THE MINISTRY OF SCIENCE AND HIGHER EDUCATION OF THE RUSSIAN FEDERATION

ISSN 2687-0517

Computing, Telecommunications and Control

Vol. 18, No. 3
2025

Peter the Great St. Petersburg
Polytechnic University
2025

The journal is included in the List of Leading PeerReviewed Scientific Journals and other editions to publish major findings of PhD theses for the research degrees of Doctor of Sciences and Candidate of Sciences.

Open access journal is to publish articles of a high scientific level covering advanced experience, research results, theoretical and practical problems of informatics, electronics, telecommunications, and control.

The journal is indexed by Ulrich's Periodicals Directory, Google Scholar, EBSCO, ProQuest, Index Copernicus, VINITI RAS Abstract Journal (Referativnyi Zhurnal), VINITI RAS Scientific and Technical Literature Collection, Russian Science Citation Index (RSCI) database Scientific Electronic Library and Math-Net.ru databases.

The journal is registered with the Federal Service for Supervision in the Sphere of Telecom, Information Technologies and Mass Communications (ROSKOMNADZOR). Certificate ЭЛ No. ФС77-77378 issued 25.12.2019.

Editorial office

Dr. Sc., Professor A.S. Korotkov – Editor-in-Chief;

Ph.Ch.S. Bastian – literary editor, proofreader; G.A. Pyshkina – editorial manager; A.A. Kononova – computer layout; I.E. Lebedeva – English translation.

Address: 195251 Polytekhnicheskaya Str. 29, St. Petersburg, Russia.

+7 (812) 552-6216, e-mail: infocom@spbstu.ru

Release date: 30.09.2025

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ

ISSN 2687-0517

Информатика, телекоммуникации и управление

Том 18, № 3
2025

Санкт-Петербургский политехнический
университет Петра Великого
2025

ИНФОРМАТИКА, ТЕЛЕКОММУНИКАЦИИ И УПРАВЛЕНИЕ

Журнал с 2002 года входит в Перечень ведущих рецензируемых научных журналов и изданий, в которых должны быть опубликованы основные результаты диссертаций на соискание ученой степени доктора и кандидата наук.

Сетевое издание открытого доступа публикует статьи высокого научного уровня, освещающие передовой опыт, результаты НИР, теоретические и практические проблемы информатики, электроники, телекоммуникаций, управления.

Сведения о публикациях представлены в Реферативном журнале ВИНИТИ РАН, в международной справочной системе «Ulrich`s Periodical Directory», в Российской государственной библиотеке. В базах данных: Российский индекс научного цитирования (РИНЦ), Google Scholar, EBSCO, Math-Net.Ru, ProQuest, Index Copernicus.

Журнал зарегистрирован Федеральной службой по надзору в сфере информационных технологий и массовых коммуникаций (Роскомнадзор). Свидетельство о регистрации ЭЛ № ФС77-77378 от 25.12.2019.

Учредитель и издатель: Санкт-Петербургский политехнический университет Петра Великого, Санкт-Петербург, Российская Федерация.

Редакция журнала

д-р техн. наук, профессор А.С. Коротков – главный редактор;

Ф.К.С. Бастиан – литературный редактор, корректор; Г.А. Пышкина – ответственный секретарь, выпускающий редактор;

А.А. Кононова – компьютерная вёрстка; И.Е. Лебедева – перевод на английский язык.

Адрес редакции: Россия, 195251, Санкт-Петербург, ул. Политехническая, д. 29.

Тел. редакции +7(812) 552-62-16, e-mail редакции: infocom@spbstu.ru

Дата выхода: 30.09.2025

Contents

Intelligent Systems and Technologies, Artificial Intelligence

Circuits and Systems for Receiving, Transmitting, and Signal Processing

Software and Hardware of Computer, Network, Telecommunication, Control, and Measurement Systems

System Analysis and Control

Содержание

Интеллектуальные системы и технологии, искусственный интеллект

Устройства и системы передачи, приема и обработки сигналов

Компьютерные сети, вычислительные, телекоммуникационные, управляющие и измерительные системы

Системный анализ и управление

Computing, Telecommunication and Control, 2025, Vol. 18, No. 3, Pp. 9–22.
Информатика, телекоммуникации и управление. 2025. Том 18, № 3. С. 9–22.

Intelligent Systems and Technologies, Artificial Intelligence

Интеллектуальные системы и технологии, искусственный интеллект

Research article
DOI: https://doi.org/10.18721/JCSTCS.18301
UDC 620.179; 004.93

APPLICATION OF NEURAL NETWORKS FOR DETECTING DEFECTS AND DAMAGE IN METAL STRUCTURES

V.L. Agamirov[1,2] ✉, *L.V. Agamirov*[3,1], *M.A. Nosikov*[4], *N.V. Toutova*[1], *A. Khaush*[1]

[1] Moscow Technical University of Communications and Informatics, Moscow, Russian Federation;

[2] Moscow Aviation Institute (National Research University), Moscow, Russian Federation;

[3] National Research University "Moscow Power Engineering Institute", Moscow, Russian Federation;

[4] BRICS University, Moscow, Russian Federation

✉ avhere@yandex.ru

Abstract. The rapid development of neural networks has led to the integration of these technologies into various industrial sectors. At the same time, improving the accuracy and efficiency of detecting defects and damages, including in real-time, remains a critical task. By combining neural networks with the Internet of Things (IoT) and technologies for data collection, storage and protection, it is possible to create a comprehensive and effective information-measurement system for surface defect detection. In this context, the present work highlights recent advances in the application of artificial intelligence for quality control, as well as the detection of defects and damages in structures. The focus is on the development and training of neural networks capable of effectively identifying and classifying various types of defects. The study demonstrates how these technologies significantly improve the speed and accuracy of diagnostics compared to traditional visual and instrumental inspection methods. The results of model testing on real industrial data confirm the high efficiency of the proposed approach. Additionally, the authors have developed an algorithm and implemented software for the automatic annotation of images in a format suitable for modern architectures such as YOLO. This approach enables the effective application of the model for detecting damages on the surfaces of structures and systems using widely available types of datasets.

Keywords: defect detection, neural networks, YOLO architecture, structural integrity monitoring, infrared thermography, automated defect analysis, machine learning for defectoscopy

Citation: Agamirov V.L., Agamirov L.V., Nosikov M.A. et al. Application of neural networks for detecting defects and damage in metal structures. Computing, Telecommunications and Control, 2025, Vol. 18, No. 3, Pp. 9–22. DOI: 10.18721/JCSTCS.18301

Научная статья
DOI: https://doi.org/10.18721/JCSTCS.18301
УДК 620.179; 004.93

ПРИМЕНЕНИЕ НЕЙРОННЫХ СЕТЕЙ ДЛЯ ВЫЯВЛЕНИЯ ДЕФЕКТОВ И ПОВРЕЖДЕНИЙ МЕТАЛЛИЧЕСКИХ КОНСТРУКЦИЙ

В.Л. Агамиров[1,2] ✉, *Л.В. Агамиров*[3,1], *М.А. Носиков*[4], *Н.В. Тутова*[1], *А. Хауш*[1]

[1] Московский технический университет связи и информатики, Москва, Российская Федерация;

[2] Московский авиационный институт (национальный исследовательский университет), Москва, Российская Федерация;

[3] Национальный исследовательский университет «МЭИ», Москва, Российская Федерация;

[4] Университет БРИКС (ЮниБРИКС), Москва, Российская Федерация

✉ avhere@yandex.ru

Аннотация. Бурное развитие нейронных сетей привело к интеграции этих технологий в различные отрасли промышленности. В то же время повышение точности и эффективности выявления дефектов и повреждений, в том числе в реальном времени, остается актуальной задачей. Комбинируя нейронные сети с Интернетом вещей (IoT), а также технологиями сбора, хранения и защиты данных, возможно создать комплексную и эффективную информационно-измерительную систему для обнаружения поверхностных дефектов. В этом контексте данная работа освещает современные достижения в применении искусственного интеллекта для контроля качества, а также выявления дефектов и повреждений в конструкциях. Основное внимание уделяется разработке и обучению нейронных сетей, способных эффективно идентифицировать и классифицировать различные типы дефектов. Исследование демонстрирует, как эти технологии значительно повышают скорость и точность диагностики по сравнению с традиционными методами визуального и инструментального контроля. Результаты тестирования моделей на реальных промышленных данных подтверждают высокую эффективность предложенного подхода. Кроме того, авторами разработан алгоритм и реализовано программное обеспечение для автоматической аннотации изображений в формате, подходящем для современных архитектур, таких как YOLO. Данный подход позволяет эффективно применять модель для обнаружения повреждений на поверхностях конструкций и систем, используя широкодоступные типы наборов данных.

Ключевые слова: обнаружение дефектов, нейронные сети, архитектура YOLO, мониторинг целостности конструкций, инфракрасная термография, автоматизированный анализ дефектов, машинное обучение для дефектоскопии

Для цитирования: Agamirov V.L., Agamirov L.V., Nosikov M.A. et al. Application of neural networks for detecting defects and damage in metal structures // Computing, Telecommunications and Control. 2025. Т. 18, № 3. С. 9–22. DOI: 10.18721/JCSTCS.18301

Introduction

In the design and operation of engineering products, special attention is given to methods for detecting macrocrack-type defects in critical structural elements, as these significantly impact the reliability and lifespan of complex technical systems. This is particularly relevant for structures operating under variable loads in aviation, railway and automotive transport, such as bridge structures, road surfaces etc.

Defects of this type can occur on the surfaces of load-bearing structural elements and lead to product failures with potentially catastrophic consequences, especially in aviation. The causes of surface defects

include both objective factors, such as structural heterogeneity of construction materials, maximum surface stresses during deformation and characteristics of surface material layers, and subjective factors, including deviations in the shape and dimensions of products, dents, scratches, other manufacturing defects and operational errors like hard aircraft landings.

The development and creation of artificial intelligence (AI) devices and systems aimed at addressing this problem is a highly relevant task. A critical requirement in this process is the transformation of raw photo and video data obtained from devices into annotated data structures suitable for processing by neural networks (NNs).

In this work, the authors have developed an information-measurement system for detecting surface defects. This includes the design and software implementation of an algorithm for automatic annotation of images into a format compatible with the state-of-the-art YOLO architecture. The algorithm processes widely used dataset types, such as run length encoded (RLE) masks on images. This solution enables the application of the model for efficient defectoscopic inspection of engineering structures and products.

Before implementation, it is essential to examine existing approaches and solutions for detecting defects in metal structures. The study [1] explores intriguing NN training methods based on various YOLO versions and other types of NNs. It also evaluates multiple methods for assessing detection efficiency across different spectra. The results demonstrate that for visible spectrum, the most effective methods are: GVN, HOG + SVM, SSD, GrabCut, cascading CNN, LBPHF + SVM, DMNN, VGG-19, LBP + ULBP, YOLO v3, DELM + LRF, SVM, Faster R-CNN, CNN, stereovision + PLAMEC. The most significant achievement was solving the key issue of inspecting power line structures (PLS) using unmanned aerial vehicles (UAV) by introducing mobility and flexibility to the process, enabling effective investigation of hard-to-reach areas.

The study [2] outlines the primary shortcomings and causes of defects in massive metal structures. It further demonstrates how infrared thermography provides a clear assessment of technical conditions, which is critical for the safe operation of structures. Infrared thermography is based on surface temperature monitoring using thermal imaging in the infrared range. Unfortunately, a suitable sensor for these studies has not yet been identified, making this method currently impractical.

In the study [3], a fiber-optic sensor was utilized to evaluate load on load-bearing elements of bridge structures. The sensor's key feature is its use of Bragg gratings and an interrogator, allowing precise measurement of relative deformation and temperature. While this method is excellent for continuous monitoring and structural reinforcement, it is not suitable for emergency integrity assessments of bridge structures.

Method

NNs and defect detection

NNs, particularly convolutional NNs (CNNs), have proven highly effective in image classification and detection tasks. In the context of structural analysis, NNs can process images and videos captured by drones or stationary cameras, automatically identifying signs of damage, such as cracks, corrosion, deformations and other defects.

YOLO (You Only Look Once) – a NN architecture designed for object detection in images [4–8].

The YOLO model employs a fully CNN that predicts bounding boxes and object classes in a single process. The input image is divided into a grid, with each grid cell predicting:

- the bounding box;
- the confidence that the bounding box contains an object;
- the probabilities of the object's belonging to specific class.

CNNs can extract and interpret complex visual features, making them an effective tool for automatic detection of defects on the surface of structures [9–14].

AI-based methods enhance defect detection across various domains. Studies have improved PCB inspection using YOLO v8 and AOI technology [15, 16]. A hybrid AI approach has been proposed for defect control [17]. Deep learning aids material classification and real-time road defect detection [18, 19].

CNN architecture

The architecture of CNNs includes several key components, each of which plays a crucial role in image analysis:

- **Convolutional layers.** These layers apply filters to the input images, extracting important features such as edges, corners and textures. For example, in the context of bridge diagnostics, convolutional layers can identify cracks, signs of corrosion and other defects.
- **Activation function.** Rectified Linear Unit (ReLU) adds non-linearity to the model, allowing it to learn more complex and abstract features of the images.
- **Pooling layers.** These layers reduce the dimensionality of feature maps while preserving the most significant information. This helps decrease the amount of data to process and makes the model less sensitive to the exact location of features.
- **Fully connected layers.** These layers combine the extracted features and use them for classification or regression, allowing the determination of the extent and type of damage in the bridge structure.

CNN-based models, particularly YOLO, have significantly advanced defect detection across industries. Studies highlight improvements in solar panel and metal surface inspection [20–22], road defect segmentation [23] and steel strip analysis [24]. The evolution from YOLO v1 to YOLO v8 further enhances CNN-driven industrial defect detection [25].

Problem of data heterogeneity

Different research groups and companies use various datasets to train their models. This can make it difficult to apply existing data for training a new NN. Different projects may require different NNs, each with specific data and annotation requirements for training. In such cases, data transformation mechanisms can be used, but in some situations, finding an acceptable solution may not be possible. For example, an RLE dataset is not directly compatible with the YOLO NN.

In this work, the authors have developed a data converter from RLE format to a format compatible with YOLO[1] [26]. RLE is a mask on the image that can label an object of any shape or type, but YOLO requires the coordinates of the top-left and bottom-right corners of a rectangle containing the object. In Fig. 1, on the right, the yellow-purple image is the mask on the picture, provided from the training data set of Severstal[2] [27, 28], while on the left is the original photo from Severstal's dataset, to which the authors applied the developed program that determines the coordinates of the defect marking squares for use in the YOLO NN.

YOLO accepts annotations in the form of a separate file with coordinates in the format: <class_label> <x_center> <y_center> <width> <height>. In the original Severstal dataset, defect annotations (labels) are stored in a separate .csv file as a string containing the pixel number and the count of subsequent pixels. As a result of applying the program developed by the authors, by referencing the image mask and using the cv2 library[3], with the parameters "CV_RETR_EXTERNAL" and "CV_CHAIN_APPROX_SIMPLE", we obtain the contours of adjacent mask pixels.

The mask (on the right in Fig. 1) is a new layer overlaid on the original image, showing only the damaged area (yellow) and the image area (purple).

---

[1] See the guide on working with YOLO in Python on the website https://docs.ultralytics.com/ru/usage/python and the source code together with instructions for running the software implementation by the authors of automatic image annotation for YOLO on the website https://github.com/Nosikmov/Kaggle-Dataset-to-YOLO.

[2] The dataset of images and defect annotations of Severstal steel sheets: https://www.kaggle.com/competitions/severstal-steel-defect-detection/data, a software implementation of defect detection on steel sheets using convolutional neural networks by a prizewinner of competitions on the Kaggle platform: https://www.kaggle.com/code/lightforever/severstal-mlcomp-catalyst-infer-0-90672.

[3] An open-source library for computer vision, machine learning, and image processing; for details see https://docs.opencv.org/4.x/d1/dfb/intro.html.

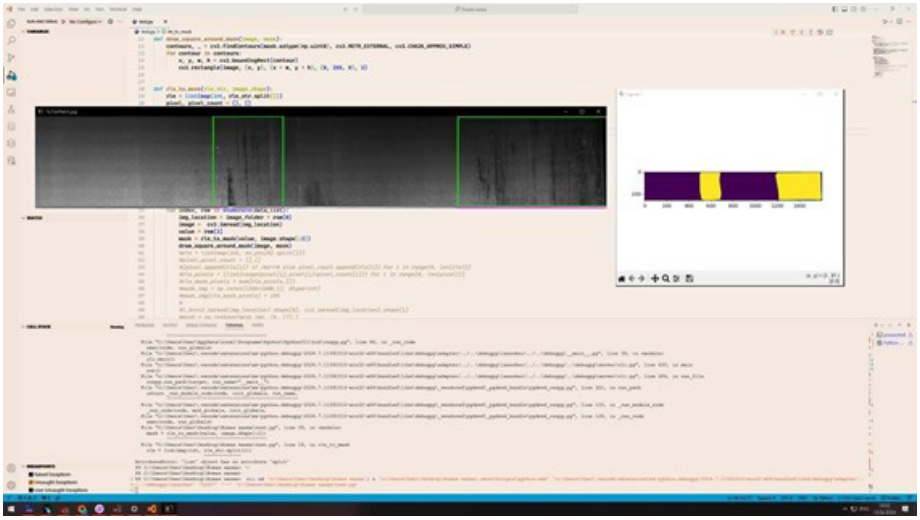


Fig. 1. The result of the data preparation program for training the YOLO v8 model

The contours (on the left in Fig. 1, green) represent the edges of the damage area in the mask. Since the damage area can have an arbitrary shape and YOLO works with rectangles, it is necessary to find the edges of the damage area in the mask to obtain the bounding rectangle.

Adjacent contour pixels are first converted into lines, using the pixel coordinates and their count in the line. If the lines are located next to each other, they are considered adjacent, and in this case, they need to be grouped into a single bounding rectangle annotation for YOLO.

The obtained contours can be converted into values suitable for the YOLO annotation format. Since the annotation requires rectangle coordinates, it is necessary to determine the center of the rectangle by averaging the corresponding values. The width and height of the rectangle are calculated as the difference between the relevant coordinates. The center, width and height are then normalized by dividing by the image dimensions. This process forms a defect annotation string, which is compatible with YOLO[4] (Figs. 2, 3).

An example of the result from the trained model on the converted data is shown in Fig. 4.

Advantages of CNN

- **Superior processing of visual data.** CNNs are capable of identifying complex visual patterns, which is especially important for detecting cracks and other structural anomalies in images and videos.
- **Generalization to new data.** After training on a large dataset, CNNs can successfully apply their knowledge to new, previously unseen images, providing high accuracy under various conditions.
- **Automation.** The use of CNNs for structural monitoring reduces the need for frequent physical inspections, which lowers costs and enhances safety.

Recent advancements in CNN-based YOLO models have significantly improved defect detection across multiple domains. Enhanced architectures have been applied to surface, packaging and PCB defect detection [29–32], as well as to steel, weld and pavement inspections [33–35]. Additionally, specialized adaptations have optimized insulator defect analysis [36].

Training NNs

To achieve high accuracy in defect detection, NNs must be trained on large datasets containing images of various types of defects. These datasets were sourced from open datasets. It is important to ensure the protection of training data using modern encryption methods and blockchain technologies, which guarantee the integrity and security of the data.

[4] See https://github.com/Nosikmov/Kaggle-Dataset-to-YOLO for details on how to transform the Severstal dataset into a suitable dataset for YOLO.

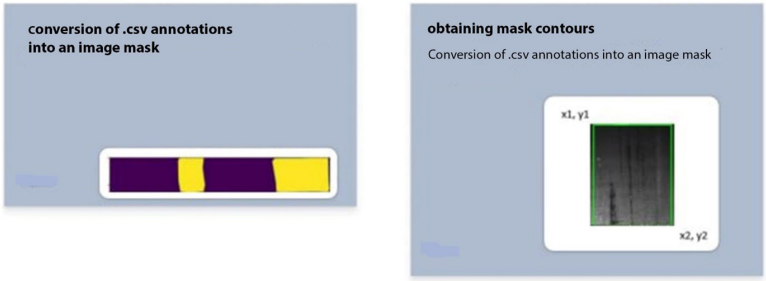


Fig. 2. Process of programmatically converting defect annotations into YOLO format

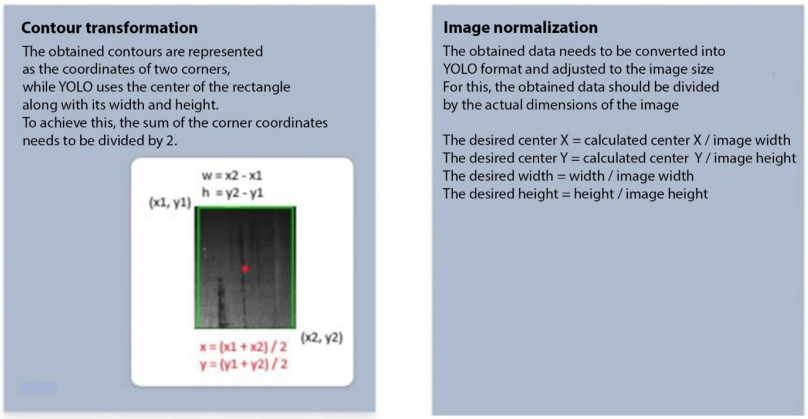


Fig. 3. Obtained data needs to be converted into YOLO format and adjusted to the image size, therefore, it should be divided by the actual dimensions of the image

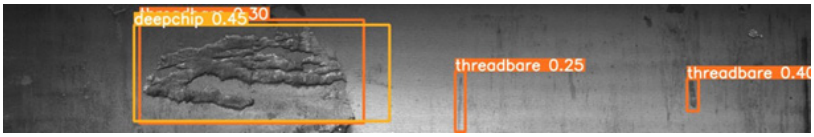


Fig. 4. Results of the YOLO model evaluation

Application of IoT for data collection

The Internet of Things (IoT) devices, such as vibration, temperature and deformation sensors, are installed on bridge structures for continuous monitoring of their condition. These sensors transmit real-time data to servers or cloud platforms, where it is analyzed by NNs to detect anomalies and assess risk.

Sensor types

- Vibration sensors detect changes in vibrations caused by cracks or other structural alterations.
- Strain sensors detect changes in shape or deformation of structural elements.
- Thermal sensors monitor temperature changes that may indicate internal issues, such as material strength degradation.
- Acoustic sensors are used to detect sounds caused by cracks or delaminations in the material.

Dash cameras, road cameras, drones

Modern technologies provide a variety of tools for monitoring road surface conditions. Among them, road cameras, dash cameras and drones stand out. These devices enable the collection of visual information about the roads, which helps improve their condition and enhance road traffic safety.

The use of drones to capture images of road surfaces offers several advantages, including the ability to inspect hard-to-reach sections of roads.

Dash cameras are installed in vehicles and record the road conditions in real time.

Advantages of IoT are as follows:

• **Continuous monitoring**: Continuous data collection allows for quick response to changes in the condition of the structure.

• **Process automation**: The exclusion of the human factor from the data collection process reduces the likelihood of errors.

• **Wide availability**: IoT devices can be installed even in hard-to-reach places, providing more comprehensive monitoring coverage.

Mobile devices for data collection and analysis

Mobile devices, such as smartphones and tablets, can be used to collect data from sensors and cameras, as well as for preliminary processing. With specialized applications, primary diagnostics of the condition of structures can be performed and data can be transmitted to central servers, utilizing blockchain technologies for more detailed analysis by NNs.

Advantages of mobile devices are as follows:

• **Convenience of use**: Mobile devices are always at hand and allow for quick collection of the necessary information.

