjects – expanding the two legged platform made previously to feature a new torso, two upper limbs and a head, to faintly resemble the human body, and creating custom electronics for the robot, mainly for replacing the PCBs in the store-bought servomotors, creating custom servomotors and for creating measuring PCBs which would be placed around different parts of the robot. These PCBs were supposed to function as individual modules, all controllable through serial half-duplex UART communication.

## III. Robot Mechanics

Going from a two-legged robot to a humanoid robot complete with arms and a head meant not only designing these new parts, but also revisiting the design of some of the old parts found on the legs of the robot. Only the most significant changes will be discussed, however.

In any case, prior to the design of the actual robot it was necessary to develop custom servomotors that would be smaller than the store-bought ones, and that would enable a logical design of the upper limbs since the servomotors used in the legs would appear disproportional at the ends of the newly made arms. The design of these servomotors will be looked at first.

### A. Custom Servomotors

Since the arms of the robot could not be constructed solely using the LX225 servomotors for reasons already discussed, custom servomotors were designed for the project. The gears, motor and potentiometer used to determine the axle position were taken from the servomotor MG90. The casing for the servomotors was designed custom and the electronics installed in these small servomotors are the same as the ones newly embedded in the bought servomotors, which will be discussed later. The casing consists of three parts – the middle part houses the motor and electronics, the top part covers the gears and the bottom part simply acts as a cover. The whole servomotor can be seen on fig. 1, and it was named the MG90I.

### B. Head and Upper Limbs

The upper limbs were constructed using two of the bought servomotors that form the shoulder joint together and two of

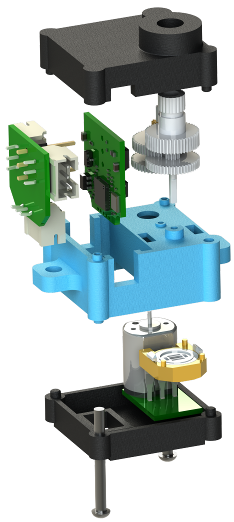


Fig. 1. An exploded view of the MG90I custom servomotor.

the smaller, custom, servomotors, which function as the wrist and a gripper. The head comprises of a camera and a LiDAR sensor. Using one more MG90I servomotor, the head can be turned – all three turning axes that a regular human neck has could not be fit in such a small space, and the turning motion was prioritized.

### C. Changes in the Bottom Limbs

The most significant change in the bottom limbs was the addition of stabilizing flaps to the foot soles of the robot, to try and compensate for the higher center of gravity caused by adding the whole upper body. Some parts were also redesigned to be more rigid, for example the *"shins"* of the robot.

## IV. Electronics

Just as the whole thesis had two main goals, the electronics of the robot also followed two main aims. One was to design a *Hardware Attached on Top* (HAT) PCB for the single board computer, the Raspberry Pi Zero 2W, controlling the robot, that would take care of powering the robot, charging the accumulators, and communication between the Raspberry Pi and all the modules (as in servomotors and measuring units). The second was to design the modules themselves, and also a board to help program them.

### A. RPi HAT Board

Previously, two extension boards were used for the robot – one for power management and one for communication between the modules and the RPi. This time, both of these functions were combined on a single board. The PCB contains a boost regulator from the 3.7 V from the accumulators, which were wired in parallel to make charging easier and increase the time the robot can function for on one charge, to 5 V to power the RPi and another boost regulator to 8.6 V to power the modules.

This second boost regulator uses the TPS-61288 boost converter IC capable of up to 15 A output current, however the input inductor is only dimensioned for about 10 A to save space on the board. A feed forward capacitor was placed between the output and the feedback pin for optimized load transient response [2].

Another circuit onboard the PCB is used for charging the accumulators – this can be done through a USB-C port located on the PCB or wirelessly, by placing the robot on a wireless charging pad. Notably, no more wires are needed to route the charging current to the IC on the PCB – the same cables that are used to power the modules are utilized for this purpose. This can be done thanks to the custom electronics of the modules, which make it possible to have the modules enter sleep mode so that even though they are powered, only a very small amount of the charging current is used by them.

A simple open drain buffer, the 74LVC2G07GM, is used to convert the full-duplex communication to half-duplex [3].

The final PCB is shown on fig. 2.


Fig. 2. The designed and manufactured HAT for power management and communication.

### B. Ecosystem of Modules

All devices in the ecosystem are based on the main parts listed below:

- ESP32-C3FH4 microcontroller
- The H7233-1 LDO voltage regulator
- the same open-drain buffer as on the HAT PCB

The ESP32 microcontroller was chosen mainly for its wireless capabilities, which were used in the IMU module.

The devices can all be chained together to be powered and controlled through one cable. In reality, five such chains are used on the robot – one for each limb and one for the head – since it would simply not be practical to chain all of the modules into one strand. A visual representation of the different types of modules being chained together is on fig. 3.

Since the modules need to be programmed prior to their usage, a separate board dedicated to this task was devised. Thanks to the ESP32-C3FH4 chips capability to be programmed directly by USB without the need to first use a USB to UART chip, the board only contains the USB-C connector, some passive components and a regulator from 5 to 3.3 V.

Let us now talk about the individual modules that were designed.

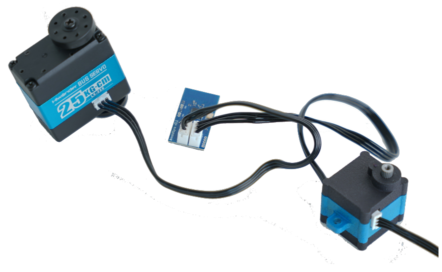


Fig. 3. The three developed module types chained together. LX225 servomotor, IMU module and custom MG90I servomotor going in order from the left.

*1) Servomotors:* The first module that was created was the servomotor control board. The microcontroller reads out the position from the potentiometer and drives the motor using an H-bridge. The H-bridge was implemented using two complementary MOSFETs to form the basic schematic and two external N channel MOSFETs to drive the P channels of the complementary MOSFETs.

A new functionality that was added is the capability to monitor motor current, using the INA181 chip. This current-sense amplifier uses an external reference voltage to be able to measuring current in both directions – using the external offset voltage of 1.65 V (half of the input range of the ESP32), current flowing one way results in an output voltage that is above this level and consequently current in the opposite direction in a lower value [4].

By designing a custom servo controller, all the motions of the servomotor can be customized; this enabled a smoother start and end of the motion, which was the initial reason why this ecosystem was developed, together with creating custom servomotors with these boards. Also, since the program could be customized, a significantly faster baud rate could be set for communication – up to 2,250,000 baud, which is nearly 20x faster than the original 115,200.

*2) IMU Modules:* Utilizing the same basic schematic, a module for measuring acceleration was developed. The LSM6DS3TR-C, an Inertial Measurement Unit (IMU), was mounted on the PCB, to measure the acceleration in all three axes – data that was of interest for future optimization of the walking.

Two inputs for externals sensors were added to this module. They are used for the force sensitive resistors in the foot soles and grippers, and also for determining when to shut down the modules for wireless charging – when the robot is being charged wirelessly, all the modules are powered, since the cables lead through the modules. To minimize the losses of this charging, the IMU module on the left foot of the robot monitors an input in the form of a photoresistor, that is facing an LED, which turns on only when wireless charging is applied. When the module detects a lower-than-normal voltage on the pin monitoring the photoresistor input, it sends a message to all the modules to shut down.

Thanks to the wireless capabilities of the ESP32, an antenna could be added to the PCB and now communication over WiFi and Bluetooth is possible with these modules.

## V. Control Application

The robot used to be controlled through a C++ computer application. However, the added complexity called for a number of functionalities that this application lacked. Therefore, a new app was programmed, this time in the language C# – it once again allows for controlling and monitoring of the robot, with the most notable differences being a window for the camera and LiDAR input and a 3D model that shows the robots current position, thereby helping with the programming of the walking sequences.

## VI. Results

The final robot is almost 45 cm tall (only about 5 cm taller than last year – even though a head was added, measures were taken to reduce excess height in other places, mainly around the *"waist"* area). Consisting of 22 servomotors, five of which are custom made, and weighing in at almost 2 kg, the robot is still capable of walking, and now also of gripping objects.

The ecosystem developed for the robot adds functionalities which were previously unthinkable, such as charging through the power cables of the modules with only low losses and monitoring motor current. The ecosystem is also completely modifiable, a big and important difference compared to the pre-programmed bought servomotors.

## VII. Conclusion

An image of the robot can be seen on fig. 4. The robot is capable of walking and gripping objects and the modules can further be used in other projects in need of numerous independent modules, that can be chained together for power and communications, and that can reprogrammed.

Fig. 4. The finished robot.

# REFERENCES


[1] S. Matoušek, "Bipedal Walking Robot" unpublished, submitted to EE-ICT in 2023.