• **Mobility**: The ability to work anywhere and anytime.

• **Interactivity**: Applications can provide users with instant feedback and recommendations.

Blockchain

Blockchain is a distributed ledger in which data is stored in the form of a chain of blocks. Each record is protected by cryptographic methods and is linked to previous records, which prevents unauthorized data alteration.

Key features of blockchain technologies are as follows:

• **Data integrity**: The data recorded in the blockchain cannot be altered without a corresponding entry in the blockchain, which prevents falsification.

• **Transparency**: Each participant in the network can verify the authenticity of the data, which increases trust in the system.

• **Security**: The use of cryptography protects data from unauthorized access.

• **Decentralization**: The absence of a central controlling authority eliminates single points of failure and enhances the system's resilience.

Let us consider the process of integrating blockchain technologies to ensure the reliability of data storage used for training NNs, using the example of bridge structure monitoring.

• **Data collection**: IoT sensors installed on bridges collect data on the condition of the structures, such as vibrations, deformations and temperature changes.

• **Data recording in blockchain**: The collected data is sent to the blockchain, where it is recorded as immutable blocks. Each block contains the hash of the previous block, creating a data chain.

• **Data analysis**: NNs access the data from the blockchain for analysis and model training. Due to the high reliability and integrity of the data, NNs are trained more accurately and efficiently.

• **Monitoring and updating**: Data is continuously updated and added to the blockchain, ensuring the relevance of the information for analysis and decision-making.

Example of usage

Let us assume that vibration sensors on the bridge detect anomalies that could indicate the beginning of a crack. The vibration data is recorded in the blockchain, where it can be verified and analyzed by NNs to determine the severity of the defect and the need for repair.

Data collection from satellites

Satellites equipped with high-resolution cameras regularly capture Earth's surfaces, including road networks. These images cover large areas and provide up-to-date information.

Defect detection from defectogram images

NNs can be trained using images from defectograms, which are graphical representations of structural defects. Such images can be obtained through various non-destructive testing methods, such as ultrasonic, radiographic and thermographic diagnostics.

Deep CNNs (DCNNs) play a crucial role in defect detection across multiple domains. Advanced models enhance glass insulator inspection and electron microscopy diagnostics [37, 38], while DCNN-based approaches improve defect identification in industrial manufacturing [39–41]. Additionally, NNs optimize defect depth estimation and automated surface analysis [42–44].

Process of image analysis

- **Image collection**: Defectograms are collected from various diagnostic methods.
- **Preprocessing**: The images are processed to improve quality and remove noise.
- **Model training**: The NN is trained on pre-labeled images, where the types and locations of defects are marked.
- **Analysis and detection**: The trained model analyzes new images and identifies defects, specifying their type and location.

In the context of the study, the following NNs were used:

- NNs trained for crack and defect recognition in steel;
- NNs trained for tuberculosis detection from fluorography images.
- The YOLO NN designed for fast and accurate detection and classification of objects in real-time image processing.

These networks were retrained to recognize cracks and defects on bridges using specialized datasets with images of bridge structure damage.

For training the NNs, datasets containing images of various bridge damages were used, including cracks, corrosion, deformations and other defects, as well as images from dash cameras and road cameras showing potholes and clean road surfaces. These images were meticulously labeled to accurately indicate the defects.

Results

After retraining, the NNs demonstrated successful defect detection on objects that needed to be identified. The accuracy of detection was high, ranging from 70 to 90%, which confirms the effectiveness of the method. The NNs were able to adapt to new conditions and deliver good results in the assigned tasks.

The application of NNs, initially trained for other tasks, and their retraining for specific defect monitoring conditions, demonstrates high efficiency and potential for real-world applications.

Use of NNs in conjunction with the IoT

The IoT and NNs, when used together, are convenient for processing real-time data, allowing for continuous receipt of up-to-date information in a processed form.

The process consists of a 3-step algorithm:

1. Acquiring information from IoT sensors.
2. Processing data through an intermediate operating system or other real-time processing system.
3. Transfer of the previously processed data to the NN.

This configuration allows for automatic data transmission without noticeable losses and adds several debugging tools to optimize and increase the accuracy of the NN model.

Implementation of a damage and defect detector on Raspberry Pi

For mobility and compactness, single-board computers with sufficient computing power, such as the Raspberry Pi Model 4B, were chosen. The board supports Linux operating systems, which allows using the Python programming language without significant limitations. The IoT sensor used is the OV5647 camera, which supports Full HD and infrared lighting for night-time recording.

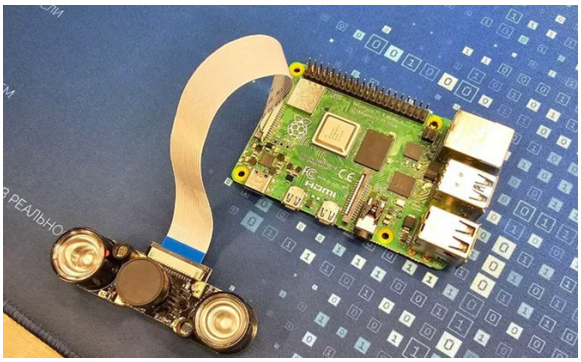


Fig. 5. Minimal configuration

Fig. 6. Assembled detector

Fig. 5 shows the minimal setup required for the detector to work, consisting of a single-board computer, a compatible camera and an SD card that stores our program code in auto-start mode, as well as the Linux operating system (Raspberry Pi OS). Additionally, a keyboard, mouse and monitor can be added for convenient use of debugging tools. If peripheral devices are unavailable for viewing results, wired or wireless (Bluetooth) headphones are recommended for listening to audio signals during defect detection, allowing for mobile operation with the device.

Fig. 6 shows the fully assembled detector, ready for operation.

The design is suitable for continuous operation in a fixed location or for mobile use in various areas. This implementation of the detector enables solving a wide range of tasks, such as automating the assessment of metal quality for various industries and evaluating the integrity of structures in hard-to-reach and poorly lit areas.

Architecture of the hardware-software complex

The architecture of the hardware-software complex of the detector is described in Fig. 7[5].

The entire architecture is hosted on a Raspberry Pi running a Linux operating system, which facilitates interaction with the camera through libraries and drivers. After receiving the image from the camera, the operating system passes the data to the virtual Python environment, specially configured for the detector. Using the picamera2 library, the image is converted to JPG format in the main thread of our program code. The image is then transformed into a matrix and passed to the NN for processing and defect detection. In the child thread, created upon launching the program, a monitoring window is available to track the image and camera angle, which will allow for optimization and improvement of the defect detection accuracy by the NN in the future.

[5] See Raspberry Pi documentation: https://www.raspberrypi.com/documentation/.

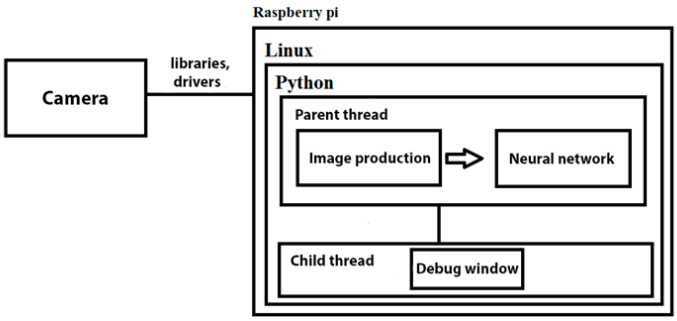


Fig. 7. Architecture of the hardware-software complex

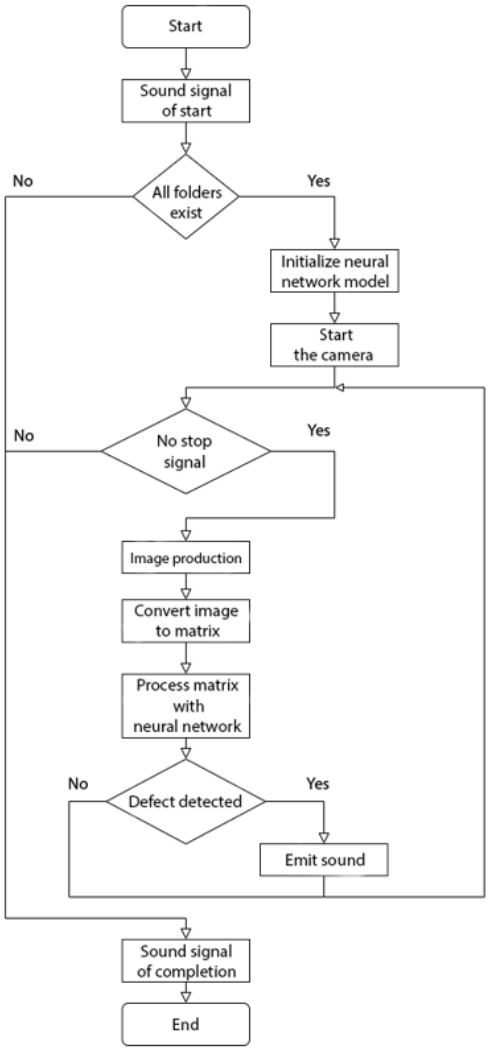


Fig. 8. Flowchart of the program code

Block diagram of the program code

The listing of the program that uses a camera to capture images and send them to the NN for processing, with added debugging tools such as a camera window and the display of NN results in the console, is available on the website https://github.com/ikly323/raspberry_metal. Fig. 8 shows the flowchart of this program code.

When the program is launched, an initial sound signal is emitted to notify the user that the detector has started. After the sound, the program checks the integrity of the project and the presence of all necessary files, which ensures the initialization of the NN model and the activation of the camera for further interaction. After all checks and service launches are completed, an infinite loop starts, responsible for defect detection and notifying the user. To do this, a frame is captured, saved as a file in the operating system, and then passed to the NN as a matrix. Based on the NN's prediction, a decision is made whether to emit an audio notification or not. The program completes its life cycle when the Ctrl + C command (kill signal) is issued or when the Raspberry Pi is powered off.

Conclusion

The use of NNs in combination with IoT, blockchain technologies and mobile devices represents a powerful tool for monitoring and assessing the process of detecting surface defects in structural elements and technical systems, ensuring high accuracy and prompt detection. The implementation of this development will contribute to improving the safety, resilience, reliability and lifespan of mechanical engineering products, road and railway surfaces. Additionally, based on the conducted research, a defect detection process for metals during the production preparation phase can be implemented in autonomous mode.

The architecture of the developed detector allows for a versatile expansion of defect detection capabilities by adding advanced analysis methods. In the future, ultrasound may be used to study macrocracks, as well as a more detailed surface analysis using time-of-flight sensors.

REFERENCES

1. **Astapova M.A., Lebedev I.V.** Review of approaches to the detection of defects in power transmission line elements in images in the infrared, ultraviolet and visible spectra. *Modeling, optimization and information technology*, 2020, Vol. 8, No. 4, Pp. 1–34. DOI: 10.26102/2310-6018/2020.31.4.036

2. **Glushkov S.P., Solovyev L.U., Borisovskaya N.E.** The detection of defects in the structures reinforced with composites by infrared thermography. *Siberian Transport University Bulletin*, 2015, Vol. 35, No. 4, Pp. 36–42.

3. **Fedotov M.Yu., Loskutov M.L., Shelemba I.S., Kolesnikov A.V., Ovchinnikov I.G.** To the issue of monitoring of metal-bearing bridges supporting fiber optic sensors. *Russian journal of transport engineering*, 2018, Vol. 5, No. 3, Art. no. 11SATS318. DOI: 10.15862/11SATS318

4. **Bao H., Wu W., Yu J., Cao Y., Liu Y., Zhang H., Li X., Shi W.** Application of YOLO model in road defect detection. *2024 5th International Seminar on Artificial Intelligence, Networking and Information Technology* (*AINIT*), 2024, Pp. 2280–2284. DOI: 10.1109/AINIT61980.2024.10581746

5. **Pei J., Wu X., Liu X.** YOLO-RDD: A road defect detection algorithm based on YOLO. *2024 27th International Conference on Computer Supported Cooperative Work in Design* (*CSCW*D), 2024, Pp. 1695–1703. DOI: 10.1109/CSCWD61410.2024.10580137

6. **Kumar A., Kalia A., Verma K., Sharma A., Kaushal M.** Scaling up face masks detection with YOLO on a novel dataset. *Optik*, 2021, Vol. 239, Art. no. 166744. DOI: 10.1016/j.ijleo.2021.166744

7. **Shi R., Li T., Yamaguchi Y.** An attribution-based pruning method for real-time mango detection with YOLO network. *Computers and Electronics in Agriculture*, 2020, Vol. 169, Art. no. 105214. DOI: 10.1016/j.compag.2020.105214

8. **Ali-Gombe A., Elyan E., Moreno-García C.F., Zwiegelaar J.** Face Detection with YOLO on edge. *Proceedings of the 22nd Engineering Applications of Neural Networks Conference. EANN 2021* (eds. L. Iliadis, J. Macintyre, C. Jayne, E. Pimenidis), 2021, Vol. 3, Pp. 284–292. DOI: 10.1007/978-3-030-80568-5_24

9. **Zanevych Yu.** Traffic light detection with YOLO models. *Grail of Science*, 2024, No. 38, Pp. 194–199. DOI: 10.36074/grail-of-science.12.04.2024.033

10. **Xu H., He J., Liang Y., Liu Y., He Y.** Detection of nectarines on trees with YOLO models. *2023 35th Chinese Control and Decision Conference* (*CCDC*), 2023, Pp. 1789–1794. DOI: 10.1109/CCDC58219.2023.10326846

11. **Viswanatha V., Chandana R.K., Ramachandra A.C.** Real time object detection system with YOLO and CNN models: A review. arXiv:2208.00773, 2022. DOI: 10.48550/arXiv.2208.00773

12. **Yu J., Ye X., Tu Q.** Traffic sign detection and recognition in multiimages using a fusion model with YOLO and VGG network. *IEEE Transactions on Intelligent Transportation Systems*, 2022, Vol. 23, No. 9, Pp. 16632–16642. DOI: 10.1109/TITS.2022.3170354

13. **Maleki P., Ramazani A., Khotanlou H., Ojaghi S., Mousavi M., Kalinin A., Mosavi A.** Object detection for vehicles with YOLO. *2024 IEEE 22nd World Symposium on Applied Machine Intelligence and Informatics* (*SAMI*), 2024, Pp. 000343–000350. DOI: 10.1109/SAMI60510.2024.10432884

14. **Sadhin A.H., Hashim S.Z.M., Samma H., Khamis N.** YOLO: A competitive analysis of modern object detection algorithms for road defects detection using drone images. *Baghdad Science Journal*, 2023, Vol. 21, No. 6, Art. no. 24. DOI: 10.21123/bsj.2023.9027

15. **Inyutin A.V., Lukashevich M.M.** PCB defect detection based on YOLOv8 architecture. *System Analysis and Applied Information Science*, 2024, No. 2, Pp. 16–24. DOI: 10.21122/2309-4923-2024-2-16-24

16. **Khodataeva T.S., Kashirin N.V., Averina A.I., Guryanov A.E.** Approaches to the development of a printed circuit board defect detection system based on AOI technology. *Trudy ISP RAN/Proc. ISP RAS*, 2023, Vol. 35, No. 4, Pp. 109–120. DOI: 10.15514/ISPRAS-2023-35(4)-5

17. **Vengerenko V.V., Inyutin A.V.** Hybrid inspection of printed board defects. *Informatics*, 2024, Vol. 21, No. 3, Pp. 63–79. DOI: 10.37661/1816-0301-2024-21-3-63-79

18. **Kuzevanov V.O., Tikhomirova D.V.** Deep learning for effective visualization and classification of recyclable material labels. *Scientific Visualization*, 2024, Vol. 16, No. 5, Pp. 179–196. DOI: 10.26583/sv.16.5.12

19. **Antonov M.O., Temkin I.O.** Real-time recognition and tracking of road surface defects based on complex using of standard computational procedures and deep neural networks. *Software & Systems*, 2024, Vol. 37, No. 3, Pp. 421–430. DOI: 10.15827/0236-235X.142.421-430

20. **Sobolev A.S.** Development of a system for detecting defects in solar panel based on the YOLO v5 algorithm. *Izvestiya Tula State University*, 2023, No. 12, Pp. 480–483. DOI: 10.24412/2071-6168-2023-12-480-481

21. **Xu Y., Zhang K., Wang L.** Metal surface defect detection using modified YOLO. *Algorithms*, 2021, Vol. 14, No. 9, Art. no. 257. DOI: 10.3390/a14090257

22. **Zhang M., Yin L.** Solar cell surface defect detection based on improved YOLO v5. *IEEE Access*, 2022, Vol. 10, Pp. 80804–80815. DOI: 10.1109/ACCESS.2022.3195901

23. **Kanaeva I.A., Spitsyn V.G.** Road damage defects segmentation based on convolutional neural network ensemble. *Tomsk State University Journal of Control and Computer Science*, 2024, Vol. 68, Pp. 75–85. DOI: 10.17223/19988605/68/8

24. **Li J., Su Z., Geng J., Yin Y.** Real-time detection of steel strip surface defects based on improved YOLO detection network. *IFAC-PapersOnLine*, 2018, Vol. 51, No. 21, Pp. 76–81. DOI: 10.1016/j.ifacol.2018.09.412

25. **Hussain M.** YOLO-v1 to YOLO-v8, the rise of YOLO and its complementary nature toward digital manufacturing and industrial defect detection. *Machines*, 2023, Vol. 11, No. 7, Art. no. 677. DOI: 10.3390/machines11070677

26. **Agamirov L.V., Agamirov V.L., Los V.S., Nosikov M., Toutova N.V.** Dataset annotation converter for training neural networks to detect defects on various surfaces. *2024 Intelligent Technologies and Electronic Devices in Vehicle and Road Transport Complex* (*TIRVED*), 2024, Pp. 1–4. DOI: 10.1109/TIRVED63561.2024.10769786

27. **Agamirov L.V., Agamirov V.L., Gaeva A.P., Vestyak K.V.** Ispol'zovanie neyronnykh setey dlya vyyavleniya defektov i povrezhdeniy v konstruktsiyakh mostov [The use of neural networks for detecting defects and damages in bridge structures]. *Tezisy dokladov XXX Mezhdunarodnogo simpoziuma im. A.G. Gorshkova.*

Dinamicheskie i tekhnologicheskie problemy mekhaniki konstruktsiy i sploshnykh sred [*Abstracts of the XXX International Symposium named after A.G. Gorshakov. Dynamic and Technological Problems of Mechanics of Structures and Continuous Media*], 2024, Vol. 1, Pp. 6–9.

28. **Khaush A.** Primenenie neyrosetey dlya resheniya zadach obnaruzheniya treshchin [Application of neural networks for crack detection]. *XV Molodezhnyy nauchnyy forum "Telekommunikatsii i informatsionnye tekhnologii – realii, vozmozhnosti, perspektivy". Sektsiya "Tsifrovaya ekonomika, upravlenie i biznes-tekhnologii"* [*XV Youth Scientific Forum "Telecommunications and Information Technologies - Realities, Possibilities, Prospects". Section "Digital Economy, Management and Business Technologies"*], 2024.

29. **Hatab M., Malekmohamadi H., Amira A.** Surface defect detection using YOLO network. *Intelligent Systems and Applications* (*IntelliSys 2020*), 2020, Vol. 1250, Pp. 505–515. DOI: 10.1007/978-3-030-55180-3_37

30. **Xie Y., Hu W., Xie S., He L.** Surface defect detection algorithm based on feature-enhanced YOLO. *Cognitive Computation*, 2023, Vol. 15, Pp. 565–579. DOI: 10.1007/s12559-022-10061-z

31. **Vu T.-T.-H., Pham D.-L., Chang T.-W.** A YOLO-based real-time packaging defect detection system. *Procedia Computer Science*, 2023, Vol. 217, Pp. 886–894. DOI: 10.1016/j.procs.2022.12.285

32. **Chen W., Huang Z., Mu Q., Sun Y.** PCB defect detection method based on Transformer-YOLO. *IEEE Access*, 2022, Vol. 10, Pp. 129480–129489. DOI: 10.1109/ACCESS.2022.3228206

33. **Zhao C., Shu X., Yan X., Zuo X., Zhu F.** RDD-YOLO: A modified YOLO for detection of steel surface defects. *Measurement*, 2023, Vol. 214, Art. no. 112776. DOI: 10.1016/j.measurement.2023.112776

34. **Liu M., Chen Y., Xie J., He L., Zhang Y.** LF-YOLO: A lighter and faster YOLO for weld defect detection of X-ray images. *IEEE Sensors Journal*, 2023, Vol. 23, No. 7, Pp. 7430–7439. DOI: 10.1109/JSEN.2023.3247006

35. **Du F.-J., Jiao S.-J.** Improvement of lightweight convolutional neural network model based on YOLO algorithm and its research in pavement defect detection. *Sensors*, 2022, Vol. 22, No. 9, Art. no. 3537. DOI: 10.3390/s22093537

36. **Yi W., Ma S., Li R.** Insulator and defect detection model based on improved YOLO-S. *IEEE Access*, 2023, Vol. 11, Pp. 93215–93226. DOI: 10.1109/ACCESS.2023.3309693

37. **Korzhov A.V., Surin V.A., Cheskidova M.A., Lonzinger P.V., Safonov V.I., Belov K.N.** Development of an algorithm for detecting defects in glass insulators based on computer vision using a neural network approach. *Bulletin of the South Ural State University. Series "Mathematics. Mechanics. Physics"*, 2024, Vol. 16, No. 4, Pp. 35–42. DOI: 10.14529/mmph240405

38. **Mukhamadiyarov R.A., Klyshnikov K.Yu., Koshelev V.A., Kutikhin A.G.** Potential of diagnostic electron microscopy of calcification, pathological neovascularization and elastolysis in combination with phenotyping of cell populations in large arteries. *Russian Journal of Cardiology*, 2024, Vol. 29, No. 8, Art. no. 5909. DOI: 10.15829/1560-4071-2024-5909

39. **Tulbure A.-A., Tulbure A.-A., Dulf E.-H.** A review on modern defect detection models using DCNNs – Deep convolutional neural networks. *Journal of Advanced Research*, 2022, Vol. 35, Pp. 33–48. DOI: 10.1016/j.jare.2021.03.015

40. **Wang T., Chen Y., Qiao M., Snoussi H.** A fast and robust convolutional neural network-based defect detection model in product quality control. *The International Journal of Advanced Manufacturing Technology*, 2018, Vol. 94, Pp. 3465–3471. DOI: 10.1007/s00170-017-0882-0

41. **Khanam R., Hussain M., Hill R., Allen P.** A comprehensive review of convolutional neural networks for defect detection in industrial applications. *IEEE Access*, 2024, Vol. 12, Pp. 94250–94295. DOI: 10.1109/ACCESS.2024.3425166

42. **Darabi A., Maldague X.** Neural network based defect detection and depth estimation in TNDE. *NDT & E International*, 2002, Vol. 35, No. 3, Pp. 165–175. DOI: 10.1016/S0963-8695(01)00041-X

43. **Singh S.A., Desai K.A.** Automated surface defect detection framework using machine vision and convolutional neural networks. *Journal of Intelligent Manufacturing*, 2023, Vol. 34, pp. 1995–2011. DOI: 10.1007/s10845-021-01878-w

44. **Xing J., Jia M.** A convolutional neural network-based method for workpiece surface defect detection. *Measurement*, 2021, Vol. 176, Art. no. 109185. DOI: 10.1016/j.measurement.2021.109185

INFORMATION ABOUT AUTHORS / СВЕДЕНИЯ ОБ АВТОРАХ

Vladimir L. Agamirov
Агамиров Владимир Левонович
E-mail: avhere@yandex.ru
ORCID: https://orcid.org/0000-0001-9181-7726

Levon V. Agamirov
Агамиров Левон Владимирович
E-mail: itno_agamirov@mail.ru
ORCID: https://orcid.org/0009-0009-6909-9399

Maksim A. Nosikov
Носиков Максим Александрович
E-mail: nosikovmaxim@yandex.ru

Natalia V. Toutova
Тутова Наталья Владимировна
E-mail: e-natasha@mail.ru
ORCID: https://orcid.org/0000-0002-2851-8472

Anton Khaush
Хауш Антон
E-mail: anton.khaush@gmail.com
ORCID: https://orcid.org/0009-0008-7929-5492

Submitted: 02.04.2025; Approved: 07.07.2025; Accepted: 10.08.2025.
Поступила: 02.04.2025; Одобрена: 07.07.2025; Принята: 10.08.2025.

Computing, Telecommunication and Control, 2025, Vol. 18, No. 3, Pp. 23–35.
Информатика, телекоммуникации и управление. 2025. Том 18, № 3. С. 23–35.

Research article
DOI: https://doi.org/10.18721/JCSTCS.18302
UDC 004.8

APPLICATION OF MACHINE LEARNING ALGORITHMS AND NEURAL NETWORKS FOR ANALYZING THE INFLUENCE OF DATA TYPE IN HATE SPEECH DETECTION

L.P. Mbele Ossiyi ✉, *P.D. Drobintsev*, *S.M. Ustinov*

Peter the Great St. Petersburg Polytechnic University,
St. Petersburg, Russian Federation

✉ lucprucell@gmail.com

Abstract. At present, communication has reached an unprecedented level of activity thanks to online social platforms that have overcome geographical and linguistic barriers. However, the shift to online communication is accompanied by the spread of hate speech, which negatively affects the social environment of these platforms. In the field of natural language processing, research is being conducted to develop models for detecting and classifying hate speech, aimed at improving the safety and quality of the online environment. However, many of these studies are based on commonly used datasets that turn out to be unbalanced and insufficiently adapted to the new grammatical features of hate speech. This article presents a comparative study of the effectiveness of machine and deep learning algorithms in detecting hate speech based on a synthetic dataset. Three separate experiments were conducted using original and synthetically perturbated data. The findings indicate that employing a synthetic dataset enhances the representation of extremely negative or infrequently encountered communication scenarios, contributing to their more effective detection. Deep learning algorithms demonstrated superior performance in all experiments. The top-performing models in the first and second experiments, both using zero-shot learning, yielded accuracies of 52.04% and 62.13%, respectively. The last experiment revealed that the BiGRU + fastText architecture outperformed other models, achieving an accuracy of 72.68%.

Keywords: sentiment analysis, emotion recognition in text, attention mechanism, embedding, CNN, LSTM, GRU

Citation: Mbele Ossiyi L.P., Drobintsev P.D., Ustinov S.M. Application of machine learning algorithms and neural networks for analyzing the influence of data type in hate speech detection. Computing, Telecommunications and Control, 2025, Vol. 18, No. 3, Pp. 23–35. DOI: 10.18721/JCSTCS.18302

Научная статья
DOI: https://doi.org/10.18721/JCSTCS.18302
УДК 004.8

ПРИМЕНЕНИЕ АЛГОРИТМОВ МАШИННОГО ОБУЧЕНИЯ И НЕЙРОННЫХ СЕТЕЙ ДЛЯ АНАЛИЗА ВЛИЯНИЯ ТИПА ДАННЫХ ПРИ ВЫЯВЛЕНИИ НЕНАВИСТНИЧЕСКИХ ВЫСКАЗЫВАНИЙ

Л.П. Мбеле Оссийи ✉ , *П.Д. Дробинцев, С.М. Устинов*

Санкт-Петербургский политехнический университет Петра Великого,
Санкт-Петербург, Российская Федерация

✉ lucprucell@gmail.com

Аннотация. В настоящее время общение достигло беспрецедентного уровня активности благодаря онлайн-социальным платформам, которые преодолели географические и языковые барьеры. Однако этот переход сопровождается распространением ненавистнических высказываний, которые негативно влияют на социальную среду этих платформ. В области обработки естественного языка ведутся исследования по разработке моделей для выявления и классификации ненавистнических высказываний, направленные на улучшение безопасности и качества онлайн-среды. Однако многие из этих исследований основаны на наборах данных, которые часто используются и оказываются несбалансированными и недостаточно адаптированными к новым грамматическим особенностям ненавистнических высказываний. В этой статье представлено сравнительное исследование эффективности алгоритмов машинного и глубокого обучения в выявлении ненавистнических высказываний на основе синтетического набора данных. Три отдельных эксперимента были проведены с использованием оригинальных и искусственно искаженных данных. Результаты показывают, что использование синтетического набора данных позволяет лучше представить крайне негативные или нечасто встречающиеся сценарии коммуникации, что способствует их более эффективному выявлению. Алгоритмы глубокого обучения продемонстрировали превосходную производительность во всех экспериментах. Лучшие модели в первом и втором экспериментах, основанные на «обучении без примеров», показали точность 52,04% и 62,13% соответственно. Последний эксперимент показал, что архитектура BiGRU + fastText превзошла другие модели, достигнув точности 72,68%.

Ключевые слова: анализ тональности текста, распознавание эмоций в тексте, механизм внимания, эмбеддинги, CNN, LSTM, GRU

Для цитирования: Mbele Ossiyi L.P., Drobintsev P.D., Ustinov S.M. Application of machine learning algorithms and neural networks for analyzing the influence of data type in hate speech detection // Computing, Telecommunications and Control. 2025. Т. 18, № 3. С. 23–35. DOI: 10.18721/JCSTCS.18302

Introduction

One of the challenges in modern online communication environments, such as forums, blogs and social media, is hate speech. Directed at individuals or groups of people, it is often based on characteristics such as skin color, religion, gender, nationality and others. The level of toxicity on the internet, measured by the amount of hate speech, has increased since the beginning of the COVID-19 pandemic in 2020 [1, 2], when a significant portion of social interactions shifted to online platforms. A number of international organizations, such as UNESCO, reported an increase in hate speech and conspiracy theories against specific communities on social media. According to a UNESCO/Ipsos report conducted in 2023 in 16 countries, 67% of internet users have encountered toxic messages and comments.

To create safe digital spaces where hate speech will be automatically detected, extensive research has been conducted [3–6]. An analysis of these studies suggests that the use of machine learning algorithms and neural networks for hate speech detection is becoming critical in modern conditions. For instance, multilayer neural network architectures enable the learning of hierarchical data representations, which is highly valuable for understanding the context and nuances of human language. Hate speech detection relies on two main approaches: supervised learning and unsupervised learning. In the context of this work and the available dataset, a supervised learning approach will be employed, where models are trained on labeled datasets containing examples of both hate speech and ordinary statements.

The work [7] explores a research direction that has not yet been widely covered in scientific literature – namely, the use of synthetic data as a non-traditional approach to overcoming the difficulties associated with collecting and annotating real data. Synthetic datasets enable the generation of a wide range of scenarios and hate speech instances that may be underrepresented in real datasets. In [8], the developed a method to maintain baseline model performance in case of future perturbations, instead of training and retraining the model on data with introduced perturbations as a mitigation method. However, this method is effective only for perturbations that preserve text semantics and exclude those that alter semantics, which are prevalent in [7]. Furthermore, this approach is suitable only for large language models with numerous parameters and high training costs. Experiments in [9] demonstrated that within fine-tuning, the performance of large language models improved by 7–19% partly due to the use of a specific synthetic dataset from [7]. Similarly, the work [10], also based on [7], focused on the automatic detection of dehumanizing statements and achieved promising results. However, it relied exclusively on large language models with extensive parameters. While the studies aim to enhance classifier performance using synthetic data with introduced perturbations, none of them investigate the impact of data type on the performance and robustness of machine learning and deep learning classifiers that do not require a large number of parameters.

Our work continues the line of research initiated in [7]. The central idea is to evaluate the influence of data type (original and synthetically perturbated) on classifier performance in binary classification, where the input is one-dimensional textual data.

The key contributions of this work are as follows:

- Utilization of a synthetic dataset.
- Application of binary classification through the training and testing of various classifiers, including traditional machine learning models (Linear Support Vector Classifier, Logistic Regression, Stochastic Gradient Descent, XGBoost) and deep learning models (CNN, LSTM, GRU, BiGRU, BiGRU + CNN) for hate speech detection.
- Investigation of the impact of static context-independent embeddings models (fastText and GloVe) on classifier performance.
- Examination of how original and synthetically perturbated data influence classifier performance, as this issue has not yet been sufficiently addressed in scientific literature.

Experimental Framework

Dataset

Hate speech detection typically involves the use of various benchmark datasets (e.g., *Wikipedia Detox*, 2016; *Jigsaw Toxic Comment Classification*, 2018; *SemEval-2019 Task 5*) for heuristic studies. However, it should be noted that most of these datasets, although some are relatively large and of high quality, gradually become outdated and lose relevance over time. Therefore, in this work we employ the *Dynamically Generated Hate Speech Dataset* from Vidgen et al. (2021), which has not yet been widely utilized or extensively discussed in scientific literature.

Dataset description

The *Dynamically Generated Hate Speech Dataset* comprises approximately 40000 entries (~10000 per round), generated and annotated by trained annotators across four rounds of dynamic data creation

using a human-in-the-loop process. The dataset is balanced, with hate speech instances constituting 54%. All entries are labeled as either *hateful* or *non-hateful*. For entries labeled as *hateful*, secondary annotations are provided, specifying the *type* and *target* of hate speech.

The dataset contains both original and synthetically perturbated data (~15000 complex perturbations). The original data consists of unmodified instances (e.g., without altered annotations). The synthetically perturbated data, available in the version used in our work, comprise statements that were initially considered *non-hateful* but, after modification (typically syntactic), were re-annotated as *hateful*.

As noted in [7, 11], perturbations are generally described as sufficient manipulations of the original text to alter the label (e.g., from *hateful* to *non-hateful*). Such perturbations can significantly change the meaning of a sentence and, consequently, the model's predictions. According to [11], perturbation-based methods applied to text remain in their early stages. Nevertheless, recent studies [8, 12–15] have proposed several semantic-preserving and semantic-altering perturbation techniques. When applied to the text, these techniques allow models for developing robustness against future or adversarial perturbations that might otherwise cause misclassification.

Dataset analysis

The dataset consists of 12 columns, including *label*, *type* and *target*.

The *label* column takes two values: *hateful* or *non-hateful*, indicating whether a given utterance constitutes hate speech.

The *type* column provides an additional annotation for hateful utterances. If an utterance is labeled *hateful*, the *type* column can take one of five values: *animosity*, *derogation*, *dehumanization*, *threatening* or *support*.

The *target* column specifies the group subjected to hate speech. Examples include *wom* (women), *bla* (black people) and *mus* (muslims). The *target* column contains more than 400 unique values. For example, the phrase "*There are so many black women at my workplace, it really annoys me*" in the dataset is labeled as *hateful*, with *type* = *animosity* and *target* = *bla.wom* (referring to black women).

The distribution of entries across the *label* and *type* columns is illustrated in Fig. 1. The presence of the "*not given*" category in relation to *type* is explained by the absence of hate-type annotation in round 1. Among the *type* categories, *derogation* (utterances that explicitly attack, demonize, humiliate or insult a group) is the most frequent, while *support* (utterances that praise or endorse events, organizations, actions that propagate hate) is the least frequent.

Fig. 2 presents the distribution of words and characters across the *label* column. The maximum utterance length does not exceed 600 characters or 150 words. Both *hateful* and *non-hateful* labels show a similar distribution in terms of word and character length. However, there is a notable difference: approximately 28% of *non-hateful* utterances contain words with lengths between 1 and 25 characters, compared to ~35% of *hateful* utterances.

Data preprocessing

The data preprocessing procedure was designed to reduce vocabulary size without removing essential content. A smaller vocabulary not only decreases the memory required for analysis, but also enhances the reliability of estimated word parameters. In this work, standard preprocessing operations were applied, albeit with some modifications. As noted in [16, 17], these operations included lowercasing, tokenization, punctuation handling, stop-word removal, part-of-speech (POS) tagging (to improve semantic understanding of text and facilitate more accurate lemmatization) and lemmatization.

However, to provide classifiers with a more favorable learning environment, we followed the approach of [18] and replaced contracted negative forms with their full equivalents. In addition, emojis were substituted with their corresponding semantic meanings. Furthermore, as part of the preprocessing pipeline, the maximum length of individual posts was limited to 100 words and 500 characters, respectively, for subsequent operations.

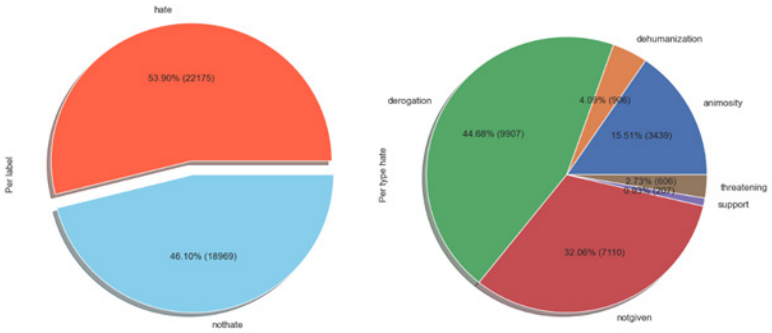


Fig. 1. Distribution of the dataset across *labels* and hate *types*

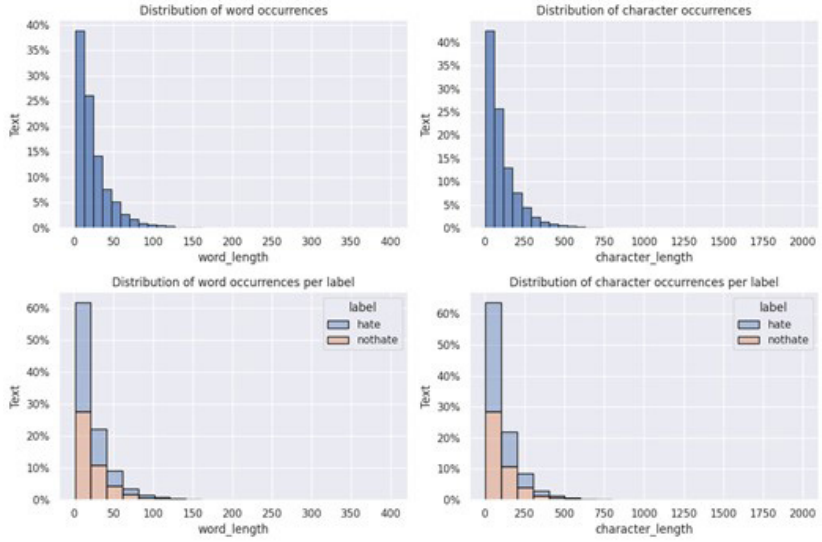


Fig. 2. Distribution of words and characters by *label*

Models

The primary focus of this work was on deep neural network architectures. For comparative analysis, several traditional machine learning methods were employed as baseline models, including Linear Support Vector Classifier (Linear SVC), Stochastic Gradient Descent (SGD), Logistic Regression (LR) and Extreme Gradient Boosting (XGBoost).

The rationale for selecting these algorithms is as follows: Linear SVC, a specific case of Support Vector Machines, assumes a linear decision boundary (effective when classes are well-separated in feature space, as in our case), handles high-dimensional spaces efficiently and thereby mitigates overfitting. SGD serves as an optimization algorithm that updates model parameters incrementally, allowing for faster convergence compared to batch gradient descent. LR is effective in tasks where the relationship between features and class labels can be approximated linearly, as demonstrated in our experiments. Finally, XGBoost excels at handling missing values, prevents overfitting and can capture complex feature interactions and non-linear relationships.

Neural networks

In this work, we employed five neural network architectures: Convolutional Neural Networks (CNNs) [19, 20], Long Short-Term Memory networks (LSTM) [19], Gated Recurrent Units (GRU),

Bidirectional GRU (BiGRU) and a hybrid CNN + BiGRU model. All these models were implemented with word embedding methods (fastText and GloVe) [19]. To enhance the effectiveness of the neural networks, we hypothesized that performance could be improved using attention mechanisms and pooling operations within the network architecture. The attention mechanism assigns different weights to sequence elements, allowing models to focus on specific parts of the input data, thereby improving their ability to generate accurate and contextually relevant predictions. Pooling, in turn, reduces computational complexity and facilitates the handling of long sequences.

The choice of deep neural network architectures is motivated by the fact that traditional machine learning methods largely rely on manual feature engineering, whereas deep learning models are capable of learning abstract data representations and automatically extracting features [18, 21].

CNN

CNNs were originally developed for computer vision tasks and are highly effective in image classification [22, 23]. However, CNNs have also demonstrated strong applicability in natural language processing (NLP), particularly for text classification tasks [24, 25]. While CNNs are primarily designed for processing data represented as matrices rather than sequences, they can outperform recurrent neural networks (RNNs) [22], especially in their ability to capture higher-level features. The role of a CNN layer is to extract meaningful substructures that are useful for solving the overall prediction task. In this work, we implemented a CNN with a global max pooling mechanism to reduce computational complexity and the number of outputs [26].

LSTM and GRU

LSTM and GRU networks are types of recurrent neural networks [24, 26]. In text classification tasks, each LSTM or GRU block processes both the embedding vector of the current word and the output of the previous block, recursively accumulating information from all other words in the text. Unlike traditional RNNs, LSTM and GRU networks are specifically designed to overcome the problems of long-term dependencies and the issues of exploding and vanishing gradients [18, 26].

These models employ more advanced mechanisms for computing hidden states at each step to mitigate gradient-related problems [27]. Both LSTM and GRU incorporate gating mechanisms that enable selective retention or forgetting of information from previous inputs. LSTMs feature a more complex structure consisting of four components: input gate, forget gate, cell state and output gate. In contrast, GRUs represent a generalized approach, with LSTMs being a special case [27]. GRUs typically require fewer filters and fewer computational operations than LSTMs [26, 27].

BiGRU and BiGRU + CNN

The concept of bidirectionality was applied in cases where the meaning of certain words depends on subsequent words in the sentence. This is particularly relevant for synthetically perturbated data, where adding a word at the end of a sentence may alter its entire meaning. In addressing this issue, a choice was made between BiGRU and BiLSTM. Ultimately, BiGRU was selected, primarily due to the simpler architecture and faster training of GRUs, as well as their ability to be effectively trained to preserve information over long sequences without loss of temporal dependencies [3]. To further improve key aspects of our work – such as addressing CNN limitations in capturing inter-word semantics, enhancing prediction accuracy, modeling complex relationships, extracting features and patterns, managing long-range dependencies and ensuring robustness to noise and outliers – we adopted a hybrid approach [23, 28] that combines CNN and BiGRU. We hypothesized that this hybrid architecture leverages the strengths of both models while compensating for their respective weaknesses.

Experimental Setup

In [16], it was demonstrated that word embedding methods (such as fastText and GloVe), which are most used in deep learning models, can also yield strong results when applied within machine learning frameworks. Following this line of reasoning, we adopted the same approach in our baseline machine learning experiments. Alongside word embedding methods, we also employed the TF-IDF bag-of-words

model [19, 29] to extract features from textual sequences. The same word embedding techniques were consistently applied across all deep learning models.

The performance of classifiers was evaluated using several metrics, including accuracy, macro-precision, macro-recall and macro-F1 score. Additionally, as in [18], to better handle the influence of true negatives – which are of limited utility in detecting hate speech – we incorporated the area under the precision-recall curve (AUC-PRC), in addition to the area under the receiver operating characteristic curve (AUC-ROC).

In this work, we focused on binary classification (hate / non-hate) and trained and evaluated the models across three experimental settings:

• Models were trained exclusively on synthetically perturbated data, but developed and tested on original data.

• Models were trained solely on original data, but developed and tested on synthetically perturbated data.

• Models were trained, developed and tested on a combination of both synthetically perturbated and original data.

The corpus was split into training, cross-validation and test sets in accordance with the nature of the data (original and synthetically perturbated).

For experimentation, we employed the Google Colab environment, which supports TensorFlow (version 2.15.0) and provides access to fast, high-performance computing resources such as GPU and TPU. The programming language used was Python 3.10, and computations were run on an MSI Katana 17 (i7-12650H, 16 GB RAM). The source code is publicly available[1].

The table below presents the configuration parameters for all classifiers. All parameters were obtained through hyperparameter tuning.

Table 1

Hyperparameter settings of the baseline models

Models	Parameters
Linear Support Vector Classification	*C = 0.1, max_iter = 1000*
Logistic Regression	*C = 1, penalty = 'l2', solver = 'liblinear', max_iter = 10000*
Stochastic Gradient Descent	*loss:'hinge', alpha: 0.0001, penalty: 'l2'*
Extreme GBOOST	*learning_rate = 0.1, n_estimators = 100, max_depth = 5, min_child_weight = 1, gamma = 0, subsample = 0.8, colsample_bytree = 0.8, objective = 'binary: logistic', nthread = 4, scale_pos_weight = 1, seed = 27*
CNN	*filters = 512, kernel_size = 6, dropout_rate = 0.5, dense_units = 512, emb_dim = = 300, optimizer = 'Adagrad', learning_rate = 0.00001*
LSTM	*lstm_units = 64, dense_units = 512, k_regularizer = 0.001, dropout_rate = 0.3, recurrent_dropout = 0.0, emb_dim = 300, optimizer = 'Adam', learning_rate = 0.001*
GRU	*gru_units = 64, dropout_rate = 0.5, k_regularizer = 0.00001, recurrent_dropout = = 0.0, emb_dim = 300, optimizer = 'Adam', learning_rate = 0.001*
BiGRU	*gru_units = 256, k_regularizer = 0.00001, dropout_rate = 0.5, recurrent_dropout = = 0.0, emb_dim = 300, optimizer = 'Adam', learning_rate = 0.001*
BiGRU + CNN	*filters = 16, kernel_size=6, dropout_rate= 0.5, dense_units= 64, gru_units = 256, k_regularizer = 0.00001, recurrent_dropout= 0.0, emb_dim = 300, optimizer = = 'Adam', learning_rate= 0.003*

[1] GitHub – LucasMbele/Hate-speech-synthetic-dataset: In this repository, we train and test some classifiers on original and perturbated data from a synthetical dataset for hate classification tasks in binary and multiclass case. Available: https://github.com/LucasMbele/Hate-speech-synthetic-dataset (Accessed 12.09.2025)

Results

Experiment 1: Training on original data, development and testing on synthetically perturbated data

The data were split as follows: 25813 for the training set, 10332 for the development set and 4429 for the test set.

Based on the results presented in Fig. 3, among all machine learning methods, the logistic regression algorithm combined with fastText generally outperforms and achieves the best results in terms of accuracy (**45.3**%), F1-score (**44.2**%) and AUC-ROC (**45.2**%). It is worth noting that the differences between the results of other machine learning algorithms and those of logistic regression are negligible.

It is evident that neural networks outperform machine learning algorithms, as expected. BiGRU + CNN + GloVe (**52.04**% accuracy, **51.09**% F1-score) achieves better performance than other models; however, its loss function value is considerably high. Measuring the difference between the model's predictions and the actual values, the loss function plays a crucial role in the efficiency of neural networks.