[2] Texas Instruments, "Optimizing Transient Response of Internally Compensated dc-dc Converters With Feedforward Capacitor", https://www.ti.com/lit/an/slva289b/slva289b.pdf?ts=1701600162194, revised 2017

[3] Analog Devicess, "Using a UART to Implement a 1-Wire Bus Master", https://www.analog.com/en/resources/technical-articles/using-a-uart-to-implement-a-1wire-bus-master.html, 2002

[4] Texas Instruments, "INAx181 Bidirectional, Low- and High-Side Voltage Output, Current-Sense Amplifiers", https://www.ti.com/lit/ds/symlink/ina181.pdf, revised 2023

# Design and Realization of a Smart Home

Petr Bednařík
*26-41-M/01 (3), SPŠEI Mohelnice*
Mohelnice, *Czech republic*
E-mail: pebednarik2005@gmail.com
Supervised by: Ondřej Krajsa
E-mail: krajsao@vut.cz

***Abstract*—This paper focuses on the design and implementation of an open-source smart home, using a central unit based on the Raspberry Pi platform and controlling the Home Assistant operating system. The smart home has smart modules relays, a set of sensors for measuring temperature, humidity and pressure, and safety features for fire prevention. The sensors are implemented using the ESP32 platform and the ESPHOME "add-on". The article also deals with the effective control of heat sources to reduce energy consumption and increase the efficiency of home heating.**

***Keywords*—smart home, home assistant, ESP32, ESPHOME, consumption monitoring**

## I. Introduction

In recent years, modern technology has changed the way we interact with our homes. The term Smart Home is no longer a surprise, but most solutions on the market nowadays have remote servers. Remote servers bring simplicity in the installation of a smart device, but the user loses the independence of his own system. As an example, when the internet connection fails, a large percentage of smart homes become unusable, not to mention the sometimes-questionable ways corporations, especially of cheaper devices, handle user data.

## II. Smart Home Design

One of the most important basic steps is the design itself. Choosing the OS (operating system) for the smart home, selecting the development platform, the "addons" that are going to be used in the implementation of the system.

### *A. Smart Home Hardware*

The choice of hardware for smart homes depends on the application. In the case of the modules that are supposed to communicate with the sensors and the smart home server, it is essential to ensure their energy efficiency, while for the server it is necessary to ensure the stability of the operating system.

### *B. Raspberry Pi 5*

For the proper function of the OS, and therefore the whole household, I have selected a minicomputer Raspberry Pi 5. The reason for my choice was mostly influenced by the existing distribution of the Home Assistant OS directly for the Raspberry platform. The Raspberry Pi unit has enough power to run a smart home for a standard user and all the needs of HA and other add-ons.

The Raspberry Pi 5 is a small single board computer about the size of a credit card. In my design, the computer is equipped with 4GB of LPDDR4X-4267 SDRAM, a 64-bit quad-core 2.4GHz Arm Cortex-A76 processor and VideoCore VII graphics. [1]


Fig. 1. Raspberry Pi 5 mounted in the server rack

### *C. ESP32*

The ESP32 is a popular module often used for smart home applications, especially for interfacing with various sensors and devices in the home. It is a small and energy efficient microcontroller with integrated Wi-Fi and Bluetooth module, making it an ideal choice for various IoT applications.

When selecting the ESP32, I mostly considered what application the modules would be used for. There are many variations of these modules, from very small ones to modules enriched with a camera and an integrated SD card to store the footage.

Fig. 2. Multiple ESP32 variants.

## D. Sensors

The modules are equipped with DHT11, BMP280, DHT22 and DS18B20 sensors for monitoring temperature and pressure in the household, Infrared sensors for fire detection, Magnetic sensors for monitoring the state of opened windows and doors.

## E. Relays

In a smart home, 1-channel relay modules are currently used to switch appliances. However, in the future up to 8 channel relay smart modules will be added to control individual rooms in the household.

# III. SMART HOME OS

When choosing an operating system for a smart home, it was decided to use Home Assistant (hereafter referred to as "HA"). HA is an open-source platform designed for home automation. It provides users with the ability to centrally control and monitor the home. Thanks to HA, the great majority of devices can be connected and controlled, even if this was not possible in the baseline.

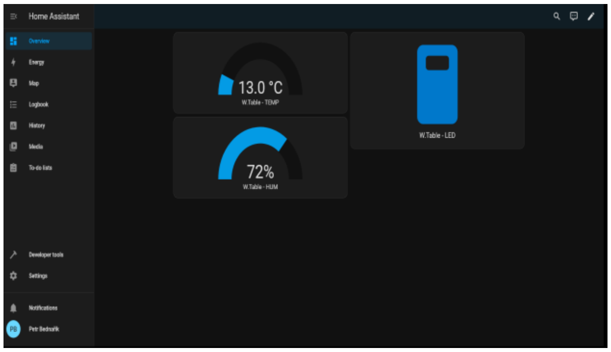



Fig. 3. Interface for smart home control using HA.

## A. ESPHOME

ESPHOME is an open-source project that allows simple control and monitoring of various devices using ESP8266 and ESP32 microcontrollers. This project provides a framework for creating custom firmware for these microcontrollers, allowing easy interfacing with home automation and IoT (Internet of Things) systems. [2]

ESPHOME provides a simple and intuitive configuration language to define the configuration of sensors, switches, relays, and other devices connected to the ESP8266 or ESP32.

```
output:
  - platform: gpio
    pin: 17
    id: lcdbacklight

switch:
  - platform: output
    name: "LCD LED WORKSHOP TABLE"
    output: lcdbacklight
    id: switch_lcdbacklight

sensor:
  - platform: dht
    pin: 15
    temperature:
      name: "Workshop - Table - Temp"
      id: "teplotadilnastul"
    humidity:
      name: "Workshop - Table - Hum"
      id: "vlhkostdilnastul"
    update_interval: 10min

i2c:
  sda: 21
  scl: 22

deep_sleep:
  run_duration: 10s
  sleep_duration: 10min

display:
  - platform: lcd_pcf8574
    dimensions: 16x2
    address: 0x27
    update_interval: 2s
    lambda: |-
      it.printf(0, 0,"Hum   :| %.1f %%", id(vlhkostdilnastul).state);
      it.printf(0, 1,"Temp  :| %.1f'C", id(teplotadilnastul).state);
```

Fig. 4. Example code for programming ESP32 in ESPHOME.

## B. DATA READING

HA allows you to store data from all sensors in a database. With this function, it is possible to compare, for example, temperature changes after and before the insulation of the building.

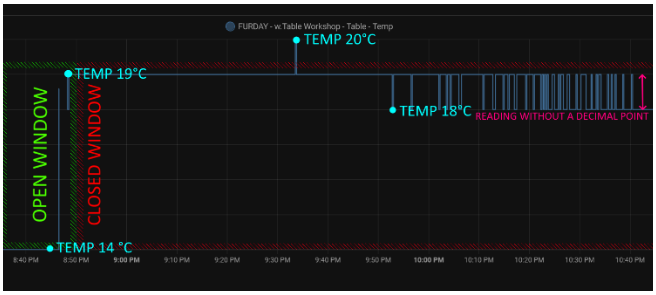


Fig. 5. Graph of the sensor placed in the workshop.

The graph above shows the temperature change in the workshop room where the sensor is installed. When the window was open, the air temperature in the room was around 14 °C, but when the window was closed, the temperature increased and ranged from 18 °C to 20 °C.

The graph also shows the characteristics of the cheaper DHT11 sensor, which can only measure in integer ranges.

## IV. REALISATION

After carefully designing and selecting the necessary hardware and operating system, the actual implementation of the smart home comes next. This phase includes the physical installation of the devices, software configuration, 3D modelling together with 3D printing and finally the interconnection of all components into a functional system.

### *A. Server*

The server and its chassis is made up of a metal housing that has been painted. A build for mounting the Raspberry Pi 5 and active cooling was modelled and then 3D printed.

Fig. 6. Server with mounted raspberry Pi 5.

### *B. Temperature and humadity sensor with integrated LCD display*

The sensor unit consists of a DHT11 temperature sensor, a Lolin ESP32 and a 1602 LCD display. The box was printed in PLA material with concrete texture.


Fig. 7. Sensor unit with display.


Fig. 8. 3D model of the sensor unit.

### *C. Hallway temperature sensor*

The module uses the HTU21D sensor to measure temperature and humidity.


Fig. 9. Hallway sensor unit.

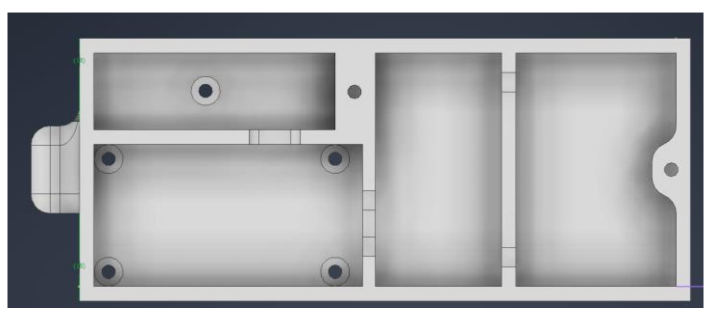


Fig. 10. 3D model of the inner part.

*D. Smart gas boiler controller*

When developing the controls for the gas boiler, I did not wanted to disturb the structure of the boiler itself, for example by replacing the control board of the device. The solution for me was to replace the thermostat itself. I set the original module to be permanently on, thus commanding the boiler control unit to continuously heat the water in the circuit.

The relay is switched when the average temperature in the home drops below the value set in the Home Assistant GUI. When the temperature drops, the relay switches on, thus switching on the original thermostat connected to the boiler. As the original thermostat is set to continuous heating, the boiler will start heating.