Models	Folds	Accuracy score			F1-score			ROC score		
		Training set	Dev	Test	Training set	Dev	Test	Training set	Dev	Test
Logistic Regression + TF-IDF	CV Fold 1	0,726			0,71			0,777		
	CV Fold 2	0,65			0,607			0,71		
	CV Fold 3	0,718			0,683			0,805		
	CV Average	**0,698**	**0,456**	**0,436**	**0,667**	**0,429**	**0,409**	**0,764**	**0,455**	**0,436**
Logistic Regression + FastText	CV Fold 1	0,619			0,576			0,684		
	CV Fold 2	0,634			0,622			0,655		
	CV Fold 3	0,689			0,667			0,676		
	CV Average	**0,647**	**0,47**	**0,453**	**0,622**	**0,46**	**0,442**	**0,672**	**0,469**	**0,452**
Logistic Regression + GloVE	CV Fold 1	0,603			0,54			0,658		
	CV Fold 2	0,629			0,601			0,633		
	CV Fold 3	0,686			0,645			0,644		
	CV Average	**0,639**	**0,453**	**0,448**	**0,595**	**0,433**	**0,425**	**0,645**	**0,452**	**0,447**
Linear SVM + TF-IDF	CV Fold 1	0,726			0,708			0,777		
	CV Fold 2	0,649			0,598			0,71		
	CV Fold 3	0,714			0,676			0,806		
	CV Average	**0,696**	**0,452**	**0,437**	**0,661**	**0,417**	**0,402**	**0,764**	**0,45**	**0,436**
Linear SVM + FastText	CV Fold 1	0,62			0,572			0,682		
	CV Fold 2	0,634			0,62			0,655		
	CV Fold 3	0,691			0,666			0,681		
	CV Average	**0,648**	**0,467**	**0,452**	**0,619**	**0,455**	**0,44**	**0,673**	**0,467**	**0,452**
Linear SVM + GloVE	CV Fold 1	0,599			0,53			0,654		
	CV Fold 2	0,627			0,597			0,632		
	CV Fold 3	0,686			0,639			0,647		
	CV Average	**0,637**	**0,453**	**0,447**	**0,589**	**0,431**	**0,419**	**0,644**	**0,452**	**0,446**
Stochastic Gradient Descent + TF-IDF	CV Fold 1	0,716			0,683			0,783		
	CV Fold 2	0,635			0,554			0,71		
	CV Fold 3	0,7			0,646			0,813		
	CV Average	**0,684**	**0,45**	**0,448**	**0,628**	**0,354**	**0,356**	**0,769**	**0,448**	**0,447**
Stochastic Gradient Descent + FastText	CV Fold 1	0,619			0,587			0,669		
	CV Fold 2	0,63			0,615			0,653		
	CV Fold 3	0,669			0,614			0,686		
	CV Average	**0,639**	**0,464**	**0,443**	**0,605**	**0,458**	**0,436**	**0,669**	**0,463**	**0,442**
Stochastic Gradient Descent + GloVE	CV Fold 1	0,603			0,511			0,643		
	CV Fold 2	0,613			0,552			0,632		
	CV Fold 3	0,693			0,504			0,663		
	CV Average	**0,636**	**0,452**	**0,442**	**0,522**	**0,401**	**0,386**	**0,646**	**0,45**	**0,441**
XGBOOST + TF-IDF	CV Fold 1	0,672			0,641			0,711		
	CV Fold 2	0,626			0,569			0,662		
	CV Fold 3	0,703			0,658			0,726		
	CV Average	**0,667**	**0,417**	**0,42**	**0,623**	**0,373**	**0,377**	**0,7**	**0,416**	**0,42**
XGBOOST + FastText	CV Fold 1	0,662			0,652			0,714		
	CV Fold 2	0,633			0,623			0,688		
	CV Fold 3	0,695			0,678			0,749		
	CV Average	**0,663**	**0,393**	**0,391**	**0,651**	**0,388**	**0,386**	**0,717**	**0,392**	**0,39**
XGBOOST + GloVE	CV Fold 1	0,658			0,648			0,714		
	CV Fold 2	0,629			0,617			0,675		
	CV Fold 3	0,697			0,678			0,746		
	CV Average	**0,661**	**0,391**	**0,406**	**0,648**	**0,387**	**0,401**	**0,712**	**0,391**	**0,406**

Models	Folds	Accuracy score				Precision score		Recall score		Test		
		CV	Loss	Test	Loss	CV	Test	CV	Test	AUPRC score	AUC score	F1-score
CNN + FastText	CV Fold 1	0,5096	0,7207			0,5083		0,3897				
	CV Fold 2	0,503	0,7286			0,4995		0,3057				
	CV Fold 3	0,5028	0,7311			0,4992		0,2919				
	CV Average/ Best Loss	***0,5051***	***0,7207***	***51,46%***	0,7205	**0,5023**	**51,62%**	**0,3291**	**51,52%**	**0,5142**	**0,5183**	**50,76%**
CNN + GloVE	CV Fold 1	0,5069	0,6988			0,5026		0,7048				
	CV Fold 2	0,4994	0,6977			0,4952		0,4014				
	CV Fold 3	0,5031	0,6996			0,4996		0,2627				
	CV Average/ Best Loss	**0,5031**	**0,6977**	**50,62%**	0,6993	**0,4991**	**50,60%**	**0,4563**	**50,51%**	**0.4952**	**0.5017**	**48,74%**
LSTM + FastText	CV Fold 1	0,4441	0,9449			0,4054		0,2555				
	CV Fold 2	0,5022	0,938			0,4991		0,582				
	CV Fold 3	0,4875	0,9916			0,4864		0,5688				
	CV Average/ Best Loss	**0,4779**	**0,938**	**49,97%**	0,9471	**0,4636**	**49,92%**	**0,46877**	**49,92%**	**0.4834**	**0.4861**	**49.58%**
LSTM + GloVE	CV Fold 1	0,4654	0,9853			0,4571		0,4072				
	CV Fold 2	0,5109	0,9413			0,51		0,3911				
	CV Fold 3	0,5063	0,9342			0,504		0,3843				
	CV Average/ Best Loss	**0,4942**	**0,9342**	**50,87%**	0,941	**0,4904**	**50,99%**	**0,3942**	**50,94%**	**0,5104**	**0,5074**	**50,21%**
GRU + FastText	CV Fold 1	0,4714	0,9326			0,459		0,3589				
	CV Fold 2	0,4808	0,9175			0,4772		0,4739				
	CV Fold 3	0,4867	0,992			0,4838		0,4971				
	CV Average/ Best Loss	**0,47963**	**0,9175**	**48,97%**	0,9878	**0,4733**	**48,97%**	**0,4433**	**48,97%**	**0,4938**	**0,4915**	**48,96%**
GRU + GloVE	CV Fold 1	0,4977	0,8744			0,4934		0,4215				
	CV Fold 2	0,4562	0,9203			0,4502		0,4289				
	CV Fold 3	0,4678	0,9602			0,4597		0,4078				
	CV Average/ Best Loss	**0,4739**	**0,8744**	**50,08%**	0,877	**0,4678**	**50,12%**	**0,4194**	**50,12%**	**0,4898**	**0,4866**	**49,82%**
BIGRU + FastText	CV Fold 1	0,4838	0,8684			0,4708		0,3161				
	CV Fold 2	0,4828	0,9147			0,4723		0,3525				
	CV Fold 3	0,494	0,946			0,4906		0,4875				
	CV Average/ Best Loss	**0,4869**	**0,8684**	**48,66%**	0,9556	**0,4779**	**48,65%**	**0,3854**	**48,65%**	**0,492**	**0,4803**	**48,65%**
BIGRU + GloVE	CV Fold 1	0,4991	0,9206			0,4944		0,373				
	CV Fold 2	0,479	0,959			0,4465		0,2042				
	CV Fold 3	0,4833	1,023			0,4783		0,4447				
	CV Average/ Best Loss	**0,48713**	**0,9206**	**50,96%**	0,913	**0,4731**	**51,08%**	**0,34063**	**51,03%**	**0,5011**	**0,5091**	**50,31%**
BIGRU + CNN + FastText	CV Fold 1	0,4383	0,9533			0,4101		0,2983				
	CV Fold 2	0,4691	1,087			0,4106		0,158				
	CV Fold 3	0,4666	1,1431			0,4174		0,1865				
	CV Average/ Best Loss	**0,458**	**0,9533**	**46,67%**	1,0986	**0,4127**	**45,10%**	**0,2143**	**46,84%**	**0,4407**	**0,4412**	**41,58%**
BIGRU + CNN + GloVE	CV Fold 1	0,4775	1,0201			0,4021		0,1064				
	CV Fold 2	0,5227	1,0387			0,527		0,3825				
	CV Fold 3	0,5099	1,1369			0,5091		0,3693				
	CV Average/ Best Loss	**0,5034**	**1,0201**	**52,04%**	1,0402	**0,4794**	**52,31%**	**0,2861**	**52,12%**	**0,5053**	**0,5159**	**51,09%**

Fig. 3. Performance of machine learning and deep learning models in Experiment 1

Models	Folds	Accuracy score			F1-score			ROC score		
		Training set	Dev	Test	Training set	Dev	Test	Training set	Dev	Test
Logistic Regression + TF-IDF	CV Fold 1	0,686			0,686			0,741		
	CV Fold 2	0,651			0,645			0,716		
	CV Fold 3	0,58			0,556			0,65		
	CV Average	**0,639**	**0,454**	**0,454**	**0,629**	**0,448**	**0,449**	**0,702**	**0,448**	**0,449**
Logistic Regression + FastText	CV Fold 1	0,627			0,627			0,661		
	CV Fold 2	0,647			0,647			0,697		
	CV Fold 3	0,579			0,576			0,609		
	CV Average	**0,618**	**0,465**	**0,46**	**0,617**	**0,461**	**0,456**	**0,656**	**0,462**	**0,456**
Logistic Regression + GloVE	CV Fold 1	0,637			0,637			0,692		
	CV Fold 2	0,627			0,626			0,667		
	CV Fold 3	0,558			0,555			0,577		
	CV Average	**0,607**	**0,472**	**0,477**	**0,606**	**0,465**	**0,469**	**0,645**	**0,465**	**0,469**
Linear SVM + TF-IDF	CV Fold 1	0,685			0,685			0,738		
	CV Fold 2	0,649			0,642			0,714		
	CV Fold 3	0,577			0,551			0,648		
	CV Average	**0,637**	**0,453**	**0,453**	**0,626**	**0,448**	**0,45**	**0,7**	**0,448**	**0,45**
Linear SVM + FastText	CV Fold 1	0,624			0,624			0,661		
	CV Fold 2	0,647			0,647			0,698		
	CV Fold 3	0,578			0,575			0,607		
	CV Average	**0,616**	**0,462**	**0,458**	**0,615**	**0,458**	**0,454**	**0,655**	**0,459**	**0,455**
Linear SVM + GloVE	CV Fold 1	0,638			0,638			0,691		
	CV Fold 2	0,627			0,626			0,666		
	CV Fold 3	0,554			0,551			0,574		
	CV Average	**0,606**	**0,475**	**0,479**	**0,605**	**0,468**	**0,47**	**0,644**	**0,468**	**0,47**
Stochastic Gradient Descent + TF-IDF	CV Fold 1	0,674			0,67			0,741		
	CV Fold 2	0,617			0,577			0,72		
	CV Fold 3	0,548			0,488			0,659		
	CV Average	**0,613**	**0,434**	**0,446**	**0,578**	**0,433**	**0,444**	**0,707**	**0,447**	**0,459**
Stochastic Gradient Descent + FastText	CV Fold 1	0,619			0,664			0,664		
	CV Fold 2	0,648			0,694			0,694		
	CV Fold 3	0,564			0,616			0,616		
	CV Average	**0,610**	**0,444**	**0,443**	**0,658**	**0,443**	**0,442**	**0,658**	**0,452**	**0,451**
Stochastic Gradient Descent + GloVE	CV Fold 1	0,648			0,503			0,699		
	CV Fold 2	0,627			0,59			0,664		
	CV Fold 3	0,549			0,557			0,589		
	CV Average	**0,608**	***0,481***	***0,485***	**0,55**	***0,47***	***0,473***	**0,651**	***0,47***	***0,473***
XGBOOST + TF-IDF	CV Fold 1	0,709			0,707			0,764		
	CV Fold 2	0,678			0,676			0,731		
	CV Fold 3	0,621			0,61			0,663		
	CV Average	***0,669***	**0,457**	**0,46**	***0,664***	**0,457**	**0,46**	***0,719***	**0,463**	**0,466**
XGBOOST + FastText	CV Fold 1	0,611			0,611			0,659		
	CV Fold 2	0,618			0,617			0,66		
	CV Fold 3	0,558			0,552			0,581		
	CV Average	**0,596**	**0,379**	**0,392**	**0,593**	**0,378**	**0,39**	**0,633**	**0,379**	**0,392**
XGBOOST + GloVE	CV Fold 1	0,636			0,684			0,684		
	CV Fold 2	0,617			0,658			0,658		
	CV Fold 3	0,569			0,59			0,59		
	CV Average	**0,607**	**0,421**	**0,427**	**0,644**	**0,419**	**0,424**	**0,644**	**0,42**	**0,424**

Models	Folds	Accuracy score				Precision score		Recall score		Test		
		CV	Loss	Test	Loss	CV	Test	CV	Test	AUPRC score	AUC score	F1-score
CNN + FastText	CV Fold 1	0,4996	0,728			0,449		0,5949				
	CV Fold 2	0,5024	0,7244			0,4482		0,5576				
	CV Fold 3	0,5035	0,7214			0,4452		0,5132				
	CV Average/ Best Loss	***0,5018***	***0,7214***	***50,72%***	**0,72**	**0,447**	**50,93%**	**0,555**	**50,95%**	**0,439**	**0,507**	**50,61%**
CNN + GloVE	CV Fold 1	0,503	0,6986			0,3934		0,2353				
	CV Fold 2	0,4887	0,6999			0,3976		0,3106				
	CV Fold 3	0,4826	0,7009			0,4027		0,3596				
	CV Average/ Best Loss	**0,4914**	**0,6986**	**50,05%**	**0,7**	**0,3979**	**46,91%**	**0,3018**	**47,11%**	**0,402**	**0,449**	**45,42%**
LSTM + FastText	CV Fold 1	0,5351	0,9364			0,4744		0,5066				
	CV Fold 2	0,5729	0,905			0,517		0,4747				
	CV Fold 3	0,5592	0,9327			0,5001		0,5289				
	CV Average/ Best Loss	**0,5557**	**0,905**	**57,39%**	**0,91**	**0,4972**	**56,52%**	**0,5034**	**56,45%**	**0,494**	**0,571**	**56,47%**
LSTM + GloVE	CV Fold 1	0,4849	0,8501			0,432		0,5352				
	CV Fold 2	0,5487	0,8494			0,4874		0,4595				
	CV Fold 3	0,5429	0,8675			0,4844		0,5733				
	CV Average/ Best Loss	**0,5255**	**0,8494**	**54,79%**	**0,85**	**0,4679**	**53,94%**	**0,5227**	**53,92%**	**0,474**	**0,545**	**53,92%**
GRU + FastText	CV Fold 1	0,5022	0,8773			0,4339		0,4241				
	CV Fold 2	0,5163	0,8892			0,4535		0,4746				
	CV Fold 3	0,5468	0,8438			0,486		0,4886				
	CV Average/ Best Loss	**0,5218**	**0,8438**	**55,58%**	**0,9**	**0,4578**	**54,47%**	**0,4624**	**54,35%**	**0,468**	**0,55**	**54,33%**
GRU + GloVE	CV Fold 1	0,5451	0,8036			0,478		0,3465				
	CV Fold 2	0,5233	0,4549			0,4549		0,4112				
	CV Fold 3	0,5629	0,5045			0,5045		0,4703				
	CV Average/ Best Loss	**0,5438**	**0,4549**	**55,73%**	**0,81**	**0,4791**	**54,88%**	**0,4093**	**54,84%**	**0,466**	**0,554**	**54,85%**
BIGRU + FastText	CV Fold 1	0,5374	0,9361			0,4756		0,4821				
	CV Fold 2	0,5035	0,9822			0,4435		0,4954				
	CV Fold 3	0,5341	1,0659			0,4639		0,3659				
	CV Average/ Best Loss	**0,5250**	**0,9361**	**53,62%**	**0,94**	**0,461**	**53,13%**	**0,4478**	**53,16%**	**0,46**	**0,528**	**53,11%**
BIGRU + GloVE	CV Fold 1	0,5286	0,8264			0,465		0,4611				
	CV Fold 2	0,5642	0,8131			0,5066		0,4355				
	CV Fold 3	0,5219	0,8792			0,4545		0,4218				
	CV Average/ Best Loss	**0,53823**	**0,8131**	**56,46%**	**0,81**	**0,4754**	**55,34%**	**0,43947**	**55,17%**	**0,476**	**0,557**	**55,15%**
BIGRU + CNN + FastText	CV Fold 1	0,55	1,0292			0,4836		0,4306				
	CV Fold 2	0,5867	1,0314			0,5318		0,4576				
	CV Fold 3	0,6197	0,9674			0,57		0,5298				
	CV Average/ Best Loss	**0,58547**	**0,9674**	**62,13%**	**0,98**	**0,5285**	**61,29%**	**0,4727**	**61,29%**	**0,5377**	**0,6054**	**61,33%**
BIGRU + CNN + GloVE	CV Fold 1	0,528	0,7952			0,4445		0,3185				
	CV Fold 2	0,5406	0,8138			0,4699		0,3948				
	CV Fold 3	0,4903	0,8455			0,4219		0,4476				
	CV Average/ Best Loss	**0,5196**	**0,7952**	**55,75%**	**0,98**	**0,4454**	**54,77%**	**0,3870**	**54,59%**	**0,47**	**0,539**	**54,48%**

Fig. 4. Performance of machine learning and deep learning models in Experiment 2

The higher the loss function value, the harder it is for the model to make accurate predictions, thereby indicating the need for further improvements. In the experiment conducted, the loss function reached **1.0201**. CNN models (CNN + fastText and CNN + GloVe, as reported in [4]), particularly due to their robust feature extraction capabilities, show a clear improvement in the loss function (0.6993 with GloVe), comparable accuracy (**51.46**% with fastText) and superior results in terms of AUPRC-score (**51.42**% with fastText) and AUC-ROC score (**51.83**% with fastText).

The perturbations introduced into the validation and test datasets proved difficult for the models to learn, resulting in poor performance.

Experiment 2: Training on synthetically perturbated data, development and testing on original data

The data were split as follows: 14761 for the training set, 19359 for the development set and 6454 for the test set.

As shown in Fig. 4, neural networks significantly outperform machine learning models. Among machine learning algorithms, XGBoost and SGD stand out. XGBoost combined with TF-IDF achieved the best results on the training set, while SGD with GloVe obtained the best results on the test set (48.5%

accuracy, which is +3.2% higher than the accuracy of the best algorithm in Experiment 1; 47.3% F1-score, which is +3.1% higher than the F1-score of the best algorithm in Experiment 1; and 47.3% AUC-ROC score, which is +2.1% higher than the AUC-ROC score of the best algorithm in Experiment 1).

BiGRU + CNN + fastText, as in Experiment 1, outperformed all other algorithms, achieving **62.13**% accuracy, **61.29**% precision, **61.33**% recall, **61.33**% F1-score, **53.77**% AUPRC and **60.54**% AUC-ROC. Training models on synthetically perturbated data and testing them on original data substantially improved the performance of neural networks compared to Experiment 1. Overall, an improvement of +**10.9**% in accuracy and +**10.24**% in F1-score was observed when comparing the best model from Experiment 2 to the best model from Experiment 1.

Experiment 3: Training, development, and testing on both original and synthetically perturbated data

The data were split as follows: 19475 for the training set, 16230 for the development set and 4869 for the test set.

As shown in Fig. 5, among all machine learning algorithms, XGBoost (combined with TF-IDF), as in Experiment 2, achieved the best performance (**65.8**% F1-score, **65.9**% AUC-ROC). Logistic regression combined with TF-IDF achieved the highest accuracy (**66.2**%).

Models	Folds	Accuracy score			F1-score			ROC score		
		Training set	Dev	Test	Training set	Dev	Test	Training set	Dev	Test
Logistic Regression + TF-IDF	CV Fold 1	0,641			0,623			0,716		
	CV Fold 2	0,654			0,635			0,729		
	CV Fold 3	0,646			0,628			0,717		
	CV Average	**0,647**	**0,656**	***0,662***	**0,629**	**0,643**	**0,651**	***0,721***	**0,645**	**0,652**
Logistic Regression + FastText	CV Fold 1	0,608			0,587			0,641		
	CV Fold 2	0,612			0,591			0,653		
	CV Fold 3	0,612			0,594			0,648		
	CV Average	**0,611**	**0,614**	**0,627**	**0,591**	**0,595**	**0,609**	**0,647**	**0,6**	**0,614**
Logistic Regression + GloVE	CV Fold 1	0,588			0,55			0,63		
	CV Fold 2	0,583			0,537			0,641		
	CV Fold 3	0,599			0,556			0,639		
	CV Average	**0,59**	**0,595**	**0,607**	**0,548**	**0,557**	**0,574**	**0,637**	**0,575**	**0,589**
Linear SVM + TF-IDF	CV Fold 1	0,638			0,614			0,715		
	CV Fold 2	0,65			0,626			0,728		
	CV Fold 3	0,639			0,617			0,716		
	CV Average	**0,642**	**0,637**	**0,655**	**0,619**	**0,637**	**0,641**	**0,720**	**0,64**	**0,644**
Linear SVM + FastText	CV Fold 1	0,606			0,583			0,64		
	CV Fold 2	0,609			0,585			0,652		
	CV Fold 3	0,613			0,593			0,648		
	CV Average	**0,609**	**0,612**	**0,624**	**0,587**	**0,591**	**0,604**	**0,647**	**0,598**	**0,61**
Linear SVM + GloVE	CV Fold 1	0,583			0,54			0,629		
	CV Fold 2	0,578			0,526			0,639		
	CV Fold 3	0,595			0,547			0,638		
	CV Average	**0,585**	**0,591**	**0,6**	**0,538**	**0,549**	**0,561**	**0,635**	**0,57**	**0,58**
Stochastic Gradient Descent + TF-IDF	CV Fold 1	0,616			0,547			0,713		
	CV Fold 2	0,619			0,552			0,724		
	CV Fold 3	0,615			0,553			0,715		
	CV Average	**0,617**	**0,618**	**0,621**	**0,551**	**0,555**	**0,56**	**0,717**	**0,592**	**0,596**
Stochastic Gradient Descent + FastText	CV Fold 1	0,596			0,52			0,639		
	CV Fold 2	0,606			0,552			0,652		
	CV Fold 3	0,608			0,483			0,65		
	CV Average	**0,603**	**0,595**	**0,607**	**0,518**	**0,552**	**0,568**	**0,647**	**0,574**	**0,587**
Stochastic Gradient Descent + GloVE	CV Fold 1	0,564			0,406			0,627		
	CV Fold 2	0,55			0,475			0,64		
	CV Fold 3	0,558			0,411			0,637		
	CV Average	**0,557**	**0,546**	**0,546**	**0,431**	**0,396**	**0,393**	**0,635**	**0,509**	**0,51**
XGBOOST + TF-IDF	CV Fold 1	0,658			0,657			0,709		
	CV Fold 2	0,664			0,664			0,715		
	CV Fold 3	0,651			0,651			0,702		
	CV Average	***0,658***	***0,661***	**0,658**	***0,657***	***0,661***	***0,658***	**0,709**	***0,662***	***0,659***
XGBOOST + FastText	CV Fold 1	0,605			0,597			0,653		
	CV Fold 2	0,604			0,597			0,662		
	CV Fold 3	0,6			0,593			0,656		
	CV Average	**0,603**	**0,605**	**0,611**	**0,596**	**0,598**	**0,604**	**0,657**	**0,598**	**0,604**
XGBOOST + GloVE	CV Fold 1	0,616			0,609			0,667		
	CV Fold 2	0,607			0,6			0,663		
	CV Fold 3	0,663			0,603			0,664		
	CV Average	**0,629**	**0,607**	**0,627**	**0,604**	**0,6**	**0,62**	**0,665**	**0,6**	**0,62**