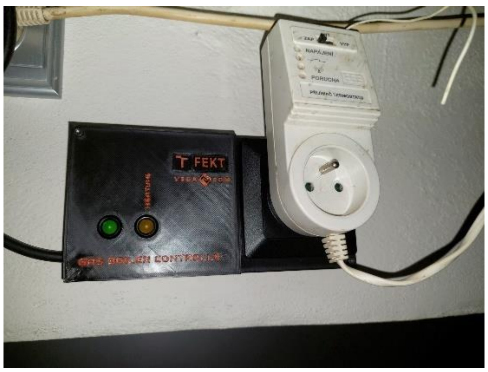



Fig. 11. 3D model of the sensor unit.

## V. CONCLUSION

Nowadays, our devices face various cyber threats and smart homes are no exception. Up to 146 cyber attacks take place every day in the Czech Republic alone, and certainly none of the users of the servers want to lose their private information. A smart home with a local server relies on its own security and, most importantly, processes user data locally. So the user doesn't have to worry about the company (from which he bought, for example, cheap smart LED strips) selling his users' data. [3]

I want to continue with the smart home design and implementation project and enrich the system with my own AI assistant, which could add some new depth to the overall system. I would also like to replace Wi-Fi communication with an Atmel LWM (LightWeight Mesh) system that communicates on the IEEE 802.15.4 protocol.

## VI. SOURCES

[1] Raspberry Pi 5 [online] [cited 2023 Dec 25] Available from: https://rpishop.cz/raspberry-pi-5/6498-raspberry-pi-5-8gb-ram.html#tab-description

[2] ESPHOME [online] [cited 2 January 2024] Available from: https://esphome.io/index.html

[3] Cyber attacks [online] [cited 20 Feb 2024] available from: https://www.statista.com/statistics/1344671/czechia-number-of-cybercrime

# School Meteorological Station


1nd Petr Pauk
*Gymnázium Židlochovice*
Židlochovice, Czech Republic
pauk.p17@zak.gymnzidlo.cz

2st Martin Knotek
*Gymnázium Židlochovice*
Židlochovice, Czech Republic
knotek.m17@zak.gymnzidlo.cz



***Abstract*—Our work deals with meteorological measurements from theoretical as well as practical perspectives. Our goal was to construct an amateur meteorological station with an emphasis on recycling components from previous projects. We created a meteorological station, which is placed in the garden of our school and continuously measures meteorological data and sends them to a server, where it is accessible online, including simple visualization. Our work is also a guide to build similar meteorological stations not only for the participating teams of the project Dotkni se vesmíru and was submitted to the project organizer. All created software is publicly available on GitHub.**



***Index Terms*—Meteorology, Project Dotkni se vesmíru, OpenSTRATOkit, Habdec, Meteorological station, Weather station Meteorology in Židlochovice**


## I. Introduction

In the school year 2021/2022 we participated in the project Dotkni se vesmíru. This involved launching a meteorological probe into the stratosphere. During our participation in the project, we have frequently discussed the fact that using the probe for a single measurement “is a waste”. We had a probe with sensors capable of measuring temperature, atmospheric pressure, relative humidity and environmental radioactivity and decided to add an anemometer and a wind direction sensor and start long-term use to measure meteorological data on the school property.

The aim of our work is therefore to build a meteorological station from available components to measure ambient temperature, atmospheric pressure, relative humidity, radioactivity and wind speed and direction. The station, located in a meteorological box in the school garden, will be sending the measured data to a server for long-term archiving. Here the data will also be made available to the public[1]. This includes a simple online display of the data in the form of graphs.

In order to reprogram the probe, to create a new program to store the data, and to modify the existing program, it was necessary to use programming languages Python, C++ and JavaScript. We encountered the latter for the first time in our lives and had to learn how to use it. We also worked with electronic components and had to learn how to wire them. The work included a summary of information about meteorological measurements mainly from publications by CHMI[2] and an interview with an expert in this field, Ing. Karel Král.


Sdružení přátel Gymnázia Židlochovice, z.s., město Židlochovice


[1] http://knotkovi.tplinkdns.com:3000/

[2] Czech hydrometeorological Institute

The motivation for our work was our passion for programming and working with computer technology. Also we had the opportunity to create a useful thing that others can use and improve.

The Dotkni se vesmíru is a project organized every year by Steam Academy Praha for active and talented students. It allows each of 45 teams of students to launch their own meteorological probe into the stratosphere and take unique photographs of more than just the curvature of the Earth from 33 kilometers high [5]. The students are tasked with launching the probe, which they can modify according to a set of rules. Our school participated in this project in 2021. After that the probe became a property of our school and had no further use. The motherboard of this probe was used to build our meteorological station.

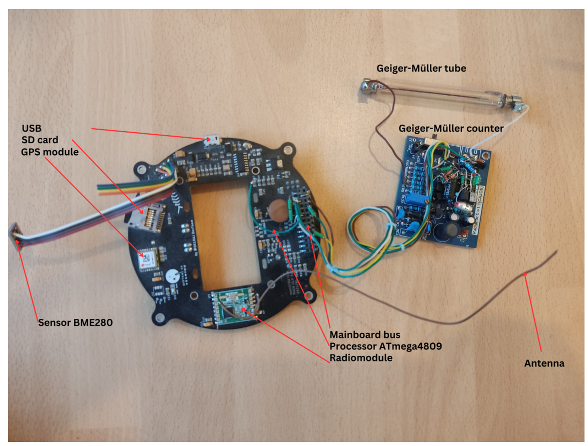



Fig. 1. The mainboard with all sensors from the probe.

## II. The meteorological station

### A. Meteorological station components

Since our goal was not to create a professional meteorological station, but rather to recycle the probe from the Dotkni se vesmíru project, we did not want to replace the integrated sensors with better ones. If we replaced them, our original intention to recycle the probe would be lost. The sensors used in the original probe can be considered sufficiently reliable. Therefore, we have added only an anemometer and a wind direction sensor.

- A very important part of the station is the booth. For our purposes we chose a standard louvered meteorological booth supplied by METEOshop. The manufacturer states that it is made of solid wood, the protective coating is provided by several layers of white paint and ventilation is allowed by a two-level ventilated floor and a double ventilated roof. The box meets almost all the requirements specified by the WMO[3] [2]. It is the best quality meteorological booth available on the market. The booth can of course be made by oneself, but it requires enough time. Instructions for construction can be found on the internet[4].

- The station is equipped with a BME280 sensor that is capable of measuring temperature, pressure and relative humidity. This sensor is very accurate and reliable. It can operate in the temperature range -40–85 °C [1].
- The wind speed is measured by the wind speed sensor WH-SP-WS01. The sensor is technically very simple, so it is expected to be reliable. Another advantage is that no additional power supply is needed [8].
- The wind direction is captured by the wind direction sensor WH-SP-WD. This is also technically very simple. No additional power supply is needed [7].
- The radioactivity of the environment is measured by a Geiger-Müller computer (type RadiationDv1.1(CAJOE)). Our device has a large operating temperature range and is very power efficient. It can capture both $\beta$ and $\gamma$ radiation [6].
- The entire station is powered by a power bank for charging mobile devices. For our purpose we chose an older functional unused piece with a capacity of 16 000 mAh. It needs to be recharged regularly, but not often, as it has a large capacity and the station is very energy efficient (we use a similar backup power supply while recharging). It needs to be recharged approximately twice a week. We will replace it with a solar panel in the future.

## III. Probe Program and Hardware Modifications

Prior to the modification, the original probe measured the ambient temperature and pressure and sent these data together with the data from the GPS[5] module (i.e. current time, latitude, longitude, altitude and velocity of the module) to the receiving station. It also took photographs and measured the radioactivity of the environment, and stored all the data on an SD card[6].

We removed the camera and added an anemometer and a wind direction sensor. We also activated the humidity sensor, which was present in the original probe, but was not used. To make everything work, we rewrote the probe program so that it sends all the measured data (ambient temperature, pressure, relative humidity, wind speed and direction, and environmental radioactivity) to a server. We also had to deactivate the GPS module (more details below). The probe program is available on GitHub[7]. The following block diagram roughly describes how the station and its program work.

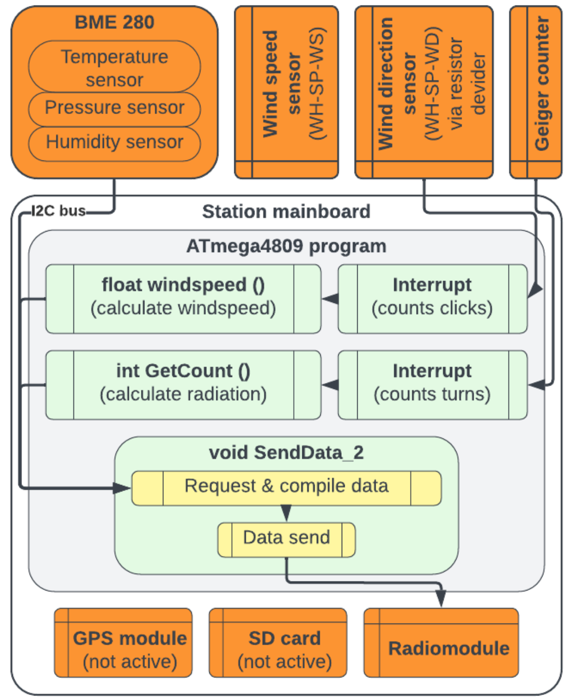



Fig. 2. Meteorological station block diagram

### A. Anemometer

The anemometer sends a signal (by connecting two contacts) every time the sensor head rotates 360 degrees. The station counts these pulses and divides the number of pulses by the length of the time interval over which it counted the pulses. This number is finally multiplied by a coefficient (this is given in the anemometer specifications). The final result is given in $\mathrm{m \cdot s^{-1}}$. To allow the station program to count pulses without interrupting the program, the CPU[8] interrupt was employed.

### B. Wind direction sensor

The wind direction sensor is a very simple component (a resistor that changes its value depending on the rotation), but we had to figure out the wiring for this particular case ourselves. The ATmega processor has an ADC[9]. This acts as a digital voltmeter, not capable of directly measuring the value of a resistor. So we needed a circuit to convert the resistance to voltage. The easiest way was to use a resistor divider, of which the wind direction sensor is a part.

[3]World Meteorological Organization

[4]https://www.globeslovakia.sk/wp-content/uploads/2020/11/meteobudka.pdf

[5]Global Positioning System

[6]Secure Digital card (small compact storage media)

[7]A cloud-based Git repository that helps developers store, manage, track and control their software, available at https://github.com/PaukPetr/Meteo-Station-Programs

[8]Central Processing Unit

[9]Analog to Digital Converter

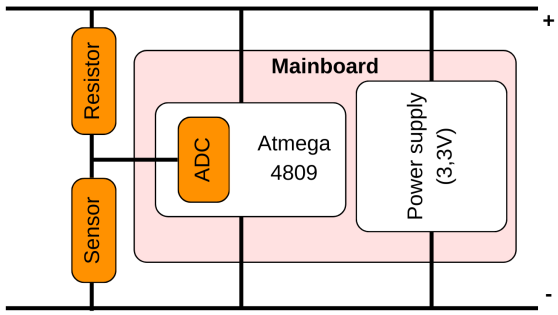



Fig. 3. The wind direction sensor wiring scheme.

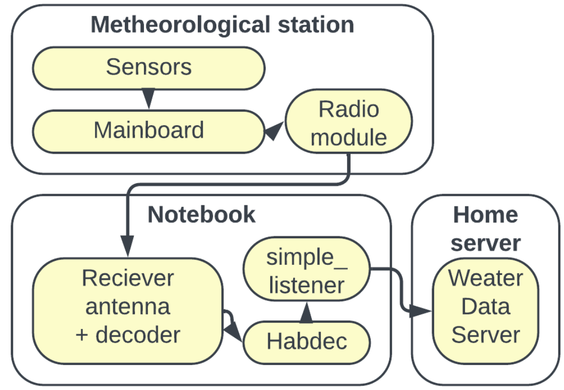



Fig. 4. Communication between station and server for storing data.

### C. Other sensors

- The GPS module (see chapter meteorological station components) of the probe was deactivated because it had a high power consumption and it also caused problems during testing the probe in a building, as the function of most of the program is dependent on access to satellites. So we removed all the parts of the original code that read GPS data and would potentially be blocked by deactivating the module.
- The data measured by the station is sent to the receiving station via a radio module. It transmits them in the form of packets. Their format is text in the form of: "ambient temperature, pressure, relative humidity, wind direction, wind speed, radioactivity of the environment". Here, the program needed to include data on humidity, wind speed and direction, and environmental radioactivity into the packet when it was transmitted.
- The original meteorological probe stored data on an SD card. However, because of the reasons described above, we had to deactivate the GPS module, which is required to obtain the exact time, which is important for orientation when storing the data. Therefore, we omitted saving to the SD card and we rely only on the server to automatically assign the time stamp to the incoming data.

## IV. Communication and Data Archiving

The meteorological station first sends the data via radio module to a laptop (an old unused piece of equipment, dedicated for this purpose), which receives it via an antenna, decodes it using the Habdec[10] program (see chapter Habdec) and sends it via the internet to our server on which the WeatherDataServer is running. The communication scheme is in figure 4.

The WeatherDataServer takes care of storing and processing the data and visualizing it on our web page. It communicates with the receiving program (Habdec) via HTTP[11], for which it uses the Express.js[12] library. A part of the program is a web page[13] with instructions and a simple tutorial for using WeatherDataServer. The program stores the data in a table, and sends the specified portion when requested by a user.

### A. Habdec

Habdec is a decoder designed for decoding data from meteorological probes. This program was used to decode data in the Dotkni se vesmíru project and continues to be used further. However, the program had to be supplemented. Habdec sends the data to the sondehub.org website where it is stored in a database. However, it doesn't send it where we need it to. We have therefore tried changing this address or adding a few more lines to the code that would accomplish this task. However, Habdec is very complex, it would take a lot of time to understand all the code, and it might not work well after intervention. Therefore, we contacted the developers of the program and found out that the Habdec program file downloaded from Github includes a short Python code (simple_listener) that, when run, connects to the Habdec communication port and reads data from it. Since this additional code is much shorter and simpler, it was only necessary to add a function to send data to the WeatherDataServer. The simple_listener itself is only capable of getting the data received by Habdec and outputting it to its console. Sending the data had to be programmed and was handled by the library requests[14], which makes it very easy to send data to a required address. The data must be in JSON[15] format. Simple_listener prints out snippets of the packet to its console continuously and finally, when it has the whole packet, it prints it out at once. These whole listings are easy to filter from the console.

### B. WeatherDataServer

WeatherDataServer is our program for data manipulation and archiving. It takes care of processing them in ten-minute intervals, then saving and sending them either in the format in which the google.charts 16 library accepts them, or the actual data values in JSON format to make them easily viewable on the web page it hosts. The data from the meteorological

[10] High Altitude Balloon decoder
[11] Hypertext Transfer Protocol - protocol for web communication
[12] Minimal and flexible Node.js web application framework
[13] http://knotkovi.tplinkdns.com:3000
[14] A Python library for web communication
[15] JavaScript Object Notation - web data interchange format: ("key: value, key: value")

station is not time-stamped because the GPS module in the station has been deactivated. Therefore, it is complemented by the WeatherDataServer. The WeatherDataServer source code is available on GitHub [16].

### C. Data download and display

The WeatherDataServer provides several ways to download and display data. For less demanding users, data for the last four days is displayed directly on our page in a graph. Another way is to download the data. This can be done by clicking on the highlighted link, or by entering the web address[17]. Finally, the raw form of our data can be programmatically accessed by other websites or servers, thanks to HTTP functions, by providing the actual values measured by our station[18] or by providing data in a user-defined time range[19]. The following figure displays correlation between temperature and humidity on 28.3.2024-29.3.2024. Only temperature and humidity is displayed because the correlation is apparent here.

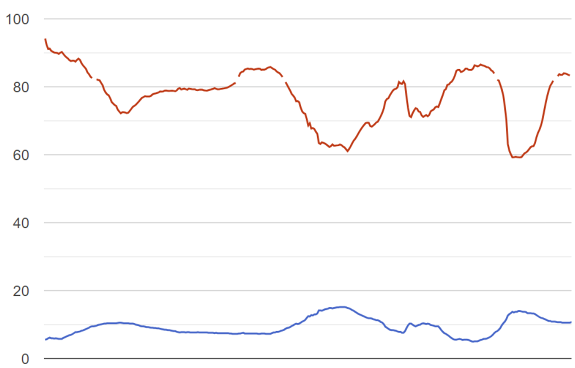



Fig. 5. Correlation between temperature (blue) in °C and relative humidity (orange) in %.

### D. Installation

When installing the station, the principles of meteorological measurements from CHMI played a very important role, so the design of the station had to meet certain parameters. The station had to be placed approximately two meters above the ground, and we had to build a support structure. For this purpose, we used a three-meter-long stainless steel rod, which we embedded 60 centimeters below ground level. The station itself was then attached to the pole using a wooden structure that we glued together from wooden planks. We placed the anemometer and wind direction sensor approximately four meters above the ground, on a wooden pole fitted into the stainless steel pole. Note that none of this is necessary unless more exact measurements are needed (We wanted the best results possible). See the installed station on figure 6.

[16]https://github.com/Mett686/WeatherDataServer

[17]http://knotkovi.tplinkdns.com:3000/data/2024.csv, where the year to be downloaded can be adjusted as needed

[18]See http://knotkovi.tplinkdns.com:3000/data

[19]For example http://knotkovi.tplinkdns.com:3000/array/2024/2025

Fig. 6. Installed station with view into inside.

## Acknowledgment

We would like to thank Ing. Lukáš Pauk, Ph.D, who helped us significantly with the programming and implementation of the project from a technical point of view. Without this help many things would not have been possible. Our thanks also belongs to Mgr. Hana Stravová for her help, encouragement and guidance in the development of this work. Without the time she gave us, we would not have been able to complete the work. We would also like to thank Mr. Ing. Karel Král for important information regarding meteorological measurements in the vicinity of Židlochovice and the town of Židlochovice itself for providing us with funds for the purchase of sensors and the meteorological booth.