Models	Folds	Accuracy score				Precision score		Recall score		Test		
		CV	Loss	Test	Loss	CV	Test	CV	Test	AUPRC score	AUC score	F1-score
CNN + FastText	CV Fold 1	0,5473	0,7001			0,5086		0,4144				
	CV Fold 2	0,5492	0,6995			0,512		0,415				
	CV Fold 3	0,555	0,6986			0,513		0,416				
	CV Average/ Best Loss	***0,5505***	***0,6986***	***55,58%***	0,69	**0,5112**	**54,39%**	**0,4151**	**54,11%**	**0,506**	**0,5608**	**53,76%**
CNN + GloVE	CV Fold 1	0,5463	0,6876			0,5104		0,3254				
	CV Fold 2	0,5468	0,6874			0,5112		0,3246				
	CV Fold 3	0,5473	0,6874			0,5122		0,3208				
	CV Average/ Best Loss	**0,5468**	**0,6874**	**55,29%**	0,69	**0,5113**	**54,45%**	**0,3236**	**53,60%**	**0,509**	**0,552**	**52,07%**
LSTM + FastText	CV Fold 1	0,6972	0,7062			0,7295		0,5427				
	CV Fold 2	0,7004	0,6966			0,7007		0,608				
	CV Fold 3	0,7047	0,698			0,7014		0,623				
	CV Average/ Best Loss	**0,7008**	**0,6966**	**70,63%**	0,69	**0,7105**	**70,58%**	**0,59123**	**70,08%**	**0,758**	**0,779**	**70,16%**
LSTM + GloVE	CV Fold 1	0,6999	0,6568			0,7112		0,5847				
	CV Fold 2	0,7002	0,6519			0,7352		0,5437				
	CV Fold 3	0,7049	0,6552			0,7394		0,5532				
	CV Average/ Best Loss	**0,7017**	**0,6519**	**70,32%**	0,65	**0,7286**	**71,05%**	**0,561**	**69,28%**	**0,75**	**0,776**	**69,23%**
GRU + FastText	CV Fold 1	0,7131	0,5806			0,7165		0,6221				
	CV Fold 2	0,7087	0,5773			0,6669		0,7321				
	CV Fold 3	0,7166	0,5774			0,7069		0,6553				
	CV Average/ Best Loss	**0,713**	**0,5773**	**71,97%**	0,57	**0,6968**	**71,85%**	**0,670**	**71,56%**	**0,781**	**0,8**	**71,64%**
GRU + GloVE	CV Fold 1	0,7126	0,577			0,7288		0,5969				
	CV Fold 2	0,7116	0,5782			0,7195		0,6108				
	CV Fold 3	0,7136	0,5818			0,7117		0,6336				
	CV Average/ Best Loss	**0,7126**	**0,577**	**71,64%**	0,58	**0,7200**	**71,59%**	**0,614**	**71,12%**	**0,77**	**0,795**	**71,21%**
BIGRU + FastText	CV Fold 1	0,7215	0,6688			0,7221		0,641				
	CV Fold 2	0,7228	0,6784			0,72		0,6497				
	CV Fold 3	0,7205	0,6625			0,7244		0,6329				
	CV Average/ Best Loss	**0,7216**	**0,6625**	**72,68%**	0,66	**0,722**	**72,65%**	**0,6412**	**72,19%**	**0,79**	**0,81**	**72,29%**
BIGRU + GloVE	CV Fold 1	0,7145	0,5873			0,7077		0,6459				
	CV Fold 2	0,7042	0,6102			0,7662		0,5133				
	CV Fold 3	0,7087	0,6066			0,7181		0,6035				
	CV Average/ Best Loss	**0,709**	**0,5873**	**72,27%**	0,59	**0,7307**	**72,16%**	**0,588**	**71,87%**	**0,772**	**0,797**	**71,94%**
BIGRU + CNN + FastText	CV Fold 1	0,7091	0,7234			0,7319		0,5797				
	CV Fold 2	0,7123	0,7479			0,744		0,5705				
	CV Fold 3	0,71	0,7673			0,7289		0,5917				
	CV Average/ Best Loss	**0,7105**	**0,7234**	**71,04%**	0,74	**0,7343**	**71,73%**	**0,5806**	**70,05%**	**0,77**	**0,788**	**70,04%**
BIGRU + CNN + GloVE	CV Fold 1	0,6477	0,622			0,7713		0,3322				
	CV Fold 2	0,6779	0,5886			0,7321		0,4724				
	CV Fold 3	0,68477	0,5764			0,6798		0,5941				
	CV Average/ Best Loss	**0,6701**	**0,5764**	**68,72%**	0,57	**0,7277**	**68,61%**	**0,4662**	**68,16%**	**0,74**	**0,76**	**68,22%**

Fig. 5. Performance of machine learning and deep learning models in Experiment 3

We observe that the best results across all experiments combined were obtained by BiGRU + fastText, achieving **72.68**% accuracy, **72.65**% precision, **72.19**% recall, **72.29**% F1-score, **79**% AUPRC and **81**% AUC-ROC. By combining original and synthetically perturbated data, we achieved the highest performance across various models.

Conclusion

This work achieved several key objectives.

First, it demonstrated the relevance of using a synthetic dataset as a novel approach for hate speech classification, offering greater flexibility compared to traditional, outdated datasets commonly employed in literature. The experiments indicate that synthetic data circumvent limitations related to sensitive content and enable training on texts featuring highly negative or rarely occurring communication scenarios that are underrepresented in real-world datasets. Consequently, the resulting models exhibit improved effectiveness in detecting hate speech.

Second, the work investigated the impact of data type on model performance. The lowest accuracy (**52.04**%) was observed when models were trained on original data and evaluated on synthetically perturbated data. Training synthetically perturbated data and evaluating original data improved performance (**62.13**% accuracy). The highest performance (**72.68**% accuracy) was achieved when models were trained and evaluated on a combined dataset, regardless of the data's original or synthetically perturbated nature.

Neural networks consistently outperformed traditional machine learning algorithms. In particular, the **BiGRU + fastText** model achieved the best overall classification results, highlighting the effectiveness of bidirectional architectures, GRU units and fastText word embeddings. The first two experiments can be interpreted as involving zero-shot learning, suggesting that further performance improvements may require alternative architecture or larger datasets.

Finally, future work could focus on leveraging pre-trained large language models [30] on expanded synthetic datasets, as proposed in [7], to further enhance model performance.

REFERENCES

1. **Waseem Z., Hovy D.** Hateful symbols or hateful people? Predictive features for hate speech detection on Twitter. *Proceedings of the NAACL Student Research Workshop*, 2016, Pp. 88–93. DOI: 10.18653/v1/N16-2013

2. **Basina P.A., Gojko E.YU., Petrov E.Yu., Bakulin V.V.** Klassifikaciya publikacij soobshchestv "VKontakte" dlya ocenki kachestva zhizni naseleniya [Classification of publications of VKontakte communities for assessing the quality of life of the population]. *Komp'yuternaya lingvistika i intellektual'nye tekhnologii: po materialam ezhegodnoj mezhdunarodnoj konferencii "Dialog"* [*Computational linguistics and intelligent technologies: based on the materials of the annual international conference "Dialogue"*], 2022, Vol. 21 (C), Pp. 1001–1016.

3. **El Koshiry A.M., Eliwa E.H.I., El-Hafeez T.A., Omar A.** Arabic toxic tweet classification: Leveraging the AraBERT model. *Big Data and Cognitive Computing*, 2023, Vol. 4, No. 7, Art. no. 170. DOI: 10.3390/bdcc7040170

4. **Ribeiro A., Silva N.** INF-HatEval at SemEval-2019 Task 5: Convolutional neural networks for hate speech detection against women and immigrants on Twitter. *Proceedings of the 13th International Workshop on Semantic Evaluation*, 2019, Pp. 420–425. DOI: 10.18653/v1/S19-2074

5. **Geet d'Sa A., Illina I., Fohr D.** Classification of hate speech using deep neural networks, *Revue d'Information Scientifique & Technique*, 2020, Vol. 25, No. 1, Art. no. hal-03101938.

6. **Smetanin S.I.** Toxic comments detection in Russian. *Computational Linguistics and Intellectual Technologies*, 2020, Vol. 26, No. 19, Pp. 1149–1159. DOI: 10.28995/2075-7182-2020-19-1149-1159

7. **Vidgen B., Thrush T., Waseem Z., Kiela D.** Learning from the worst: Dynamically generated datasets to improve online hate detection. *arXiv:2012.15761*, 2020. DOI: 10.48550/arXiv.2012.15761

8. **Bitton J., Pavlova M., Evtimov I.** Adversarial text normalization. *arXiv:2206.04137*, 2022. DOI: 10.48550/arXiv.2206.04137

9. **Hartvigsen T., Saadia G., Palangi H.** ToxiGen: A large-scale machine-generated dataset for adversarial and implicit hate speech detection. *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics*, 2022, Vol. 1, Pp. 3309–3326. DOI: 10.18653/v1/2022.acl-long.234

10. **Saffari H., Shafiei M., Zhang H., Harris L., Moosavi N.S.** Beyond hate speech: NLP's challenges and opportunities in uncovering dehumanizing language, *arXiv:2402.13818*, 2024. DOI: 10.48550/arXiv.2402.13818

11. **Zhao X., Lu Z., Xu D., Yuan S.** Generating Textual adversaries with minimal perturbation. *arXiv:2211.06571*, 2022. DOI: 10.48550/arXiv.2211.06571

12. **Roth T., Gao Y., Abuadbba A., Nepal S., Liu W.** Token-modification adversarial attacks for natural language processing: A survey. *arXiv:2103.00676*, 2021. DOI: 10.48550/arXiv.2103.00676

13. **Wang B., Xu C., Liu X., Cheng Y., Li B.** SemAttack: Natural textual attacks via different semantic spaces. *arXiv:2205.01287*, 2022. DOI: 10.48550/arXiv.2205.01287

14. **Gutiérrez-Megías A., Jiménez-Zafra S.M., Ureña L.A., Martínez-Cámara E.** Smart lexical search for label flipping adversial attack. *Proceedings of the Fifth Workshop on Privacy in Natural Language Processing*, 2024, Pp. 97–106.

15. **Badri N., Kboubi F., Chaibi A.H.** Combining FastText and Glove word embedding for offensive and hate speech text detection. *Procedia Computer Science*, 2022, Vol. 207, Pp. 769–778. DOI: 10.1016/j.procs.2022.09.132

16. **Kosykh N.E., Molodkin I.A., Khomonenko A.D.** Features of text preprocessing for performing sentiment analysis. *Intellectual Technologies on Transport*, 2022, No. 3, Pp. 68–73. DOI: 10.24412/2413-2527-2022-331-68-73

17. **Zinovyeva E., Härdle W.K., Lessmann S.** Antisocial online behavior detection using deep learning. *Decision Support Systems*, 2019, Vol. 138, Art. no. 113362. DOI: 10.1016/j.dss.2020.113362

18. **Minaee S., Kalchbrenner N., Cambria E., Nikzad N., Chenaghlu M., Gao J.** Deep learning-based text classification: A comprehensive review. *ACM Computing Surveys* (*CSUR*), 2022, Vol. 54, No. 3, Art. no. 62. DOI: 10.1145/3439726

19. **Widiastuti N.I.** Convolution neural network for text mining and natural language processing. *IOP Conference Series: Materials Science and Engineering*, 2019, Vol. 662, No. 5, Art. no. 052010. DOI: 10.1088/1757-899X/662/5/052010

20. **Alkomah F., Ma X.** A literature review of textual hate speech detection methods and datasets. *Information*, 2022, Vol. 13, No. 6, Art. no. 273. DOI: 10.3390/info13060273

21. **Soni S., Chouhan S.S., Rathore S.S.** *TextConvoNet*: a convolutional neural network-based architecture for text classification. *Applied Intelligence*, 2023, Vol. 53, Pp.14249–14268. DOI: 10.1007/s10489-022-04221-9

22. **Georgakopoulos S.V., Tasoulis S.K., Vrahatis A.G., Plagianakos V.P.** Convolutional neural networks for toxic comment classification. *SETN '18: Proceedings of the 10th Hellenic Conference on Artificial Intelligence*, 2018, Art. no. 35. DOI: 10.1145/3200947.320806

23. **Maslej-Krešňáková V., Sarnovský M., Butka P., Machová K.** Comparison of deep learning models and various text pre-processing techniques for the toxic comments classification. *Applied Sciences*, 2020, Vol. 10, No. 23, Art. no. 8631. DOI: 10.3390/app10238631

24. **Budyl'skij D.V.** GRU i LSTM: sovremennye rekurrentnye nejronnye seti [GRU and LSTM: Modern Recurrent Neural Networks]. *Molodoj uchenyj* [*Young Scientist*], 2015, Vol. 95, No. 15, Pp. 51–54.

25. **Beniwal R., Maurya A.** Toxic comment classification using hybrid deep learning model. *Sustainable Communication Networks and Application*, 2021, Vol. 55, Pp. 461–473. DOI: 10.1007/978-981-15-8677-4_38

26. **Mironenko V.V., Saveleva A.A., Sodikov S.A.** Overview of methods for analysis of natural language based on machine learning models. *Aktual'nye problemy aviacii i kosmonavtiki* [*Current issues in aviation and astronautics*], 2021, Vol. 2, Pp. 169–170.

27. **Tan K.L., Lee C.P., Lim K.M.** RoBERTa-GRU: A hybrid deep learning model for enhanced sentiment analysis. *Applied Sciences*, 2023, Vol. 13, No. 6, Art. no. 3915. DOI: 10.3390/app13063915

28. **Ivanovs M., Kadikis R., Ozols K.** Perturbation-based methods for explaining deep neural networks: A survey. *Pattern Recognition Letters*, 2021, Vol. 150, Pp. 228–234. DOI: 10.1016/j.patrec.2021.06.030

29. **Platonov E.N., Rudenko V.Y.** Identification and Classification of Toxic Statements by Machine Learning Methods. *Modelling and Data Analysis*, 2022, Vol. 12, No. 1, Pp. 27–48. DOI: 10.17759/mda.2022120103

30. **Kureichik V.V., Rodzin S.I., Bova V.V.** Deep learning methods for natural language text processing. *IzvestiyaSFedU. Engineering Sciences*, 2022, Vol. 226, No. 2, Pp. 189–199. DOI: 10.18522/2311-3103-2022-2-189-199

INFORMATION ABOUT AUTHORS / СВЕДЕНИЯ ОБ АВТОРАХ

Luc Prucell Mbele Ossiyi
Мбеле Оссийи Люк Прюсель
E-mail: lucprucell@gmail.com
ORCID: https://orcid.org/0009-0002-3090-2809

Pavel D. Drobintsev
Дробинцев Павел Дмитриевич
E-mail: drobintsev_pd@spbstu.ru

Sergey M. Ustinov
Устинов Сергей Михайлович
E-mail: usm50@yandex.ru
ORCID: https://orcid.org/0000-0003-4088-4798

Submitted: 15.01.2025; Approved: 11.04.2025; Accepted: 25.04.2025.
Поступила: 15.01.2025; Одобрена: 11.04.2025; Принята: 25.04.2025.

Computing, Telecommunication and Control, 2025, Vol. 18, No. 3, Pp. 36–45.
Информатика, телекоммуникации и управление. 2025. Том 18, № 3. С. 36–45.

Research article
DOI: https://doi.org/10.18721/JCSTCS.18303
UDC 004.932

AUTOMATION OF BIOLOGICAL CELL IMAGE PROCESSING

K.A. Turchinskii ✉ , *A.Ye. Krasnov*

Russian State Social University (RSSU), Moscow, Russian Federation

✉ turchin.sky@yandex.ru

Abstract. When analyzing images of biological cells, automated methods for segmentation and result storage are becoming increasingly in demand. Manual annotation is extremely labor-intensive and does not scale to large volumes of data, while conventional segmentation algorithms create binary masks of substantial size. The objective of this work is to develop a software pipeline that combines local threshold filtering and morphological post-processing to obtain an accurate binary mask and then encodes the result using run-length encoding (RLE) to reduce storage space. Methods used are as follows: at the segmentation stage, local statistical criteria are applied, followed by morphological closing. For storing the result, several modifications of RLE (standard, Foreground-Only, DRLE and Z-order) are implemented along with their comparative analysis. The scientific novelty of the work lies in the comprehensive integration of block filtering and morphology with subsequent compression of binary segmentation masks in the task of erythrocyte (and other cells) segmentation. This approach significantly reduces storage requirements without substantial loss of accuracy. The proposed solution demonstrates high metrics (Accuracy, IoU, Dice) while substantial memory savings. The practical significance is that the developed software pipeline can be easily integrated into biomedical data analysis systems, accelerating the mass processing of cell images and reducing the demands on storage infrastructure.

Keywords: segmentation, biological images, run-length encoding, local threshold filtering, morphological post-processing, automation, accuracy

Citation: Turchinskii K.A., Krasnov A.Ye. Automation of biological cell image processing. Computing, Telecommunications and Control, 2025, Vol. 18, No. 3, Pp. 36–45. DOI: 10.18721/JCSTCS.18303

Научная статья
DOI: https://doi.org/10.18721/JCSTCS.18303
УДК 004.932

АВТОМАТИЗАЦИЯ ОБРАБОТКИ ИЗОБРАЖЕНИЙ БИОЛОГИЧЕСКИХ КЛЕТОК

К.А. Турчинский ✉ , *А.Е. Краснов*

Российский государственный социальный университет (РГСУ),
Москва, Российская Федерация

✉ turchin.sky@yandex.ru

Аннотация. При анализе изображений биологических клеток все более востребованными становятся автоматизированные методы сегментации и хранения результатов. Ручная разметка чрезвычайно трудоемка и не масштабируется на большие объемы данных, а обычные алгоритмы сегментации создают бинарные маски значительного объема. Целью работы является разработка программного конвейера, который сочетает локальную пороговую фильтрацию и морфологическую постобработку для получения точной бинарной маски, а затем кодирует результат отрезками разной длины (RLE) для уменьшения занимаемого пространства. Используемые методы: на этапе сегментации применяются локальные статистические критерии, за которыми следует морфологическое закрытие. Для хранения результата внедряются несколько модификаций RLE (стандартная, Foreground-Only, DRLE и Z-order) с их сравнительным анализом. Научная новизна работы заключается в комплексном объединении блоковой фильтрации и морфологии с последующим сжатием бинарных сегментационных масок в задаче сегментации клеток, что позволяет существенно сократить объем хранения без значимого ущерба точности. Полученное решение демонстрирует высокие метрики (Accuracy, IoU, Dice) при существенной экономии памяти. Практическая значимость исследования состоит в том, что разработанный программный конвейер легко интегрируется в системы анализа биомедицинских данных, ускоряя массовую обработку изображений клеток и снижая требования к инфраструктуре хранения.

Ключевые слова: сегментация клеток, биологические изображения, кодирование отрезками разной длины, локальная пороговая фильтрация, морфологическая постобработка, автоматизация, точность

Для цитирования: Turchinskii K.A., Krasnov A.Ye. Automation of biological cell image processing // Computing, Telecommunications and Control. 2025. Т. 18, № 3. С. 36–45. DOI: 10.18721/JCSTCS.18303

Introduction

Digital processing of cell images is becoming increasingly important in biological and medical practice: from visualization of erythrocytes and leukocytes to analysis of tissues and micro-objects [1, 2]. Manual image annotation is an extremely labor-intensive process, which becomes more complicated as data volumes grow. Therefore, there is a demand for automated pipelines that can not only segment target regions (e.g., cell boundaries), but also efficiently store the results as binary masks.

However, traditional approaches to obtaining and storing masks face the problem of data redundancy: even single objects in large fields of view can lead to significant memory costs [3, 4]. To solve this problem, run-length encoding (RLE) and its various modifications – Foreground-Only, DRLE, Z-order [2, 3, 4, 5] – are widely used. At the same time, not only the RLE method itself is important, but also the quality of pre-segmentation, which directly affects the structure of the binary mask and potential compression efficiency.

Medical and biological image processing imposes additional accuracy requirements: incorrectly defined cell boundaries lead to errors in subsequent analyses (e.g., in cell concentration calculation, cell shape estimation etc.) [2]. Consequently, in order to develop reliable automated processing systems, it is necessary to combine high segmentation accuracy with optimized storage. Methods such as local threshold filtering and morphological post-processing, which have proved their effectiveness when working with heterogeneous images[1], can improve segmentation results and prepare them for more efficient compression.

The relevance of the topic under study is driven by the need to improve performance and reduce memory costs in image processing systems that require long-term storage of a large number of segmented cell images. The development of this approach facilitates the scaling of laboratory and clinical studies, enable rapid processing of microscopic observations and support the creation of databases for subsequent analysis, diagnosis and training of neural network models [4]. This creates a practical demand for a comprehensive solution: automatic cell segmentation followed by efficient mask encoding and compressed storage.

The aim of this work is to develop a software pipeline that combines local threshold filtering and morphological post-processing to obtain an accurate binary mask, and then encodes the result using RLE to reduce storage space.

To achieve this goal, the following tasks were solved:

1. To design an algorithm that combines adaptive segmentation and morphological operations with further RLE.

2. To create a prototype of a software package including filtering, contour extraction, binary mask packing and quality metrics computation functions.

3. To test the pipeline on real biological images and evaluate the achieved compression and the compliance of the segmented regions to the reference data.

4. To analyze the influence of filtering and morphology parameters on the error magnitude and volume of encoded masks in order to formulate practical recommendations.

Description of the proposed algorithm

A system processing a discrete image $Im_{m,n}$ of size $M \times N$ pixels (where $m = 1, 2, ..., M$ is the row index and $n = 1, 2, ..., N$ is the column index) will face the problem of brightness inhomogeneities (e.g., uneven illumination and noise). To solve this problem, we introduce a local threshold filtering stage.

Block partitioning is performed as follows. A field of size $M \times N$ is divided into non-overlapping fragments (blocks) of square $B \times B$ or rectangular shape $B_h \times B_w$. In this case, the indices of pixels belonging to the block with number (i, j), approximately satisfy the following expression:

$$i \in \left[(i-1) \cdot B_h + 1, i \cdot B_h\right], \quad j \in \left[(j-1) \cdot B_w + 1, j \cdot B_w\right], \tag{1}$$

where B_h and B_w are the block height and width, and the total number of blocks vertically is equal to M/B_h, horizontally – to N/B_w.

Fig. 1 shows the partitioning of the original image into blocks of size $B \times B$ with a "step" of B, meaning adjacent blocks follow one another without overlap. For each block $\{i, j\}$, the average brightness $\overline{\mathrm{Im}}_{i,j}$ and local standard deviation $\sigma_{i,j}$ are calculated. Next, the local threshold $\theta_{i,j}$ is formed using the following formula:

$$\theta_{i,j} = \overline{\mathrm{Im}}_{i,j} - \alpha\sigma_{i,j}, \tag{2}$$

[1] Voxel Compression, Available: https://eisenwave.github.io/voxel-compression-docs/rle/space_filling_curves.html (Accessed 19.09.2025).

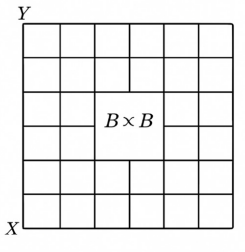


Fig. 1. Schematic partitioning of the original image of size $M \times N$ into non-overlapping blocks $B \times B$ (step = B)

where α is the coefficient adjusting noise sensitivity. Similar expressions are used in a number of adaptive thresholding methods of object extraction[2] [5].