## References

[1] BME280 Combined humidity and pressure sensor [online]. itbrainpower.net THE ALPHABET PROJECT - Arduino and Raspberry PI compatible IoT modems, shields and sensors [cit. 2022-11-7]. Available: https://itbrainpower.net/downloadables/BST-BME280-DS002-1509607.pdf

[2] Stevensonova meteorologická budka malá - bez stojanu. METEOshop [online]. [cit. 2023-10-05]. Available: https://www.meteoshop.cz/produkt/stevensonova-meteorologicka-budka-mala-bez-stojanu/797/

[3] Návod pro pozorovatele automatizovaných meteorologických stanic. 2., upr. vyd. Praha: Český hydrometeorologický ústav, 2014. ISBN 978-80-87577-34-9.

[4] Návod pro pozorovatele meteorologických stanic. 2., upr. vyd. Praha: Český hydrometeorologický ústav, 2014. ISBN 978-80-87577-33-2.

[5] Projekt Dotkni se vesmíru. Dotkni se vesmíru [online]. [cit. 2023-01-09]. Available: https://dotknisevesmiru.cz/

[6] Senzor radiace - Geigerův počítač. Laska Kit [online]. [cit. 2022-11-07]. Available: https://www.laskakit.cz/senzor-radiace-geigeruv-pocitac/

[7] WH-SP-WD čidlo směru větru. Laska Kit [online]. [cit. 2022-11-09]. Available: https://www.laskakit.cz/wh-sp-wd-cidlo-smeru-vetru/

[8] WH-SP-WS01 čidlo rychlosti větru - anemometr [online]. [cit. 2022-11-09]. Available: https://www.laskakit.cz/wh-sp-ws01-cidlo-rychlosti-vetru-anemometr/

# Simplifying Plant Care with Plantiful: A Comprehensive Plant Care App

1st Karol Raffay
*SPŠ Brno, Purkyňova*
Brno, Czech Republic
k.raffay@protonmail.com

2nd Robin Jarůšek
*SPŠ Brno, Purkyňova*
Brno, Czech Republic
robinjarusek@gmail.com

3rd Valon Mavriqi
*SPŠ Brno, Purkyňova*
Brno, Czech Republic
valon.m@seznam.cz

4th Tomáš Caha
*Department of Telecommunications*
*Brno University of Technology*
Brno, Czech Republic
tomas.caha1@vut.cz

***Abstract*—Caring for plants is not easy for many people. It requires knowledge, it takes time and each plant has different requirements. Platinful is a mobile app developed by three high school students designed to simplify plant care. The app is connected to an external plant recognition API, Plant.id, which provides the app with comprehensive plant care information, watering intervals, and general recommendations for indoor and outdoor plants. This article discusses the development of the Plantiful app, its features, and its potential impact on conservation efforts.**

***Index Terms*—plant care, mobile application, plant recognition, watering reminder, plant health**

## I. Introduction

Plants are an integral part of our environment, but not everyone knows how to look after them properly. For one thing, a layperson may not recognise a plant in order to know how to look after it properly. Secondly, people are pressed for time and can simply forget about watering plants.

Our goal was to create a mobile app that users can keep close at hand at all times, making it easier and more accessible to care for plants so that fewer plants die and more are planted. In the app, we provide users with all the information they need to care for their plants - recommended room temperature, sun exposure, watering intervals and the amount of water a given plant needs, all by connecting to the external API called Plant.id.

In the following sections, we will go over our technical solution where we describe the developed mobile app and connections to external services including the technology behind it and our vision for the future in this paper.

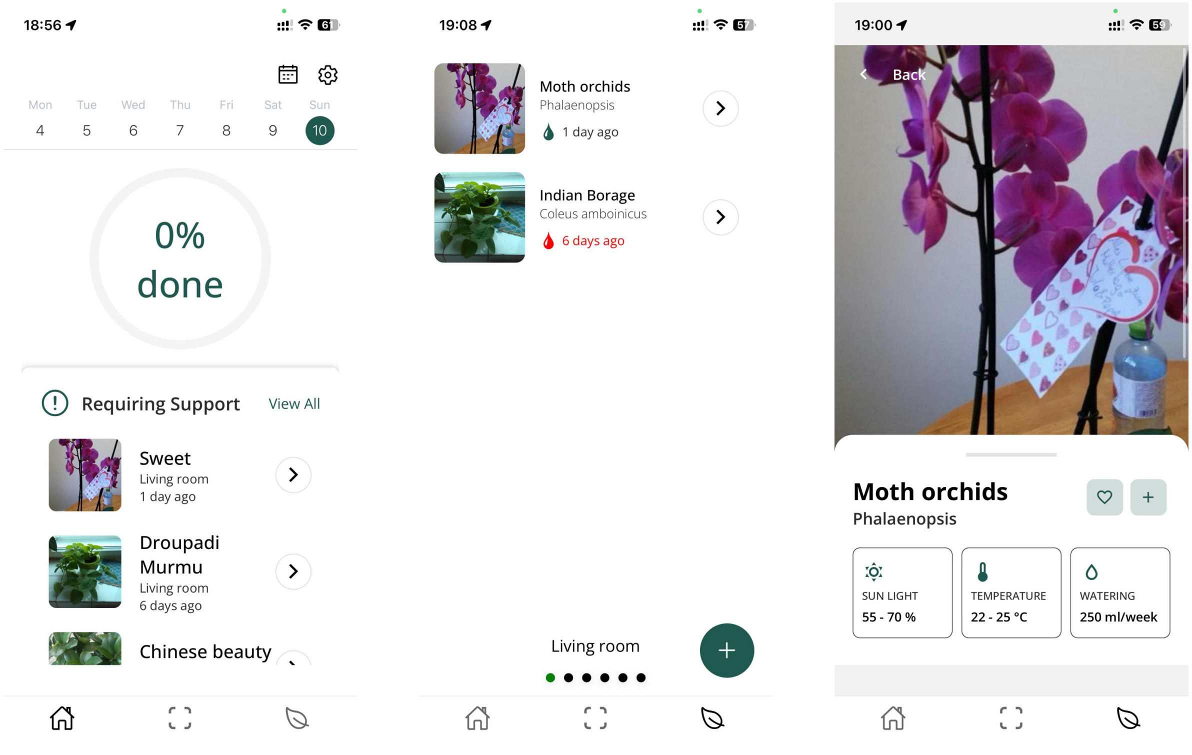
## II. Description of developed mobile app

Plant recognition is an essential part of our application. The process of recognising a plant is done through an AI (artificial intelligence) model that is trained on a large amount of specific data (images of plants from different angles, in different lighting conditions, in different climates, with varying health and the corresponding plant's name). To create a model like this ourselves was an impossible solution for us – three high school students; simply gathering all of the necessary images would take ages [1]. So we decided to use Plant.id's API (application programming interface) [2] with their own respective plant database and AI model. Our reasons were their competitive pricing, experience in the industry (launched in 2018), and the fact that they are also a company from Brno.

We have created the following list of key features of our application:

- **Plant and Disease Recognition:** Utilising AI technology, Plantiful can identify a wide range of plants and their potential diseases, providing users with immediate care solutions. See Figure 1a.
- **Personalised Watering Reminders:** Based on the type of plant and its specific needs, the app sends timely watering reminders to users, ensuring optimal hydration levels.
- **Plant Information:** A large database offers valuable information about your plants, including sunlight requirements, soil types, and fertilisation schedules.
- **Categorization by Room:** Users can organise their plants based on their location within their home, allowing for tailored care recommendations that consider the unique environmental conditions of each room. See Figure 1b.
- **Cross-platform:** Our app is designed to make the experience seamless whether you are using an Android or iOS device without sacrificing performance or user experience.
- **Beginner-Friendly:** Designed with simplicity in mind, Plantiful offers an intuitive and easy-to-navigate interface. With step-by-step guides and helpful tips, users can learn at their own pace and gain the confidence needed to care for their plants successfully. See Figure 1c.

We wanted to have a unified code base and ultimately decided to use React Native, which allows us to develop native apps for both iOS and Android without sacrificing performance or code complexity. The downsides are definitely dependencies on third-party libraries and complexity in navigation [3], which is currently split between four different files due to the approach react-navigation uses [4]. Then we had to figure out how to store all of the user data, their plants and rooms, and authorise users. We found a clear favourite, making use of Firebase [5], which is a set of backend cloud computing services and application development platforms provided by Google. It hosts databases, services, authentication, and integration for a variety of applications, including Android, iOS, JavaScript, and many more, which is perfect for our use case.




(a) Home screen. (b) List of plants. (c) Detail of selected plant.

Fig. 1: Screens of Plantiful app

## III. Use Case Example: From Recognition to Care

The objective of this use case example is to demonstrate the application of Plantiful in identifying an unknown plant and setting up a customised care schedule to ensure its optimal health.

The primary actor is a regular user with limited plant knowledge, referred to as "the user" for this scenario. The user has an unidentified indoor or outdoor plant requiring regular care. The whole process can be seen in Fig. 2. When the user wants to identify the plant, he opens the Plantiful application and selects the "Scan Screen". With the device's camera, the user takes a clear image of the plant in question. The application bundles the image along with other necessary data, such as the base64-encoded image, its dimensions, and whether the API should respond with similar images of the scanned plant, into a JSON payload. The payload is sent to the Plant.id API, awaiting a request. The user sees a loading spinner until we get a response from the API (approx. 3 seconds). The response from the Plant.id API contains data for the backend, such as "is_plant", signifying if Plant.id recognized the image as a plant or the probability of a correct prediction, although most of the data, such as the plant's name, description, watering intervals, or if there is a need for fertilising, is for the user himself.

After successful identification, Plantiful prompts the user to select the room or general area the plant is located in or to create a new one. The user then adds the plant to one of the two options mentioned and gets taken to the "Plant's Detail Screen", which details important information such as optimal sunlight exposure or watering frequency. The user can later find all of his plants in the "Plants Screen". Plantiful prompts the user to activate personalised reminders, such as push notifications or general reminders within the application. These reminders are specific to the plant's needs.

## IV. Future Work

Future work should focus on integrating a health assessment feature to match results of other competitors, full use of Plant.id's data and infrastructure, extensive user testing to ensure the right user experience, and refinement of the app design. Images and plant recognition (meta)data collected from users will be stored securely and anonymously to build our own AI model for plant recognition to reduce dependency on third parties and optimize the app's operational cost per API query.

## V. Conclusion

Plantiful can accurately recognize different plants based on data obtained from the external Plant.id API and our own user data. Users can create a list of their own plants, which is stored in cloud storage to their account, and get the information they need to care for their plants, receive watering reminders through our app. The development process was described in Chapter 2, the whole application is built on the React Native framework, which allows us to develop cross-platform for iOS

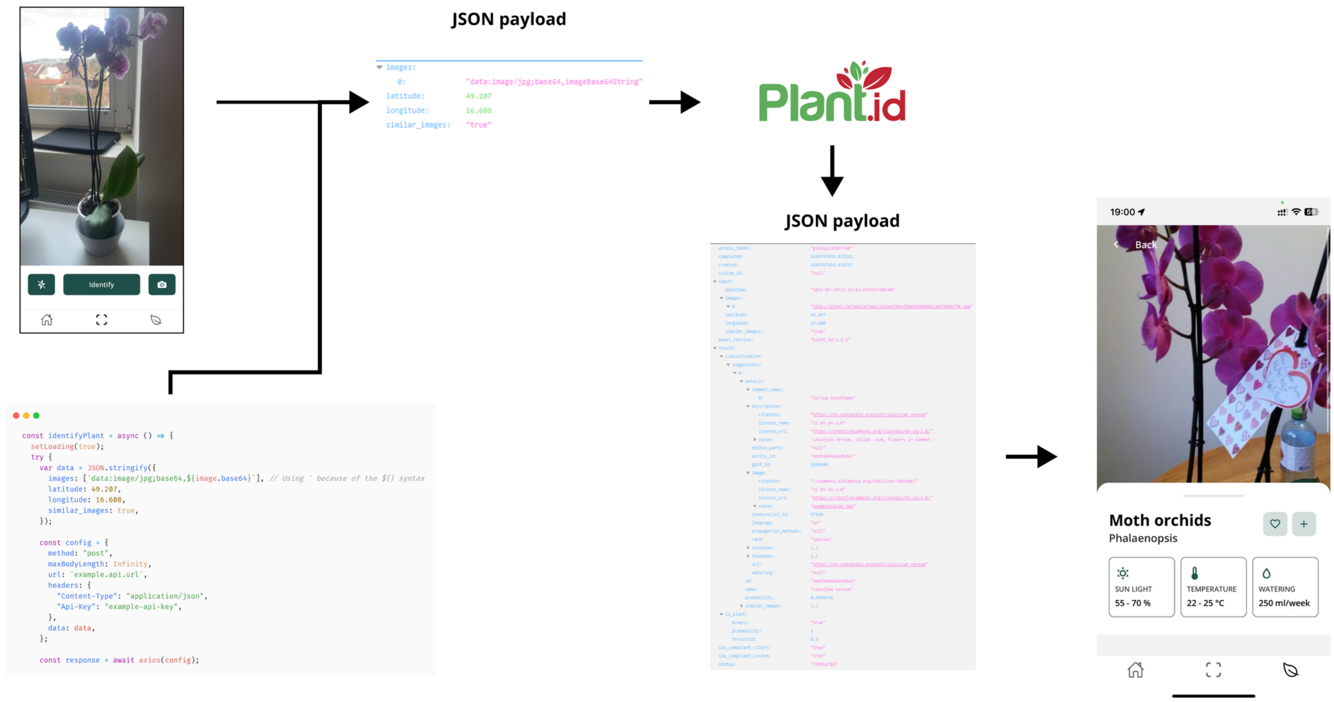



Fig. 2: Use case: Plant identification using Plantiful app

and Android operating systems. Next, the basic idea and its transformation into its current state was described. In Chapter 3, the process of taking a picture of a plant, recognizing it, processing the response from the API and using the data in our application was demonstrated. In the future, it would be good to use more information from the API and focus on reducing the amount of API requests.

Development of this app started in 2022 as part of the Red Hat summer camp competition and after that it secured a year-long internship at Kyndryl. Lastly, we participated in the JA EXPO 2023 competition, winning the Best Business Pitch and Best STEM Project categories.

## ACKNOWLEDGMENT

We would like to thank Kindwise, who develops and operates the Plant.id API, without whose data this application would not work. They kindly reached out to us and offered to increase our API usage credits to make our development work smoother.

## REFERENCES

[1] REYES, Angie K.; CAICEDO, Juan C.; CAMARGO, Jorge E. *Fine-tuning Deep Convolutional Networks for Plant Recognition*. CLEF (Working Notes), 2015, 1391: 467-475.

[2] *Plant.id - Plant identification app*. Accessed March 10, 2024. [Online]. Available: https://plant.id/

[3] *Navigating Between Screens · React Native*. Accessed March 10, 2024. [Online]. Available: https://reactnative.dev/docs/navigation

[4] *React Navigation — React Navigation*. Accessed March 10, 2024. [Online]. Available: https://reactnavigation.org

[5] *Firebase — Google's Mobile and Web App Development Platform*. Accessed March 10, 2024. [Online]. Available: https://firebase.google.com

# Design of model rocket with integrated flight control system


1st Ondřej Kadlec
*Gymnázium Moravský Krumlov, příspěvková organizace*
Moravský Krumlov, Czech Republic
ondrej.kadlec@post.cz

2nd Petr Walla
*Gymnázium Moravský Krumlov, příspěvková organizace*
Moravský Krumlov, Czech Republic
walla.petr@mkgym.cz



***Abstract*—This article is describing design and construction of rocket model with solid propellant engine, which communicates in real-time with server. Purpose of this rocket is to engage in vertical flight and afterwards safely return back to the ground, while sending data and recieving orders. In scope of this thesis we developed desktop application versatile in use for controlling flights of model rockets.**

***Index Terms*—rocket, solid propellant rocket motor, rocket recovery, flight software, ground control**


## I. Introduction

This article is partly based on our theoretical and practical knowledge gained from participating in two years of the Czech Rocket Challenge competition, where we came to the conclusion that the basic prerequisite for a development of advanced features on model rockets is increase in reliability and quality of telemetry data collection and its subsequent processing, as well as real time tracking information about the condition of the rocket in real time.

Our goal was to design and build a single-stage rocket with the following properties:

- simple construction with usage of accessible and cheap parts,
- modularity, flexibility and versatility of components,
- reliable recovery system,
- reusable solid propellant motor,
- real-time communication with the server, collection, transmission and recording of telemetry data, status information and visual data.

The other main goal was to develop a universal desktop application that displays the acquired data in real time to the user.

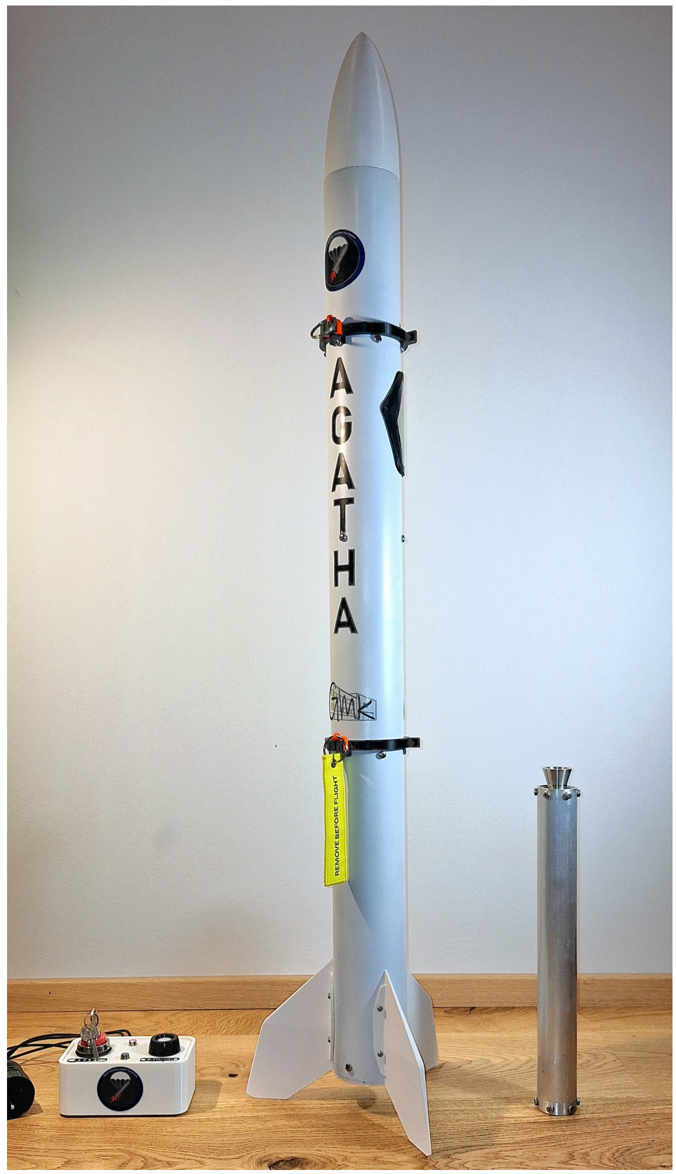



Fig. 1. Agatha rocket (center), K-I450 rocket motor (right) and launch control panel (right)