Each pixel (m_i, n_i) inside the current block $\{i, j\}$ is compared with the threshold $\theta_{i,j}$ (3). If

$$\mathrm{Im}_{m_i,n_i} \le \theta_{i,j}, \tag{3}$$

then the pixel is considered to belong to the "cell" region; otherwise, it is assigned to the "background". As a result of this step, an intermediate binary map A is generated, where each pixel is assigned a value of 1 (object) or 0 (background).

The output of the local threshold filtering is a binary map $A(M \times N)$ (4), where

$$A_{m,n} = \begin{cases} 1 : \mathrm{Im}_{m,n} \le \theta_{i,j} \\ 0 : \mathrm{Im}_{m,n} > \theta_{i,j} \end{cases}. \tag{4}$$

Subsequently, morphological processing is used to clean A and eliminate "gaps" or "outliers".

Morphological closing (Close) is usually applied when it is necessary to "connect" adjacent white regions (objects). In its classical form, the closing operation is defined by the following formula:

$$\mathrm{Close}(A) = (A \oplus S) \ominus S, \tag{5}$$

where A is the original binary map (see above); S is a structuring element, such as a circle (disk) or ellipse; $\oplus$ is morphological dilatation (expansion); $\ominus$ is morphological erosion (reduction).

Visually, closing "fills in" small gaps inside an object, thereby merging segments.

Morphological opening $\left(\mathrm{Open}(A) = (A \ominus S) \oplus S\right)$ is used to "clean" the map from random small noises (outliers) that have a small area and do not belong to real cells [6].

The choice between closing, opening or their combination is determined by the characteristic sizes of the cells and the noise content. If it is known that cells can be closely packed, applying closing is advisable. If it is necessary to get rid of "point-like" fragments, adding opening is useful. The final contour of each region, and therefore the structure of the final binary mask, directly depends on the morphology settings.

When the morphologically processed binary map A (the "final mask") is ready, it is subjected to packing using RLE method.

Before encoding, a pixel traversal scheme is defined, for example:

1. Row-major – left to right, top to bottom (by rows).

[2] Voxel Compression, Available: https://eisenwave.github.io/voxel-compression-docs/rle/space_filling_curves.html (Accessed 19.09.2025).

2. Z-order (Morton-order) – pixels are rearranged into a special sequence, more favorable for certain object configurations [3, 8].

During RLE, each consecutive pixel fragment with the same value (0 or 1) is replaced with a "run". In its classical form, this is a pair (r, v), where r is the number of consecutive identical bits, and v is the bit itself (0 or 1). In Foreground-Only encoding [4], where only the "runs" of units are stored, records of the form (s, l) are used, where s is the start index of a block of unit pixels, and l is the length of this continuous block. A variant of differential RLE additionally encodes the difference between the lengths of consecutive runs, which sometimes reduces the overall data volume [6]. Z-order encoding is often useful if objects are highly "clustered" and arranged in compact groups.

The criterion for choosing the encoding method (DRLE, Z-order, Foreground-Only etc.) depends on the number of cells, their sizes, noise density and decoding frequency. If decoding needs to be performed frequently, then the simplest format RLE(r,v), which does not require complex permutations or calculations, is more convenient.

The final stage is packing and saving the run-length data (i.e., a set of pairs like (r, v) or (s, l) etc.) in a form convenient for storage. In applied scenarios, this may be:

– structured record;
– serialization;
– general "stream" compression.

The essence of "run-length data" is a list (or other structure) of segments with the same bit. Because there are often extensive areas of background outside the cells, RLE achieves significant memory savings. When there are many disparate tiny objects, other options (e.g., Foreground-Only) are chosen to avoid storing long chains of zeros.

The modular architecture of the developed solution provides flexibility:

1. Local threshold filtering is easily adjustable to cell size and contrast.
2. Morphological operations can be varied (use opening, closing or their combination, changing the type of structuring element).
3. RLE system can be replaced or supplemented with other compression methods (e.g., Quadtree, Octree or ZIP archiving).

This simplifies the adaptation of the algorithm to different types of biological images, as well as scaling for large data volumes. If other tools (non-Python) have to be used in the future, it is still convenient to have run-length data, as it is easy to import and unpack in most environments.

It is convenient to describe the technological pipeline using a diagram reflecting the sequence of blocks and principles of interaction (Fig. 2).

The loading module accepts a list of source files stored in any available graphic formats and sends each image sequentially to the local threshold filtering stage. A block-based or adaptive strategy is used to flexibly adjust the brightness threshold when different frame zones have uneven intensity. The morphological module eliminates minor defects that may arise from random noise or optical system artefacts. Closing with a suitable structuring element allows merging adjacent objects to form more solid cell regions, while opening filters out excess details protruding beyond the intended contours.

The resulting system is capable to operate in batch mode, sequentially processing an extensive collection of images and generating grouped results in a convenient format for analysis or further transfer. The practical significance lies in resource savings: storage capacity, transfer speed and processing requirements become more manageable, while the preservation of contours allows for statistics on cell number and shape to be conducted.

Architecture and implementation of the software package

The foundation of the work is implemented through a set of modules, each performing its own functions for segmentation, packing the results and interacting with the user interface. One of the key

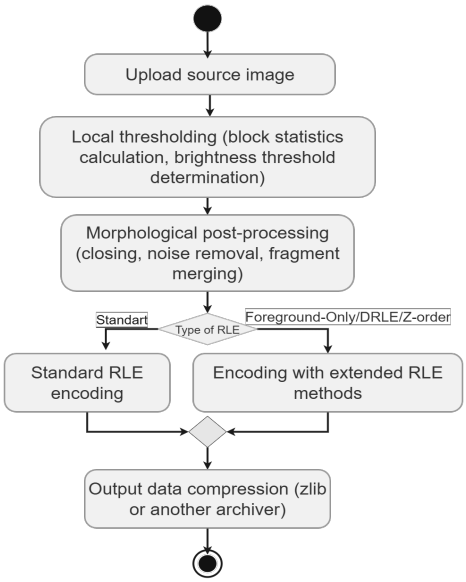


Fig. 2. Algorithm for automation of biological image processing

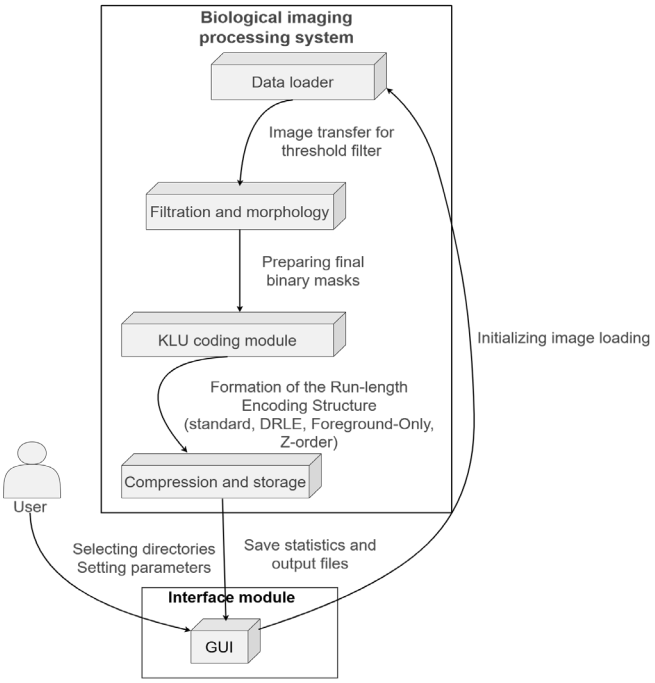


Fig. 3. Architecture of the software package

requirements is that the listed stages, including morphological correction and various RLE schemes, are flexible in configuration and can be extended with additional procedures if necessary. The architecture diagram in Fig. 3 reflects the overall flow of transformations.

The calling process, initiated by user actions, involves reading the necessary files in the "Data Loader" block. The local or remote directory access module is configured to read images in batch

mode, and all intermediate results and service metadata are collected into a single structure. Next, the data enters the node responsible for threshold filtering and morphological operations. It is at this stage that local brightness estimation, adaptive threshold calculation, closing of detected gaps and removal of noise inclusions occur, if they do not match the expected cell contours.

Encapsulation of different RLE schemes within a separate module simplifies further maintenance, as it is easy to add new packing variants or special optimization modes to the working pipeline, depending on the features of the images under study. All types of RLE conclude with the formation of a structure, which is packed using zlib or other compression methods. The "Compression and Storage" module outputs the result in a form convenient for storage and transfer, while informational responces, e.g., statistics on the number of processed frames, final packet size, conversion time, are sent back to the "Interface Module".

The software implementation relies on image processing libraries (OpenCV, Pillow), as well as tools for mathematical calculations (NumPy) and serialization (pickle). Such a suite ensures operation both on local machines and on server systems, where it is necessary to scale the processing of large sets of microphotographs. At the same time, the same encoding module can work on data that has undergone different types of morphological correction. This method is suitable for situations where different cell types have a particular contour structure and require the selection of specific morphology or filtering settings.

Experimental results and analyses

The system pipeline was tested on a dataset consisting of 343 annotated biological images (BloodImage_XXXX). For each image, stored alongside the corresponding XML annotation, a binary mask reflecting the location of cells (RBCs) in the frame was automatically generated. Local threshold filtering with fixed block size and morphological closing to correct small gaps within regions were the key steps. Subsequently, the resulting maps were subjected sequentially to classical RLE, Foreground-Only, DRLE, Z-order RLE and a combined method incorporating local threshold filtering with RLE. It was important to capture computational cost, compressed file size and accuracy metrics to understand the relationship between packing efficiency and cell structure preservation. Table 1 shows the main results.

Table 1

Comparative analysis with the proposed method

Parameters	Standard RLE	Foreground-Only RLE	DRLE	Z-order RLE	RLE Enhanced Local Filtering + RLE
Total encoding time, seconds	9.4137	7.7363	9.3730	9.8173	10.2493
Total decode time, seconds	0.3608	0.1910	0.4326	0.9448	0.3835
Average size, bytes	406.36	2791.14	512.78	1525.84	391.92
Average accuracy, %	100.00	100.00	100.00	100.00	99.54
Average *IoU*	1.00	1.00	1.00	1.00	0.9873
Average *Dice*	1.00	1.00	1.00	1.00	0.9936

The full sample included 343 XML descriptions. The processing time recorded by the system took approximately one minute at an average speed of about 6 images per second. Sequential encoding-decoding stages were considered, and the average size in bytes was calculated to get an idea of typical compressed mask volumes. Improved local threshold filtering in the last method (the fifth variant) additionally allowed for the estimation of accuracy (acc), IoU and Dice, as its own procedure modified the original boundaries.

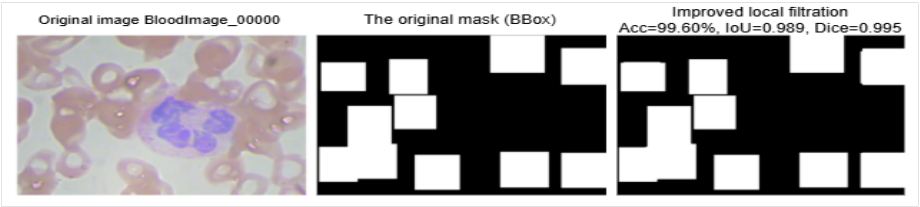


Fig. 4. Visualization of image processing by the proposed method

Combined analysis of encoding costs and file size confirmed that classical RLE yielded a fairly modest volume indicator (406.36 bytes on average) with a noticeably lower decoding time (0.3608 sec). However, the Foreground-Only scheme presented a failure: the average size increased to 2791.14 bytes, indicating a high number of disparate fragments requiring separate description. DRLE maintained a moderate compactness (512.78 bytes) with a slight increase in decoding time (0.4326 sec). Z-order increased the size to 1525.84 bytes, most likely because mixing local mask blocks did not always lead to longer runs in the overall structure. The standard schemes (Standard RLE, Foreground-Only RLE, DRLE and Z-order RLE) are lossless: they do not alter the binary mask and hence exhibit maximum accuracy (acc = 100%, IoU = 1.00, Dice = 1.00). Nevertheless, their average size varies from 406 to 2791 bytes. The proposed "local threshold filtering + morphology + RLE" complex yields the smallest average size of 391.92 bytes with near-maximum accuracy (acc = 99.54%, IoU = 0.9873, Dice = = 0.9936), which we consider as an optimal compromise between memory savings and the preservation of diagnostically significant contours. For each scheme, the accuracy metrics are equal to one because the encoding-decoding operation does not change the original partitioning. We have included these metrics in Table 1 to ensure a correct comparison by size + accuracy criteria.

Contour overlap (Dice = 0.9936) was particularly impressive, indicating that on average the final mask matched the reference data almost without distortion. Hence, local filtering and morphology not only preserved the integrity of cell regions, but also contributed to reducing the volume of the recorded mask due to larger homogeneous areas. The assumption that closing eliminates internal gaps and thus increases run lengths was confirmed empirically. The data obtained for Z-order indicated that Morton permutation is not always optimal, especially when cells are located in different regions and their mutual geometry does not create large connected clusters.

A visual comparison of the segmentation is presented in Fig. 4, where the original frame contains a blood fragment with individual cellular elements, and the resulting masks demonstrate the importance of merging closely located areas to improve boundary consistency.

The above example confirms the conclusions drawn from the encoding time and size statistics. The improved local filtering that corrects illumination variations within the image eliminates minor noise objects, resulting in a clearer binary map and longer sequences of 1's or 0's during subsequent RLE.

The results indicate the high efficiency of the proposed integrated approach, where filtering and morphology prepare the image for RLE. The obtained metrics facilitate the large-scale automation of cell analysis with big data volumes and help save memory or transmission channels without the risk of losing key diagnostic details. This strategy is particularly appropriate in the biomedical field, where precise contours are required while the compactness of the recorded structures is also important.

Conclusions

The testing of the complex solution based on local threshold filtering, morphological post-processing and subsequent mask encoding using different RLE schemes demonstrates a significant gain when

working with large arrays of microscopic images. Statistical data for 343 annotated images confirm that careful noise removal within segmented regions and correct fragment smoothing have a noticeable effect on reducing the final volume, while improving accuracy metrics and preserving cell configuration. Applying classical RLE to masks that have undergone additional local thresholding and morphological closing procedures results in a relatively small file size and an almost complete match with the reference boundaries.

The observed advantage over alternative methods is explained by the fact that local filtering adjusts to the details of each image fragment, eliminating global thresholding errors, while closing merges disparate pixels into large regions where the runs become longer. Morton (Z-order) permutation can be convenient for strictly localized clusters, but in some cases does not provide improved compression. Foreground-Only scheme is effective when the background occupies almost the entire area and cells are concentrated in one region, but leads to redundant intervals for fragmented objects. The results confirm the need for an integrated approach: each module (filtering, morphology, RLE) enhances the positive effect of the other elements. The studied pipeline simplifies the large-scale automation of medical analysis tasks and opens up opportunities for integration into laboratory and clinical systems oriented to the rapid processing and storage of a large number of biological images.

REFERENCES

1. **Haggere L.R., Alagarswamy R.** HWCD: A hybrid approach for image compression using wavelet, encryption using confusion, and decryption using diffusion scheme. *Journal of Intelligent Systems*, 2023, Vol. 32, No. 1, Art. no. 20229056. DOI: 10.1515/jisys-2022-9056

2. **Rahman M.** Comparative analysis of run-length encoding techniques: Standard, modified, two-dimensional, and bitwise approaches for efficient data compression with numerical example, 2025.

3. **Thong Y.J.** An introduction to bitmask representations and encodings: RLE vs REE. Datature, 2024. Available: https://www.datature.io/blog/an-introduction-to-bitmask-representations-and-encodings-rle-vs-ree (Accessed 19.09.2025).

4. **Schmidt K., Horowitz J.** *Method and apparatus for double run-length encoding of binary data*. European Patent, No. EP0783208A2, 1997.

5. **Haghighi K.G., Mirnia M.K., Navin A.H.** Optimizing run-length algorithm using octonary repetition tree. *arXiv:1611.09664*, 2016. DOI: 10.48550/arXiv.1611.09664

6. **Samson A.Ch., Sastry V.U.K.** An improved run length encoding scheme for image compression. *International Journal of Engineering and Computer Science*, 2017, Vol. 6, No. 3, Art. no. 57. DOI: 10.18535/ijecs/v6i3.57

7. **York T.** Quadtrees for image processing. *Medium*, 2020. Available: https://medium.com/@tannerwyork/quadtrees-for-image-processing-302536c95c00 (Accessed 19.09.2025).

8. **Shen J.** Run-length encode and decode. *Medium*, 2021. Available at: https://medium.com/@ccshenyltw/run-length-encode-and-decode-a33383142e6b (Accessed 19.09.2025).

INFORMATION ABOUT AUTHORS / СВЕДЕНИЯ ОБ АВТОРАХ

Kirill A. Turchinskii
Турчинский Кирилл Александрович
E-mail: turchin.sky@yandex.ru

Andrey Ye. Krasnov
Краснов Андрей Евгеньевич
E-mail: krasnovmgutu@yandex.ru

Submitted: 21.04.2025; Approved: 24.07.2025; Accepted: 18.09.2025.
Поступила: 21.04.2025; Одобрена: 24.07.2025; Принята: 18.09.2025.

Computing, Telecommunication and Control, 2025, Vol. 18, No. 3, Pp. 46–57.
Информатика, телекоммуникации и управление. 2025. Том 18, № 3. С. 46–57.

Research article
DOI: https://doi.org/10.18721/JCSTCS.18304
UDC 004.89

TEXT AUGMENTATION METHOD VIA PARAPHRASTIC CONCEPT EMBEDDINGS: A CASE STUDY ON AZERBAIJANI LANGUAGE

A.F. Aghayev ✉ , *S.A. Molodyakov* , *S.M. Ustinov*

Peter the Great St. Petersburg Polytechnic University,
St. Petersburg, Russian Federation

✉ agaev.af@edu.spbstu.ru

Abstract. A novel data augmentation method – paraphrastic concept embeddings – is presented, designed to address the problem of insufficient labeled data in Azerbaijani natural language processing (NLP). This method generates high-quality paraphrastic sentences by encoding semantic concepts into a continuous vector space and decoding them into diverse textual realizations. This approach is the first to utilize concept-level paraphrasing for the Azerbaijani language, yielding substantial improvements in applied tasks. The theoretical foundations of the method, including its mathematical formulation and implementation within NLP pipelines, are proposed. In text classification experiments, the method outperforms standard augmentation techniques in accuracy and robustness. The method does not require external lexical resources, making it especially useful for low-resource languages. It scales for various types of tasks, including sentiment analysis, entity extraction and text generation. It is concluded that the proposed approach significantly advances the level of Azerbaijani NLP and has the potential to be extended to other low-resource languages.

Keywords: natural language processing, low-resource language, data augmentation, paraphrastic embeddings, concept embedding, text classification

Citation: Aghayev A.F., Molodyakov S.A., Ustinov S.M. Text augmentation method via paraphrastic concept embeddings: A case study on Azerbaijani language. Computing, Telecommunications and Control, 2025, Vol. 18, No. 3, Pp. 46–57. DOI: 10.18721/JCSTCS.18304

Научная статья
DOI: https://doi.org/10.18721/JCSTCS.18304
УДК 004.89

МЕТОД АУГМЕНТАЦИИ ТЕКСТОВ С ПОМОЩЬЮ ПАРАФРАЗНЫХ ВЕКТОРНЫХ ПРЕДСТАВЛЕНИЙ НА ПРИМЕРЕ АЗЕРБАЙДЖАНСКОГО ЯЗЫКА

А.Ф. Агаев ✉ , *С.А. Молодяков* , *С.М. Устинов*

Санкт-Петербургский политехнический университет Петра Великого,
Санкт-Петербург, Российская Федерация

✉ agaev.af@edu.spbstu.ru

Аннотация. Представлен новый метод аугментации данных – парафразные концептуальные векторные представления, – предназначенный для решения проблемы нехватки размеченных данных в азербайджанской обработке естественного языка. Метод генерирует качественные парафразные предложения, кодируя семантические концепты в непрерывное векторное пространство и декодируя их в разнообразные текстовые формы. Это первый подход, использующий концептуальное парафразирование для азербайджанского языка, обеспечивая заметные улучшения в прикладных задачах. Предложены теоретические основы метода, его математическая модель и интеграция в конвейеры обработки данных. В экспериментах по классификации текста метод превосходит стандартные техники аугментации по точности и устойчивости. Метод не требует внешних лексических ресурсов, что делает его особенно полезным для малоресурсных языков. Метод масштабируется для различных типов задач, включая анализ тональности, извлечение сущностей и генерацию текста. Делается вывод, что предложенный подход существенно продвигает уровень обработки естественного азербайджанского языка и имеет потенциал расширения на другие малоресурсные языки.

Ключевые слова: обработка естественного языка, малоресурсный язык, аугментация данных, парафразные векторные представления, контекстные векторные представления, классификация текстов

Для цитирования: Aghayev A.F., Molodyakov S.A., Ustinov S.M. Text augmentation method via paraphrastic concept embeddings: A case study on Azerbaijani language // Computing, Telecommunications and Control. 2025. Т. 18, № 3. С. 46–57. DOI: 10.18721/JCSTCS.18304

Introduction

Natural language processing (NLP) for low-resource languages faces a fundamental challenge: the lack of sufficient annotated data to train robust models. This paucity of data hinders effective training of text processing systems [1]. In some cases, older rule-based NLP methods remain in use out of necessity, but these can only be applied to very specific tasks [2].

Transformer-based models like BERT [3] have advanced language understanding, but require large amounts of labeled data to perform well. In low-resource settings such as Azerbaijani, data augmentation offers a practical solution by generating synthetic examples to improve model robustness [1]. Specifics of the language are important for augmentation. In this paper, the Azerbaijani language is used for experiments.

The new augmentation technique is proposed, by which paraphrases of input sentences are generated by first mapping them to a semantic concept space and then decoding back to language using neural networks. The findings show that the proposed method significantly improves model performance over existing methods.

Related work

Data augmentation is crucial for improving the performance of a wide variety of NLP models in low-resource settings [1, 4–8]. Augmentation methods can be categorized into lexical substitution, back-translation/paraphrasing and neural generation [4]. Methods like EDA [9] use synonym replacement, insertion and deletion [1]. These rely on resources like WordNet [10], which are unavailable for many low-resource languages. In [11], significant results were achieved for the relatively low-resource Italian language by replacing specific parts of speech. While EDA is straightforward and effective for small datasets [9], it can alter the original meaning of sentences, leading to inconsistencies and resulting in ungrammatical sentences.

Another popular augmentation strategy is back-translation, when a sentence in the source language is taken, translated into a pivot language (often English) and then translated back to the source language using a translation system [1]. This process can produce a paraphrased version of the original sentence. Back-translation can be successfully used for text augmentation [12]. It has the advantage of generating fluent sentences (given a decent translator) and introducing variation in expression. Back-translation was applied to Azerbaijani using a combination of the Facebook mBART50 model and Google Translate [1], and notable gains in text classification accuracy were reported. Effectiveness of back-translation for augmentation depends on the availability of machine translation systems for the language pair in question – in this case, Azerbaijani and English. Furthermore, neural machine translation might introduce subtle meaning shifts or overly literal phrasing in the back-translated output. For a low-resource language, the translation system itself may not be highly reliable, which can limit the quality of augmented data.

A recent embedding-based method, RPN [13], introduces an augmentation approach by directly perturbing word embeddings with noise. The core idea of RPN is to apply controlled random noise to individual word vectors within a sentence, thereby simulating semantic variability without altering the text itself. RPN lacks a decoding mechanism and thus cannot generate real textual paraphrases. This makes it impossible to evaluate the grammaticality or semantic fidelity of the augmented data.