## II. Rocket Functions

As we outlined in the introduction, the function of our rocket is to perform an uncontrolled stable vertical flight and post apogee releasement of the recovery system (parachute), on which it safely descends onto the ground. Throughout the flight, including descent, the following data is collected:

- barometrically detected altitude,
- atmospheric pressure,
- temperature,
- yaw, pitch and roll
- forward-facing acceleration,
- coordinates obtained by GNSS.

This data is sent to the server and also saved inside the rocket. The camera tilted at a 45° angle to the rocket axis records video at resolution of 720p and a frame rate of 60 frames per second, which is stored on an SD card. Along with other data, it sends images from the camera of the same resolution. The reason behind not transmiting entire video is the low computational power determined by the less powerful onboard computer, although scaling the system for higher performance is valuable option. The data package also contains information about the condition of the rocket.

All functions are controllable via the desktop application. In case of signal loss, the rocket is capable of independent operation.

## III. Rocket Parameters

Here we present a table of basic parameters

TABLE I
BASIC PARAMETERS OF THE ROCKET

| | |
|---|---|
| Weight at start | 1950 g |
| Weight at descent | 1550 g |
| Weight without motor | 1100 g |
| Length (without stabilizers) | 1050 mm |
| Body diameter | 75 mm |
| Maximum diameter | 250 mm |
| Number of stabilizers | 3 |
| Stability (at start) | 1.8 cal |
| Drag coefficient | 0.32 |

## IV. Structure

The external aerodynamics are designed so that the center of pressure (the common action point of the drag force of all rocket components) is located below the center of gravity. This ensures high stability flight, because with an increasing deviation of the rocket axis relative to its movement direction, the normal component of the drag force increases. [6]

The body of the rocket was created by using part of a PP waste pipe, while most of other parts is 3D printed.

Internal components are attached using system of bulkheads, avionics are located in removable plug in module. The rail buttons are attached to clamps that fall of after the rocket leaves the launching pad.

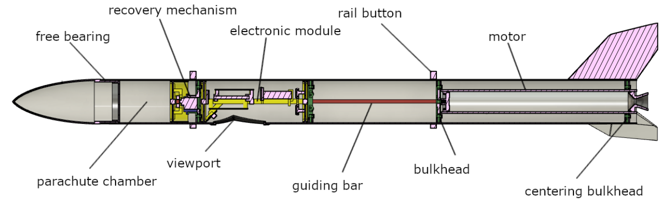



Fig. 2. Layout of components in Fusion 360

## V. Motor

The rocket is powered by a solid propellant motor, where the propellant is stored directly in the elongated combustion chamber. Body of the motor is made of aluminium, remaining parts of are made of steel. For simulation of the fuel burning, we used opensource software openRocket. [2]

TABLE II
K-I450 MOTOR PARAMETERS

| | |
|---|---|
| Class | I |
| Total impulse | 450 Ns |
| Running time | 1.4 s |
| Average thrust | 380 N |
| Average pressure | 3.8 MPa |
| Propellant weight | 440 g |
| Outer diameter | 40 mm |
| Diameter of the propellant core | 34 mm |
| Length of the propellant core | 300 mm |
| Initial diameter of the combustion channel | 12 mm |
| Profile of the combustion channel | circular |
| Fuel | KNSB |

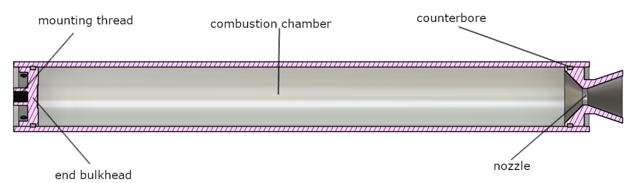



Fig. 3. Motor design in Fusion 360

KNSB is a compound of potassium nitrate and sorbitol; it is one of the simplest and safest rocket propellants, commonly used by amateurs. [4] The motor is ignited electrically using launch control panel.

## VI. Recovery System

### A. Recovery System Deployment Mechanism

We wanted to aviod usage of explosives, therefore is the mechanism based on springs with total stifness of 0.25 N·m$^{-1}$.

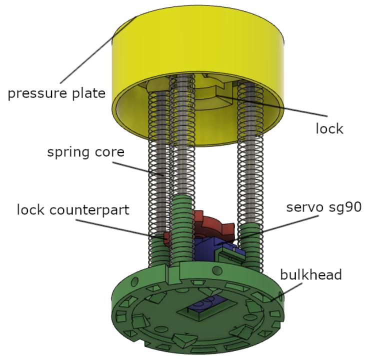



Fig. 4. Design of recovery system deployment machanism in Fusion 360

### B. Parachute

We use a so-called two-stage parachute. It consists of a smaller parachute, that ensures the correct deployment of the main parachute. For detemination of decsent velocity, we use simple equation:

$$v_d = \sqrt{\frac{2m_r g}{\pi \rho_{vz} C_p r^2}}$$

where $C_p$ is drag coefficient of the parachute, for spherical parachute it has approximate value 1.75. [3] Diameters of the parachutes are 700 mm and 500 mm, which should provide descend velocity of 10m m·s$^{-1}$

## VII. Avionics

The entire system is powered through the onboard computer. It is connected to a Li-pol battery using a step up converter.

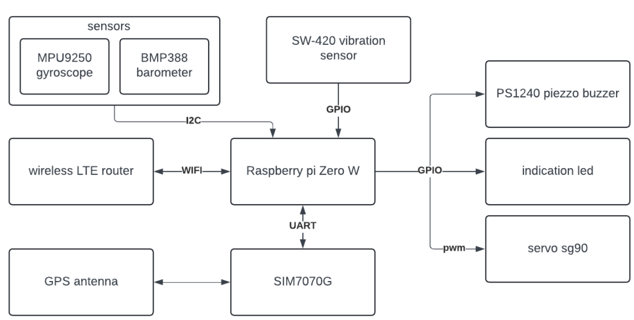



Fig. 5. Wiring of avionics

In addition, an independently functioning ESTES altimeter is located inside the plug-in module for comparison of values.

## VIII. Software

The onboard program is written in Python and is based on cycles that iterate until an external interruption. Their launching and termination is ensured by a control script. We divided the program into several concurrent threads, which share hardware resources. The control script ensures most of the logic and communication. A special thread ensures reading and storing data from sensors, another two threads control the indicator LED and piezzo.

A parameter influencing the behavior of the rocket is its state.

- In the *disconnected* state, the indicator LED blinks. The state is detected according to the success of establishing communication with the server.
- In the *connected* state, the indicator LED lights up and the rocket receives remote commands. The state is detected in the same way as the previous one.
- The rocket is manually put into the *ready* state. This state means that the rocket is placed on the launching pad and will no longer be manipulated with it and therefore the calibration of the sensors is started. The state also allows the detection of the start according to vibrations.
- The *start* state is detected in the ready state according to vibrations. It has only an informational function.
- The *apogee reached* state is detected in all states (including the *disconnected* state) barometrically and in the *ready* state also according to vibrations and in the *in flight* state also according to inclination, if calibration was performed. Transition to this state triggers the activation of the rescue device.
- The *landing* state is detected according to the altitude and vibrations (both variants). In this state, the rocket already lies motionless on the ground. The piezzo is activated.

The last three states are detected regardless of the connection, so they take precedence over the *disconnected* state. In all states, beginning with the *connected* state, it is possible to manually switch on and off data collection, release and secure the rescue device mechanism and switch on and off the piezzo. Furthermore, the rocket can receive a command to send the entire video, send telemetry data as csv, delete image records and delete telemetry data.

### A. Calibration

The sensors have factory calibration, the exception is the gyroscope, for which the average basic value is calculated from the average of values from 2000 iterations, which is then subtracted from the raw values. Next, the altimeter must be calibrated, there the detected altitude of the start place is subtracted. The basic value of the altitude according to GNSS is determined in the same way.

### B. Data Collection

The thread reads all sensor data and captures an image with a delay of 200 ms. The image is saved in jpeg format, while the average of the last five data is added to the common csv file after five cycles.