Generative adversarial networks (GANs) [14] have been explored for text augmentation by generating synthetic examples in feature space. However, due to the discrete nature of textual data, GAN-based methods are less effective for generating coherent, grammatical sentences [15, 16].

Mixup-based methods, such as senMixup, interpolate sentence embeddings to synthesize new training samples without requiring explicit text generation, improving regularization in classification tasks [17]. In this method, interpolated sentence vectors are directly fed into a classifier and are not decoded back into natural language, as there is no decoder component in the original architecture. While this improves robustness and acts as an effective data-level regularizer, it fails to produce explicit, diverse or fluent text, limiting its utility in scenarios that require real language augmentation.

Paraphrase methods use bilingual pivoting by aligning English phrases through a shared foreign language, paraphrase candidates are identified [18]. Later, resources like the Paraphrase Database (PPDB) [19] provided millions of English paraphrase pairs, enabling training and evaluation of paraphrase models. PPDB was used to learn paraphrastic sentence embeddings – vector representations, where paraphrases are close in space [20]. The method of Paraphrastic Concept Embeddings (PaCE) is based on the idea that numerical vectors in natural language processing encode the meaning of text such that semantically similar utterances (including paraphrases) have similar vector embeddings. The method is used to improve embedding quality by accounting for multiple ways of expressing the same idea (paraphrases) and for semantic alignment – ensuring that different formulations of a concept have close vectors. It has been applied, for example, in sentiment analysis tasks [20]. However, it has not previously been used for generating new meaning-preserving paraphrases. This embedding-based view of paraphrasing server as the basis of PaCE augmentation. The method enables

expansion of training corpora for NLP models. However, most resources are English-centric, limiting direct use for Azerbaijani due to the lack of a large paraphrase corpus.

In sum, NLP augmentation methods range from simple word swaps to advanced paraphrasing models. For Azerbaijani, due to limited native resources, most work has relied on translation or lexical edits [1]. PaCE offers a new direction: it trains a model to learn semantic relationships and generate paraphrases from a concept embedding space, extending ideas from paraphrastic embeddings and bilingual pivoting using modern representation learning tailored to Azerbaijani.

Methodology

The PaCE augmentation pipeline consists of two main components:

1) a concept embedding model that encodes sentences into a semantic vector space,

2) a paraphrase generation mechanism that decodes or transforms vectors in this space back into novel sentences.

In contrast to traditional word-level methods like synonym replacement and back-translation, PaCE operates directly on semantic concepts, which allows generating more semantically coherent and linguistically accurate paraphrases. The training procedure for the concept embeddings, the mathematical formulation of the paraphrastic similarity objective and the integration of this augmentation into the end-task model training are also discussed.

PaCE Model

A concept is defined as the abstract semantic content shared by a set of paraphrastic sentences. Formally, consider two sentences s_1 and s_2 in Azerbaijani that are paraphrases of each other (denoted $s_1 \approx s_2$). They express the same concept (meaning) using different wording. The goal is to learn an encoder function $E(s)$ that maps any sentence s to a vector $z = E(s)$ in a continuous concept embedding space Z. For any pair of sentences s_1, s_2 that are true paraphrases, their embeddings should be close. Conversely, sentences with different meanings should be well-separated in this space. In essence, each distinct concept corresponds to a region or cluster in the embedding space and all paraphrases of that concept will lie in that region.

Model Architecture

To implement $E(s)$, a sequence-to-sequence autoencoder architecture is adopted. The encoder is a neural network (in this case, a transformer encoder similar to BERT's encoder [3]) that produces a fixed-size vector representation of the input sentence. The decoder is another neural network (a transformer decoder) that attempts to reconstruct the original sentence from the embedding. The combined encoder-decoder is first trained as an autoencoder on Azerbaijani text data: given a sentence s, the encoder produces $z = E(s)$, and the decoder generates $\acute{s} = D(z)$, which is trained to match s. This ensures that $E(s)$ retains enough information to reconstruct the sentence, effectively learning a latent representation of the sentence. However, a standard autoencoder alone does not guarantee that paraphrastic sentences map to similar embeddings. Therefore, an additional training signal using paraphrase pairs is introduced. The full architecture is illustrated in Fig. 1–3, which decomposes the PaCE process into modular components.

Paraphrastic pair training

A set of paraphrase pairs $\{(p_i, q_i)\}$, where $p_i \approx q_i$ (sentence p_i is a known paraphrase of q_i), is leveraged. Such pairs can be obtained through various means in a low-resource setting: one approach is to use back-translation or bilingual pivoting on available parallel corpora to produce candidate paraphrases (for example, translate an Azerbaijani sentence into English and back, obtaining a paraphrase). A paraphrase corpus for training was curated by translating a subset of Azerbaijani sentences into English and back into Azerbaijani using a high-quality neural translator, and then manually filtering for true paraphrase equivalence. During training, for each paraphrase pair (p, q), the encoder is encouraged to produce similar

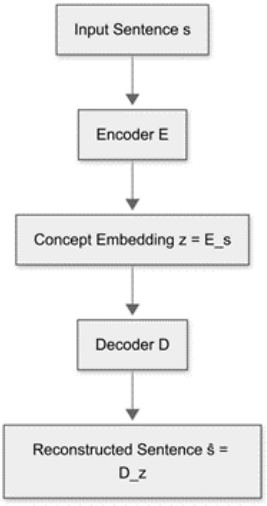


Fig. 1. Encoder-decoder autoencoder used to train the PaCE space.
Given a sentence s, the encoder outputs $z = (s)$, which the decoder then reconstructs as $\acute{s} = D(z)$

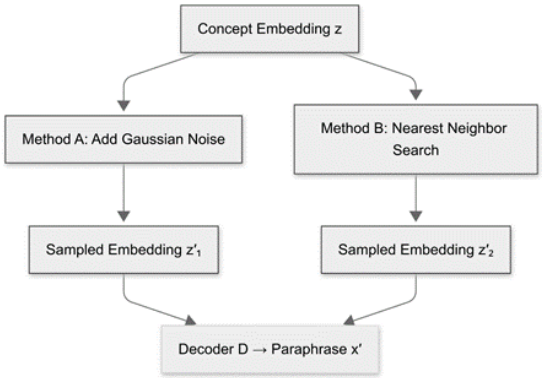


Fig. 2. New sentence embeddings z' are generated by sampling near
z – either by noise injection or by retrieving a nearby neighbor in embedding space.
These embeddings are decoded back into paraphrased sentences x'

embeddings $E(p)$ and $E(q)$. This can be done with a contrastive loss or a Siamese network setup: the distance $\left\|E\left(p\right)-E\left(q\right)\right\|_2$ for each paraphrase pair is minimized, while for non-paraphrase pairs (p, t) the distance could optionally be maximized or a margin used. In practice, a contrastive loss L_{para} defined as:

$$L_{para} = \sum_{(p,q)paraphrase} \left\|E\left(p\right)-E\left(q\right)\right\|_2^2,$$

is used, which pulls paraphrase embeddings together (it was found that explicitly pushing away non-paraphrase pairs was not necessary when combined with the autoencoder objective and the inherent separation of distinct sentences). The autoencoder reconstruction loss $L_{AE} = \sum_S L\left(D\left(E\left(S\right)\right), s\right)$ (where L is a token-level cross-entropy between output $\acute{s} = \left(D\left(E\left(S\right)\right)\right)$ and original) runs in parallel. The combined training objective is:

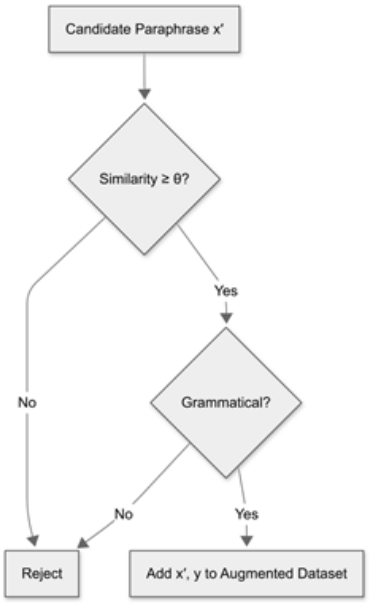


Fig. 3. Each generated paraphrase x' is checked for semantic similarity to the original and for grammatical correctness. If both conditions pass, it is added to the augmented dataset with the same label

$$L_{total} = L_{AE} + \lambda L_{para},$$

where λ is a weighting factor that balances reconstruction fidelity and paraphrase clustering. λ is chosen based on validation performance; it controls how strongly paraphrase similarity is enforced in the embedding space.

By training with this objective, the encoder E learns a vector space Z where sentences are embedded according to their semantic content. After training, if $E(s_1)$ and $E(s_2)$ are close, s_1 and s_2 are expected to be paraphrases. $E(s)$ is referred to as the PaCE model.

Paraphrase generation by concept embeddings

Once the concept embedding model E (and decoder D) is obtained, it is used to generate new sentences for data augmentation. Fig. 4 illustrates the overall PaCE augmentation pipeline. Starting with a labeled dataset, each sentence x is passed through the encoder E to produce a semantic embedding z. To generate paraphrases, nearby points z' are sampled around z using two methods: adding Gaussian noise or retrieving a nearest neighbor from existing embeddings. The first method encourages diversity; the second retrieves high-quality paraphrases if similar examples exist. Each z is decoded by the decoder D into a candidate paraphrase x'. A two-stage filter then ensures quality.

The method consists of the following stages:

- encoding sentences into semantic vectors,
- identifying their underlying concept,
- perturbing the vector to obtain a new point with similar meaning,
- decoding this into a paraphrased sentence,
- verifying semantic similarity and grammatical correctness.

The proposed method differs from prior approaches by explicitly modeling paraphrastic similarity in a learned semantic space, rather than relying on surface-level edits or translation-based transformations. Unlike synonym substitution, which often breaks grammaticality in morphologically rich languages, or back-translation, which introduces uncontrolled variations, PaCE generates fluent paraphrases with preserved meaning through concept-level perturbations.

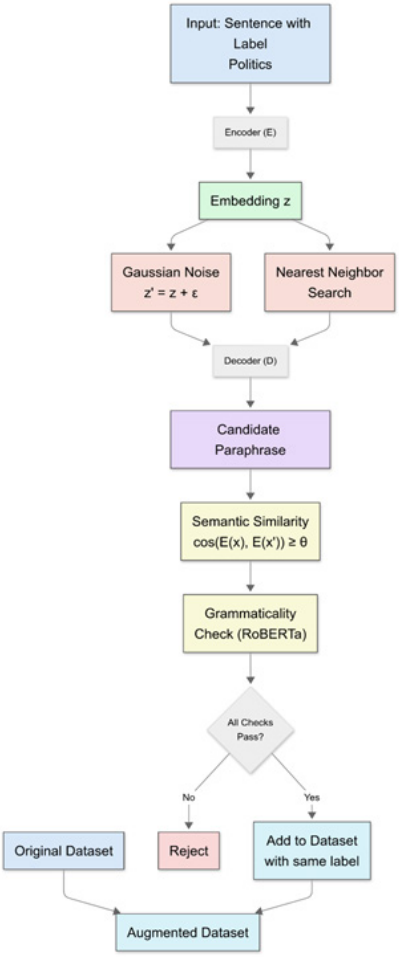


Fig. 4. PaCE augmentation pipeline: encoding, sampling, decoding, filtering and dataset expansion

First, cosine similarity between $\cos\left(E\left(x'\right),E\left(x\right)\right)$ must exceed a threshold ($\theta = 0.8$) to retain semantic meaning.

Second, grammaticality is checked using a RoBERTa-based perplexity filter trained on Azerbaijani. We compute pseudo-perplexity scores using a RoBERTa [21] model fine-tuned on Azerbaijani corpora. Given that RoBERTa is a masked language model (MLM), we adopt the pseudo-perplexity approach [22], where each token in a sentence is masked individually and the model predicts the masked token based on its context. The average log-likelihood across all tokens provides a pseudo-perplexity score, which serves as a proxy for grammaticality. Sentences with scores exceeding a predefined threshold are filtered out to ensure grammatical correctness.

Only paraphrases passing both checks are added to the augmented dataset with the same label y. This enriched dataset is then used to train the final classifier.

Implementation details

The models were built using the PyTorch deep learning framework and HuggingFace Transformers library for ease of implementation. The concept embedding model's encoder and decoder were initialized from a multilingual pre-trained model (mBART50) which is trained for many-to-many translation including Azerbaijani; this provided a strong starting point for Azerbaijani encoding/decoding. This model was fine-tuned on Azerbaijani autoencoding and paraphrase objectives. This cross-utilization of a translation model for paraphrasing is an example of transfer learning and it aligns with the idea of bilingual pivoting – the mBART model's latent space already has some notion of aligning Azerbaijani with other

languages, aiding concept space learning. The classifier model for evaluation was a RoBERTa-based Azerbaijani language model (pre-trained on news data [1]) fine-tuned on the specific classification task. All hyperparameters (such as learning rates, noise levels, thresholds) were optimized on a development set.

By encapsulating this methodology as a software toolkit, a reusable augmentation module for Azerbaijani NLP tasks is contributed. The entire PaCE pipeline – from concept embedding training to dataset augmentation – represents a form of software support for computing systems handling language data. It is essentially an add-on component that can integrate with existing NLP training workflows, providing a mathematical and algorithmic enhancement to the data processing stage.

Experiments and results

The effectiveness of PaCE augmentation is evaluated on a text classification task in Azerbaijani.

Dataset

For the experiments, a publicly available Azerbaijani news classification dataset, AZERNEWS, derived from the Azertac news agency corpus [1] is used. It consists of news article sentences labeled by category.

Dataset statistics

The subset used in the experiments contains 10000 labeled instances in total. The distribution is somewhat imbalanced: Politics (3500 sentences), Economy (2800), Sports (1700), Culture (2000). A class-balanced evaluation set was maintained to fairly assess performance across categories. The data was split into 8000 training examples, 1000 validation and 1000 test. The average sentence length is about 15 words (with significant variance, as news sentences can range from short headlines to longer explanatory sentences).

Before augmentation, basic text preprocessing was performed: all text was lowercased (Azerbaijani is typically written in Latin script with special characters like ə, ı, etc., which were preserved) and some typical OCR or spelling errors found in the dataset were corrected.

Models and training

Baseline classifier

As the baseline model for classification, a pre-trained multilingual RoBERTa model that had been further tuned on Azerbaijani news [1] was used. This model, referred to as Az-RoBERTa, is an encoder-only transformer model capable of producing contextualized embeddings for Azerbaijani text. A classification layer was added on top of Az-RoBERTa to predict the news category label. This baseline already leverages transfer learning (from unlabeled news data via RoBERTa pre-training), which is known to improve performance in low-data regimes. However, it is expected that data augmentation can further improve results by providing additional labeled variations.

Augmentation strategies compared

The following training setups are compared:

a. No Augmentation (Baseline) – training on the original 8k training sentences only.

b. Synonym Augmentation – a lexical augmentation method akin to EDA [9] is applied, replacing 1–2 nouns or adjectives in each sentence with synonyms. Since Azerbaijani lacks a WordNet, a small synonym dictionary was built from a bilingual English-Azerbaijani dictionary: English WordNet synonyms were mapped to their Azerbaijani translations where possible. This provided a limited resource to replace some words (for example, “böyük” could be replaced with “iri” for “big/large”). One augmented sentence per original was generated with this method.

c. Back-Translation Augmentation – the Facebook mBART50 model is used to translate each Azerbaijani sentence to English and back to Azerbaijani generating one paraphrase per sentence.

d. PaCE Augmentation – using the concept embedding method, up to 2 paraphrases per sentence are generated as described in Fig. 4.

All augmented datasets (b, c and d) roughly double the amount of training data (to ~16k instances, except synonym augmentation which resulted in slightly fewer augmentations for some sentences where no synonym was found). The Az-RoBERTa classifier is trained on each augmented training set with the same hyperparameters as the baseline for a fair comparison.

Model performance was evaluated on the held-out test set of 1000 instances, using accuracy and macro-averaged F1 score as the primary metrics. Accuracy measures overall correctness, while macro-F1 gives equal weight to each class, which is important given the class imbalance. We also report per-class precision and recall to understand where improvements are coming from. All results are averaged over three training runs with different random seeds to ensure robustness; we report the mean and standard deviation. We perform statistical significance testing (paired t-test) between the baseline and PaCE-augmented model to verify if improvements are significant.

The results of text classification across four categories are presented in Table 1. The baseline model (without augmentation) already achieves decent accuracy, considering the use of a pre-trained model. However, augmentation methods yield clear improvements. Augmentation using PaCE shows the best results, significantly outperforming both synonym-based and back-translation augmentation.

Table 1

Classification performance with different training data augmentations

Training data	Accuracy (%)	Macro-F1 (%)	Politics F1	Economy F1	Sports F1	Culture F1
No augmentation (8k)	76.8 ± 0.5	74.3 ± 0.6	78.1	72.5	69.0	77.5
+Synonym augment (EDA)	79.4 ± 0.7	76.1 ± 0.8	80.0	75.0	71.2	78.5
+ Back-translation	81.0 ± 0.6	78.0 ± 0.5	82.3	77.4	74.1	78.2
+ PaCE augment	**84.5 ± 0.4**	**81.7 ± 0.5**	85.9	80.5	78.3	82.0

As shown in Table 1, performance is consistently improved by augmenting the data over the no-augmentation baseline. Synonym replacement provides a modest boost of around 2.6% in accuracy, which indicates that even simple lexical variety helps the model generalize better. Back-translation performs better, with about 4.2% accuracy gain over baseline, likely because the paraphrases generated are more diverse and contextually richer than the limited synonym sets. The proposed PaCE augmentation delivers a further jump, achieving 84.5% accuracy, approximately 7.7% higher than the baseline and 3.5% higher than back-translation. The macro-F1 score shows a similar trend, with PaCE > back-translation > synonym > baseline. These improvements are found to be statistically significant ($p < 0.01$ for PaCE vs baseline, and $p < 0.05$ for PaCE vs back-translation, under a paired t-test across the three runs).

In terms of per-category performance, F1 scores across all news categories were improved by PaCE augmentation, with the largest gains in the Sports category (+9.3 points over baseline F1) and Economy (+8.0 points). These two categories had relatively fewer training examples initially, so the additional paraphrased examples had a pronounced effect on the model's ability to recognize varied expressions of sports and economic news. For instance, in Sports, the baseline might have learned keywords like "qalib gəldi" ("won") or "oyun" ("game"), but with augmentation, alternative phrasings like "məhz qazandı" ("secured victory") or "qarşılaşma" ("match") were also seen, reducing the model's reliance on any single phrasing. The Culture category, interestingly, showed a smaller improvement (F1 from 77.5 to 82.0) compared to others; this could be because the model already performed well on Culture, or because some paraphrases in cultural context (e.g., names of artistic works or terms) are harder to generate without loss of meaning, so augmentation helped slightly less. Nevertheless, every category saw an increase in F1, indicating that PaCE augmentation is broadly effective and not limited to specific content.

Semantic similarity between original sentences and their PaCE-generated paraphrases was also measured to ensure that augmentation did not drift from the intended meaning. Using the encoder EE cosine similarity, the average similarity was 0.89 for accepted paraphrases (by design it had to be ≥ 0.8), compared to 0.95 for the trivial identity paraphrase. For back-translation outputs, an average similarity of 0.83 was measured, confirming that PaCE's filtering indeed produced paraphrases that were closer in meaning to the source than unfiltered back-translations. This likely contributed to the classifier's superior performance, as training data augmented with PaCE had less noise (fewer label inconsistencies or off-topic sentences).

Overall, the experimental results confirm that PaCE augmentation leads to superior model performance on our Azerbaijani classification task. By injecting diverse yet semantically consistent training examples, the model generalizes better and is more robust to linguistic variations. In the next section, we delve deeper into the implications of these results, analyze why PaCE outperforms the alternatives, and discuss any limitations observed.

Conclusion

This paper proposed PaCE augmentation, a novel data augmentation method for Azerbaijani NLP. The method differs from existing approaches by operating at the sentence level through semantic concept embeddings, ensuring paraphrased outputs maintain full semantic coherence and grammatical correctness, crucial for morphologically rich languages like Azerbaijani.

The experiments on Azerbaijani news text classification demonstrated that PaCE significantly improves performance, achieving a 7–8% absolute accuracy gain over strong baselines and conventional methods like synonym replacement and back-translation. The method consistently enhanced model robustness across multiple categories, effectively addressing both data scarcity and linguistic variability.

PaCE's key novelty is bridging representation learning and data augmentation, enabling controlled and meaningful paraphrase generation without external lexical databases or translation tools. This results in high-quality, natural-language paraphrases inspectable by engineers, unlike vector perturbation methods (e.g., RPN). Additionally, PaCE is task-agnostic and thus broadly applicable across various NLP applications. A specific mathematical formulation was developed, and a corresponding software component implemented. Practically, NLP engineers can readily incorporate PaCE into training pipelines to enhance system performance and precision.

The approach is applicable to other low-resource languages, given minimal paraphrase training data, making it valuable beyond Azerbaijani. However, a distinctive advantage for Azerbaijani is the method's natural handling of its complex morphology, ensuring grammatical accuracy in augmented sentences.

In conclusion, PaCE augmentation provides a significant methodological advancement for low-resource NLP, particularly for Azerbaijani, encouraging further exploration and broader integration into NLP workflows.

REFERENCES

1. **Ziyaden A., Yelenov A., Hajiyev F., Rustamov S., Pak A.** Text data augmentation and pre-trained Language Model for enhancing text classification of low-resource languages. *PeerJ Computer Science*, 2024, Art. no. 10:e1974. DOI: 10.7717/peerj-cs.1974

2. **Aghaev A.F., Molodyakov S.A.** Lemmatization of nouns in the Azerbaijani language. *Modern Science: Actual Problems of Theory & Practice*, 2023, No. 7, Pp. 12–17. DOI: 10.37882/2223-2966.2023.07.01

3. **Devlin J., Chang M.-W., Lee K., Toutanova K.** BERT: Pre-training of deep bidirectional transformers for language understanding. *arXiv:1810.04805*, 2019. DOI: 10.48550/arXiv.1810.04805

4. **Feng S.Y., Gangal V., Wei J., Chandar S., Vosoughi S., Mitamura T., Hovy E.** A survey of data augmentation approaches for NLP. *arXiv:2105.03075*, 2021. DOI: 10.48550/arXiv.2105.03075

5. **Taylor L., Nitschke G.** Improving deep learning with generic data augmentation. *2018 IEEE Symposium Series on Computational Intelligence* (*SSCI*), 2018, Pp. 1542–1547. DOI: 10.1109/SSCI.2018.8628742

6. **Mikołajczyk A., Grochowski M.** Data augmentation for improving deep learning in image classification problem. *2018 International Interdisciplinary PhD Workshop* (*IIPhDW*), 2018, Pp. 117–122. DOI: 10.1109/IIPHDW.2018.8388338

7. **Bao F., Neumann M., Vu N.T.** CycleGAN-based emotion style transfer as data augmentation for speech emotion recognition. *Proceedings Interspeech 2019*, 2019, pp. 2828–2832. DOI: 10.21437/Interspeech.2019-2293

8. **Wen Q., Sun L., Yang F., Song X., Gao J., Wang X., Xu H.** Time series data augmentation for deep learning: A survey. *arXiv:2002.12478*, 2020. DOI: 10.48550/arXiv.2002.12478

9. **Wei J., Zou K.** EDA: Easy data augmentation techniques for boosting performance on text classification tasks. *arXiv:1901.11196*, 2019. DOI: 10.48550/arXiv.1901.11196

10. **Miller G.A.** WordNet: a lexical database for English. *Communications of the ACM*, 1995, Vol. 38, No. 11, Pp. 39–41. DOI: 10.1145/219717.219748

11. **Amin M., Anselma L., Mazzei A.** Data augmentation for low-resource Italian NLP: Enhancing semantic processing with DRS. *Proceedings of the 10th Italian Conference on Computational Linguistics* (*CLiC-it 2024*), 2024, Pp. 29–38.