### C. Communication

The thread sends a package of data obtained by reading the last line of the csv file and the image via an http request with a delay of 1000 ms. In the http response, the thread can receive a command to change the state, or perform an action, which the control cycle evaluates and performs.

## IX. Launch Ramp

The launch ramp consists of a long aluminum profile attached to a wooden beam. The beam is anchored in the ground and secured with straps tied to pegs hammered into the ground. The sliding surface of the pad is slightly less than 150 cm.

## X. Flight Control Application

### A. Purpose

We developed the control system to increase the safety and reliability of flights. In our previous rockets, all data was stored only locally. This also applied to the control of the recovery system, so even a little problem with the sensor prevented the parachute from being released. Therefore, it was necessary to ensure a way of remote communication for controlling the rescue system in case of its activation malfunction.

## XI. Flight Control Application

### A. Purpose

We developed the control system to increase the safety and reliability of flights. In our previous rockets, all data was stored only locally. This also applied to the control of the recovery system, so even a little problem with the sensor prevented the parachute from being released. Therefore, it was necessary to ensure a way of remote communication for controlling the rescue system in case of its activation malfunction.

### B. Application Functions

*1) Display of Data and Status Information:* The application can display onboard data from all sensors in real time. Displaying this data can bring us closer details of current onboard situation. This data is read from the server in real time and displayed on the user interface of the application. Displayed are also data about the condition of the rocket and its systems. The last output displayed to the user is visual data. As mentioned, this output is demonstrated in the form of a stream of images, not continuous video, due to the low computational performance of the onboard computer. All this data can serve important role for the user to decide about the need for intervention in the flight and for maintaining a general overview of its course. This is important addition because due to the high velocity, large altitude and apogee and small dimensions of rocket models, direct visual contact from ground is often completely lost during flight.

*2) Rocket Control:* The application allows sending commands to the server, which the onboard computer uses to change the state after reading. These commands can remotely activate the rescue system if the computer incorrectly evaluates the situation and does not perform it automatically, they can command the onboard computer to delete data stored in memory, immediately send a file in mp4 format or a data packet with the history of the entire flight, switch sending data to the server and turn on the piezzo.

### C. Use

Thanks to its modular system, this application allows the connection of a large number of rockets and the control of many flights. This can monitor flights, provide storage and also access to data in the form of an integrated flight control system. This system can be scaled to a full ground flight communication and monitoring dispatch, which takes care of the operation and safety of rocket transport with sensors monitoring the state of the onboard situation and in case of a crisis, it will take safety steps allowing the rescue of material and onboard data. Scalability also applies to the number of rockets and flights. This application also allows long-term storage of flight data, ensures higher safety and reliability and helps with the control and management of individual flights and entire flight series.

### D. Implementation

*1) Desktop Part:* The frontend is divided into several parts. The logic is always separated from the display. The logic is divided into so-called Viewmodels, which handle data for display, Services, which handle the API, and models, serving as a container for specific data. We focused on ensuring that classes perform a specific function according to the paradigm of object-oriented programming.

When the application is launched, it loads the main page, where there is a layout of data display and the possibility to select a rocket and flight, or create a new flight. By flight, we mean the association of onboard data that falls under one rocket launch. After selecting a rocket and flight, the application calls the server address using an HTTP client, which returns the requested onboard data via the API, which the application processes and converts into objects using AutoMapper, which it displays on the screen. Each time new onboard data is loaded, a new request for onboard data is sent. Shortly after the onboard data is saved, a request for visual display data is sent. They immediately call for another and newer one after loading and displaying. In case of connection loss, the calling cycle is interrupted, in this case, the user can press a button that calls the API and thus restores the cycle. Last but not least, there are buttons needed to send a command, i.e., an HTTP request with the necessary data to the API. The entire application is capable and ready for any modification of the incoming data packet and can easily be adapted to the possibilities of other types of data than sensory and safety data and commands for remote control of the rocket. In the future, it will be possible to extend the packages for financial possibilities, statistical data on the cost of a specific rocket, and the possibility of monitoring multiple flights simultaneously.

*2) API:* The API is based on ASP.NET CORE and operates on the principles of RESTful API. It contains endpoints that the desktop application can call and perform the necessary actions according to them. Onboard data is handled by 5 endpoints. The first GET endpoint /api/onboard-data is used by the desktop application to read all data from a specific flight. The second GET endpoint /api/onboard-data/download-photo is used to download binary data of the visual file sent by the onboard computer, stored on the server. The third and last GET endpoint /api/onboard-data/download-csv sends the data of the csv file with all onboard data of a specific flight. The other two endpoints are of the POST type. The first /api/onboard-data/upload-record is used to store data sent by the onboard computer. The data that does not come in the package is replaced by the computer with the data from the last saved data of a specific flight. Finally, this point returns an array of command objects from the command queue, which are sent to the onboard computer for evaluation. The second post /api/onboard-data/upload-photo, on the other hand, serves the onboard computer as a way to upload binary data of individual frames of the rocket's visual output.

The others are commands, which contain only two endpoints, namely GET /api/commands, which returns an array of all existing commands that can be entered and POST /api/rockets/{rocketId}/commands/{commandId}, which adds a specific command to the command queue of a specific rocket.

*3) Server:* Due to the need for cost accessibility, the server hardware was approached using VPS. This decision allows us to scale the project as needed and increase parameters such as CPU computing power (number of cores), operating memory capacity, or SSD storage size. Thanks to the simple architecture, a smooth transition to our own integrated server is possible. This server is running the Linux operating system Ubuntu, which takes care of running the API, database, and web server. Security is mostly taken care of by UFW. This firewall blocks all ports except selected ones, thereby increasing the security of the entire ecosystem. NginX was chosen as the web server due to the possibility of creating a proxy server, which redirects requests from URL addresses and the allowed port to the application port 5000, which is then able to process the request.

*4) Database:* A classic SQL database MariaDB, which is a branch of MySQL, was used as the database server. This database is suitable for simple write and read operations, which allowed us to reduce the response time between the rocket, server, and application [1].

The data is divided into several tables, each storing specific data. The database can be divided into several parts. The first takes care of commands and their sorting into queues for individual rockets, it contains a table of existing commands and a table for a specific queue. The second takes care of the management of rockets, their categorization, and accesses. The third takes care of all onboard data and data. There is a table for a specific flight data and the creation of individual data, which are therefore easily adaptable for use on individual rockets with various sensors. It contains a table of flights, to which individual records of onboard data are assigned.

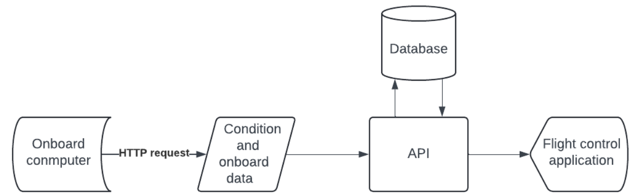



Fig. 6. Flight control application function

## XII. Flight Simulation using openRocket Software

To determine the aerodynamic properties of the rocket and simulate the flight, we use the open source software openRocket.

The program can also simulate some external influences, such as wind. However, the calculated values are usually too optimistic, which may be partly due to manufacturing inaccuracies of the rocket and motor. Typically, the calculated ascend time is greater the real ascend time [5].

TABLE III
Values obtained by simulation

| | |
|---|---|
| Apogee | 900 m |
| Ascend time | 14 s |
| Maximum acceleration | 290 m·s$^{-2}$ |
| Maximum velocity | 210 m·s$^{-1}$ |
| Mach number at maximum velocity | 0,617 |

## XIII. Conclusion

In the presented article, we designed and built a functional single-stage model rocket and developed an application for controlling the flight of rocket models. During the work, we acquired knowledge in the field of rocket aerodynamics, design of solid propellant engines, recovery systems, and flight software, and we gained skills in the field of development of integrated systems for controlling unmanned vehicles.

In the future, we plan to test our rocket more, perform several flights, improve data collection and processing, and refine flight predictions. We also want to expand the functionality of the application system and subsequently publish the source code of the application and flight software.

## References


[1] Mariadb foundation. Available at: https://mariadb.org/. [cit. 28-2-2024].
[2] openmotor. Available at: "https://github.com/reilleya/openMotor?tab=GPL-3.0-1-ov-file. [cit. 16-2-2024].
[3] Velocity during recovery. Available at: https://www.grc.nasa.gov/WWW/k-12/VirtualAero/BottleRocket/airplane/rktvrecv.html. [cit. 16-2-2024].
[4] R. Nakka. Knsb propellant. Available at: http://nakka-rocketry.net/sorb.html. [cit. 16-2-2024].
[5] S. Niskanen. Openrocket technical documentation for openrocket version 13.05. pages 5–19, 41–43, 56, 72, 84–88, 2013.
[6] C. Prof. Ing. Pavel Konečný. *Vnější balistika raket*. Univerzita obrany, 2006.


BRNO UNIVERSITY OF TECHNOLOGY
FACULTY OF ELECTRICAL ENGINEERING AND COMMUNICATION
TECHNICKÁ 3058/10, 616 00 BRNO, CZECH REPUBLIC

BRNO UNIVERSITY OF TECHNOLOGY
FACULTY OF ELECTRICAL ENGINEERING AND COMMUNICATION

www.eeict.cz

www.fekt.vut.cz