12. **Goodfellow I.J., Pouget-Abadie J., Mirza M., Xu B., Warde-Farley D., Ozair S., Courville A., Bengio Y.** Generative adversarial nets. *Advances in Neural Information Processing Systems* (*NIPS 2014*), 2014, Vol. 27, Pp. 2672–2680.

13. **Yuan Z., Zhao Z., Wang Y., Hou X., Xue H., Zhao Z., Liu Y.** RPN: A word vector level data augmentation algorithm in deep learning for language understanding. *arXiv:2212.05961*, 2022. DOI: 10.48550/arXiv.2212.05961

14. **Sennrich R., Haddow B., Birch A.** Improving neural machine translation models with monolingual data. *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics*, 2016, Vol. 1, Pp. 86–96. DOI: 10.18653/v1/P16-1009

15. **Chen L., Dai S., Tao C., Shen D., Gan Z., Zhang H., Zhang Y., Carin L.** Adversarial text generation via feature-mover's distance. *arXiv:1809.06297*, 2018. DOI: 10.48550/arXiv.1809.06297

16. **De Rosa G.H., Papa J.P.** A survey on text generation using generative adversarial networks. *Pattern Recognition*, 2021, Vol. 119, Art. no. 108098. DOI: 10.1016/j.patcog.2021.108098

17. **Guo H., Mao Y., Zhang R.** Augmenting data with Mixup for sentence classification: An empirical study. *arXiv:1905.08941*, 2019. DOI: 10.48550/arXiv.1905.08941

18. **Bannard C., Callison-Burch C.** Paraphrasing with bilingual parallel corpora. *Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics* (*ACL'05*), 2005, Pp. 597–604. DOI: 10.3115/1219840.1219914

19. **Ganitkevitch J., Van Durme B., Callison-Burch C.** PPDB: *The paraphrase database. Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 2013, Pp. 758–764.

20. **Wieting J., Bansal M., Gimpel K., Livescu K.** Towards universal paraphrastic sentence embeddings. *arXiv:1511.08198*, 2015. DOI: 10.48550/arXiv.1511.08198

21. **Liu Y., Ott M., Goyal N., Du J., Joshi M., Chen D., Levy O., Lewis M., Zettlemoyer L., Stoyanov V.** RoBERTa: A robustly optimized BERT pretraining approach. *arXiv:1907.11692*, 2020. DOI: 10.48550/arXiv.1907.11692

22. **Salazar J., Liang D., Nguyen T.Q., Kirchhoff K.** Masked language model scoring. *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, 2020, pp. 2699–2712. DOI: 10.18653/v1/2020.acl-main.240

INFORMATION ABOUT AUTHORS / СВЕДЕНИЯ ОБ АВТОРАХ

Aslan F. Aghayev
Агаев Аслан Фахри оглы
E-mail: agaev.af@edu.spbstu.ru

Sergey A. Molodyakov
Молодяков Сергей Александрович
E-mail: samolodyakov@mail.ru
ORCID: https://orcid.org/0000-0003-2191-9449

Sergey M. Ustinov
Устинов Сергей Михайлович
E-mail: usm50@yandex.ru
ORCID: https://orcid.org/0000-0003-4088-4798

Submitted: 15.01.2025; Approved: 11.04.2025; Accepted: 25.04.2025.
Поступила: 15.01.2025; Одобрена: 11.04.2025; Принята: 25.04.2025.

Computing, Telecommunication and Control, 2025, Vol. 18, No. 3, Pp. 58–67.
Информатика, телекоммуникации и управление. 2025. Том 18, № 3. С. 58–67.

Research article
DOI: https://doi.org/10.18721/JCSTCS.18305
UDC 004.89

METHOD FOR AUTOMATED ENRICHMENT OF A KNOWLEDGE BASE ON GLASS COMPOSITIONS AND PROPERTIES BASED ON DATA FROM SCIENTIFIC PUBLICATIONS

E.A. Pavlov ✉ , *P.D. Drobintsev, V.A. Klinkov, A.V. Semencha, I.G. Chernorutskiy*

Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russian Federation

✉ pavlov_ea@spbstu.ru

Abstract. Automating the extraction of glass composition and property data from scientific literature is critically important for accelerating the development of new material. This work presents a method integrating: 1) the collection of full-text articles using the Elsevier Research Products APIs, 2) text preprocessing, 3) context-dependent extraction of structured data using a large language model (LLM) and a domain-specific prompt, 4) enrichment of a knowledge base on glasses. The key achievement is the development of a prompt that yields an F1-score of 0.99 for extracting chemical compositions, their properties and correctly establishing relationships between them on a sample of 50 articles. The proposed method significantly simplifies the automatic creation and continuous updating of knowledge bases on glasses, thereby eliminating the traditional reliance on manually curated, potentially outdated resources and providing a robust, data-driven foundation for the efficient designing of glasses with target properties using machine learning.

Keywords: data extraction, natural language processing, LLM, prompt engineering, knowledge base, glass, glass properties

Citation: Pavlov E.A., Drobintsev P.D., Klinkov V.A. et al. Method for automated enrichment of a knowledge base on glass compositions and properties based on data from scientific publications. Computing, Telecommunications and Control, 2025, Vol. 18, No. 3, Pp. 58–67. DOI: 10.18721/JCSTCS.18305

Научная статья
DOI: https://doi.org/10.18721/JCSTCS.18305
УДК 004.89

МЕТОД АВТОМАТИЗИРОВАННОГО ПОПОЛНЕНИЯ БАЗЫ ЗНАНИЙ О СОСТАВАХ И СВОЙСТВАХ СТЕКОЛ НА ОСНОВЕ ДАННЫХ ИЗ НАУЧНЫХ ПУБЛИКАЦИЙ

Е.А. Павлов ✉ , *П.Д. Дробинцев, В.А. Клинков, А.В. Семенча, И.Г. Черноруцкий*

Санкт-Петербургский политехнический университет Петра Великого, Санкт-Петербург, Российская Федерация

✉ pavlov_ea@spbstu.ru

Аннотация. Автоматизация извлечения данных о составах и свойствах стекол из научной литературы критически важна для ускорения разработки новых материалов. В работе представлен метод, интегрирующий: 1) сбор полнотекстовых статей с помощью Elsevier Research Products APIs, 2) предобработку текста, 3) контекстно-зависимое извлечение структурированных данных с помощью большой языковой модели (LLM) и доменно-специфичного промпта, 4) пополнение базы знаний о стеклах. Ключевым достижением стала разработка промпта, обеспечивающего точность F1=0,99 для извлечения химических составов и их свойств, а также корректного установления связей между ними на выборке из 50 статей. Предлагаемый метод значительно упрощает автоматическое создание и непрерывное обновление баз знаний о стекле, тем самым устраняя традиционную зависимость от вручную отобранных, потенциально устаревших ресурсов и обеспечивая надежную, управляемую данными основу для эффективного проектирования стекол с заданными свойствами с помощью машинного обучения.

Ключевые слова: извлечение данных, обработка естественного языка, большая языковая модель, промпт-инжиниринг, база знаний, стекло, свойства стекла

Для цитирования: Pavlov E.A., Drobintsev P.D., Klinkov V.A. et al. Method for automated enrichment of a knowledge base on glass compositions and properties based on data from scientific publications // Computing, Telecommunications and Control. 2025. Т. 18, № 3. С. 58–67. DOI: 10.18721/JCSTCS.18305

Introduction

Contemporary materials science research, particularly in designing functional glasses for optics, electronics and energy, faces an exponential growth of published data. Critical data on chemical compositions and physicochemical properties needed for predicting and designing new glasses remain “locked” in unstructured text of scientific publications [1]. Manual data collection and sorting requires a lot of labor-intensive: an expert may require up to one hour to thoroughly analyze a single article. This approach not only slows progress but also makes real-time analysis of thousands of publications impossible. Extracting data for multicomponent glass systems is especially challenging, where compositions are described in diverse, often non-standard formats (linear combinations, variable-based formulas, at.%/mol.% etc.) and property values heavily depend on synthesis and measurement methods. This specificity hinders traditional automated extraction methods. Large-scale, well-characterized data is fundamental for ensuring the accuracy and reliability of material inferences [2]. For example, the SciGlass database [3], long considered the gold standard for glass properties, is no longer updated, rendering it unsuitable for analyzing modern research. The lack of up-to-date structured glass composition and property data impedes the development of new materials with tailored properties. Given the rapidly expanding volume of scientific literature on glasses, automated extraction of composition and property data has become increasingly vital.

Breakthroughs in natural language processing (NLP), driven by the transformer architecture and attention mechanisms [4], have opened new avenues for textual data analysis [5–10]. Large language models (LLMs) like GPT [11], BERT [12] and Falcon [13] demonstrate unique capabilities in identifying semantic relationships and extracting hidden patterns from weakly structured texts. LLM-based methods often surpass classical NLP approaches in tasks requiring deep contextual understanding. Traditional entity extraction pipelines (e.g., OSCAR [14], ChemicalTagger [15], ChemDataExtractor [16]) are effective for text mining in chemistry and materials science but are limited by rigid templates and poor adaptation to semantic nuances. A key advantage of LLMs – particularly valuable for domains with complex data structures – is flexible extraction adaptation via prompt engineering techniques, enabling task-specific tailoring without resource-intensive fine-tuning [17, 18].

Current research confirms LLM effectiveness for scientific publication data extraction [19–21]. For example, in [19], BERT is refined on battery publications, combining question-answering with classical NLP to extract battery components (cathode, anode, electrolyte) and link them to characteristics like capacity and cycle stability, improving specialized database population accuracy by 18%. In [20], a ChemPrompt Engineering approach is proposed using ChatGPT to automate extraction of metal-organic framework (MOF) synthesis conditions. In [21], a pipeline for extracting reticular material (MOF, COF) synthesis parameters from Elsevier API-sourced PDFs using language models is implemented. While existing studies confirm LLM efficacy for chemistry and materials science data extraction, current solutions either require domain-specific model fine-tuning or lack adaptation to key features of glass system descriptions, such as diverse composition formats, mandatory unit handling and measurement method dependencies.

This article describes an automated knowledge base enrichment method combining an LLM guided by a specially designed domain-specific prompt without additional training or fine-tuning. The method enables intelligent digital platforms capable of near-real-time analysis of thousands of publications. This is particularly relevant for AI-driven design of glasses with target properties, where data quality and volume directly impact accuracy.

Method

The ultimate goal of this work is to create a continuously updated knowledge base on glasses. The target data schema is designed to store not only individual facts but also the complex relationships critical for analysis and predictive modeling. Its core consists of the following entities: detailed records of chemical compositions, specifying the concentration type (at.%, wt.%, mol.%), material properties, each accompanied by a value, unit of measurement and, crucially, the measurement method, ensuring correct comparison of data from different sources and enriched publication metadata (authors, journal, year). To transform the unstructured text of scientific articles into such a relational structure, an end-to-end pipeline was developed. Its key task is not merely to extract mentions of substances and numbers, but to correctly interpret them in context, identify relationships between them and package them into a strict, formalized format (JSON) suitable for automatic analysis using the following pipeline.

Pipeline

The proposed method for automated glass property knowledge base formation is implemented as follows (Fig. 1):

1. Article Retrieval: Full-text articles in XML format are downloaded from ScienceDirect via Elsevier Research Products APIs.

2. Preprocessing: Article text is processed. Only main text and metadata are extracted, with irrelevant elements removed (figure references, bibliographies, etc.).

3. LLM Text Processing: Preprocessed text is fed into an LLM with a prompt generating structured JSON of a specified format.

4. Postprocessing: The LLM response is processed, and extracted information is saved to a database.

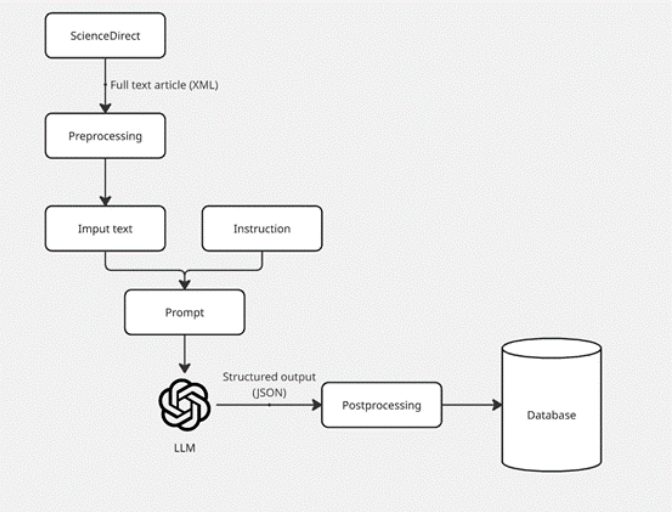


Fig. 1. Automated knowledge base formation pipeline

Article extraction via Elsevier API

The Elsevier API is utilized to access full-text scientific publications from the ScienceDirect library to extract information on glass compositions and properties. When developing automated scientific article processing systems using language models, the selection of an input data format that ensures accurate interpretation of information, particularly tables, is crucial. Although several formats (JSON, PDF, XML, TEXT) are available via the Elsevier API, the XML format was selected in this study. This choice was driven by the issue of incorrect recognition of missing values in tables encountered when using other data formats, which adversely affected information extraction quality.

To collect relevant publications, a search query was executed via the Elsevier API, incorporating the following keyword combination: “glass”, “chalcogenide glass”, “As-Se”, “Tg”, “density”. To minimize noise and enhance the relevance of input data fed into the language model, article text preprocessing was implemented. This preprocessing involves removing information unrelated to material compositions and properties. Specifically, only tables and the main body text are extracted from the source XML. This approach is necessary to ensure the model focuses on the core article content – where methods, results and discussion are presented – and to reduce the volume of input data processed by the model. The output is a cleaned text containing exclusively data relevant for analysis.

Information extraction

The Qwen 3 LLM was selected for automating structured glass composition/property extraction. Qwen 3 employs a hybrid approach with two modes:

- Thinking Mode: The network spends time reasoning step-by-step before final output.

This mode is essential for decomposing intricate material descriptions, resolving ambiguities in terminology or units, and inferring implicit relationships between composition and properties – common challenges in scientific text extraction.

- Non-Thinking Mode: Delivers rapid “near-instant” responses (suited for simple queries prioritizing speed).

Qwen 3 offers revolutionary advances in logical reasoning, instruction following, agent capabilities and multilingual support. The explicit reasoning chain in Thinking Mode significantly enhances the reliability and accuracy of extracted data from dense, research papers. It is open-source and free to use.

Prompt engineering

Prompt engineering is crucial for extracting glass composition/property data from publications. It precisely formulates queries for LLMs, minimizing hallucination risks [22] and improving accuracy. For

complex data structures, prompts must explicitly instruct the LLM to output extracted information in a strict format. JSON is optimal due to its ubiquity in LLM training data, ensuring correct structuring and simplifying downstream processing. Thus, effective prompt design optimizes extraction and reduces error risks.

An iterative prompt development approach was applied. The prompt underwent several refinement cycles using a validation set of 10 articles, which were excluded from the final test sample. The key stages of this refinement process are described below. The initial prompt defined for the LLM requested extraction of all glass compositions and properties in JSON format.

Initial prompt:

Analyze the provided text and extract information about the compositions and their properties. Present the result in JSON format, grouping the properties by each composition. If the text mentions multiple compositions, create a separate object for each one. For each composition, specify its name and a list of properties, where each property contains a name and a value (if specified). If the property value is not specified, leave the field empty or use null.

This prompt formulation revealed two key limitations:

1. Inconsistent JSON formatting. The model returned JSON objects with varying structures, complicating subsequent processing.

2. Incorrect handling of chemical formulas. The model failed to interpret complex composition expressions, such as $(100 - x)[70GeO_2 - 12BaF_2 - 10Ga_2O_3 - 3BaO - 3La_2O_3 - 2Y_2O_3] - xDy_2O_3$, and to calculate elemental composition percentages.

To overcome these issues, the prompt was significantly enhanced. The core modifications included:

• Explicit instructions for composition parsing.

The LLM was explicitly tasked to identify different compositional formats (e.g., linear combinations, percentage-based notations) and to calculate elemental percentages using provided formulas. A critical addition was the requirement to determine and output the type of percentage (wt%, mol%, at%) for each composition, as this is essential for interpreting the data correctly.

...

Instructions:

1. Composition Identification

– Detect all compositional formats:

a) Linear combination: $(1 - x)\cdot$[Base] + $x\cdot$[Additive] (e.g., "$(1 - x)\cdot Ge_{20}As_{20}Se_{40} + x\cdot Te$")

b) Percentage-based notation: $Ge_{12}As_{24}Se_{64}$ or $As_{30}Se_{40}Te_{30}$

– Determine percentage type (wt/atomic/mol.%)

– For linear combinations:

** Calculate wt/atomic/mol.% using:*

`Element% = (1 – x)\*(base_%) + x\*(additive_%)`

** Round to 2 decimal*

– For percentage-based notation: Keep original wt/atomic/mol. percentages

...

• Strict JSON schema definition.

A comprehensive JSON example was embedded into the prompt to enforce a consistent output structure. The schema requires each composition entry to include the following keys:

• “type”: “calculated” for expressions requiring computation (e.g., linear combinations), “raw” for direct notations.

• “percentage_type”: The unit of the composition's concentration (“mol%”, “wt%”, “at%”).

• “formula”: The original compositional expression as found in the text.

• “x”: The value of the parameter x for parametrized compositions.

• “composition”: A key-value object where each key is a chemical element/compound and its value is the calculated or extracted percentage.

• “properties”: A key-value object where each key is a property name. Each property must be an object containing the keys “value” and “unit”.

...

JSON Structure

```
“‘json
{{
"compositions": {
"GAST-1": {
"type": "calculated",
"percentage_type": "mol%",
"formula": "(1 − 0.2)·Ge25As25Se50 + 0.2·Te",
"x": 0.2,
"composition": {"Ge": 20.0,"As": 20.0,"Se": 40.0,"Te": 20.0},
"properties": {"density": {"value": 3.45,"unit": "g/cm3" },
"Tg": {"value": 285,"unit": "°C" }}
},
```

...

However, this prompt formulation also proved suboptimal. Further testing revealed the following issues:

• Hallucination on missing compositions.

When no chemical compositions were found in the article text, the LLM occasionally hallucinated by returning the example response provided in the prompt.

• Omission of duplicate properties.

The LLM skipped properties with identical names. For instance, if glass transition temperature was measured using multiple methods in an article, only one value was extracted. The same behavior occurred with properties like refractive index.

These issues were resolved by adding explicit instructions to the prompt:

• Return empty JSON if no compositions are detected.

...

4. JSON Structure
– If NO COMPOSITIONS detected: return {}
– For each detected composition:
“‘json

...

• For multiple measurements of the same property, require the LLM to include the full property name and measurement method in its output.

...

– Include the measurement method from the text. If method is unspecified → "measurement_method": "Not specified".
– Map abbreviations to full names (e.g., "Eg" → "Optical Bandgap", "HV" → "Vickers Hardnes")
– Format properties as key-value pairs:

```
“‘json
"properties": {"property_abbreviation": {
"property_full_name": "...",
"value": ...,
"unit": "...",
"measurement_method": "..."}}“‘
```

...

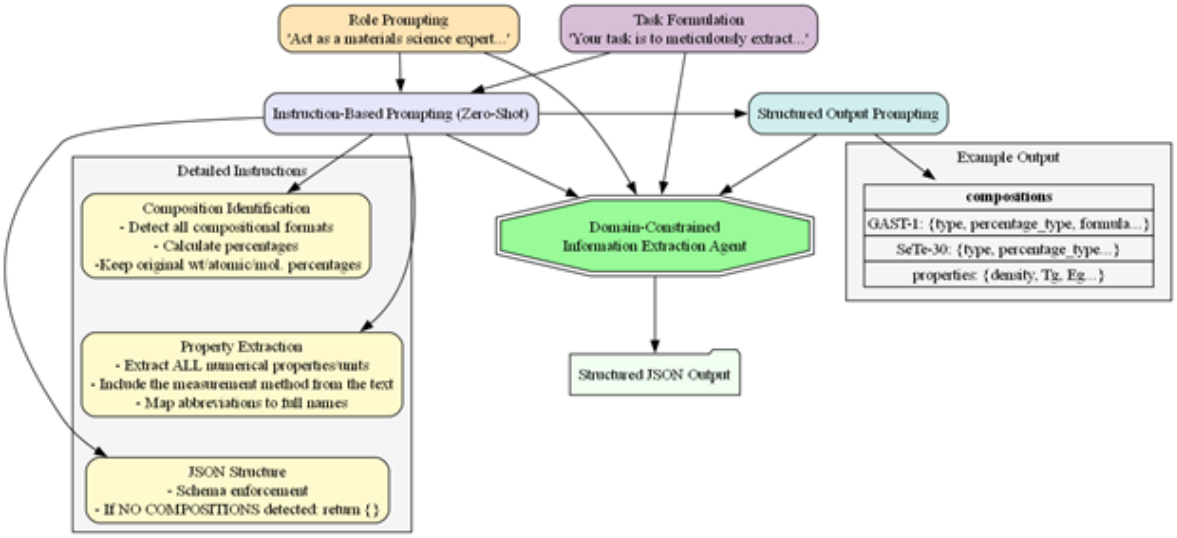


Fig. 2. Prompting architecture

Consequently, the final prompt for extracting glass compositions and properties was formulated and is available in the GitHub repository associated with this work[1].

The extraction of precise compositional and property data from unstructured glass science literature necessitates overcoming significant challenges, including variable notation formats, domain-specific terminology and specific requirements for outputs. To address these, a multi-faceted prompt engineering strategy was implemented, combining established techniques to optimize LLM performance.

Role prompting was employed to instill domain-specific expertise, achieved through the explicit directive: “Act as a materials science expert specializing in glasses”. A precise task formulation followed, stating: “Your task is to meticulously extract compositional data and property values from research papers”. This explicit framing focused the model’s processing exclusively on targeted data extraction, minimizing diversion toward ancillary content within source texts. Instruction-based prompting under a zero-shot paradigm provided detailed procedural rules without in-context examples. Structured output prompting enforced rigorous schema compliance through a predefined JSON structure. The synergistic integration of these techniques transformed the LLM into a domain-constrained information extraction agent, as visualized in Fig. 2. Role prompting established foundational expertise, task formulation defined the operational objective, instruction-based rules governed the technical execution, and structured output prompting ensured machine-processable structured output. This layered methodology enabled reliable, automated curation of materials data from complex research texts.

Postprocessing

The JSON response from the LLM is merged with article metadata extracted from the source XML text and undergoes further processing before storage in the knowledge base. This processing includes JSON parsing, data validation, linking of the LLM’s JSON response to article metadata and persistence of validated records to a relational database.

Core stored entities comprise:

1. Publication Metadata:
    - Article title
    - Author list
    - Journal name
    - Publication year

[1] https://github.com/EvgenDI/automated-glass-data-extraction

- Page range

2. Extracted Data:
   - Chemical compositions
   - Properties (value, unit of measurement, measurement method, full descriptive name)

Results and discussion

To evaluate the quality of glass composition and property extraction from scientific publications, 50 articles were selected. A materials science expert manually extracted glass composition data, corresponding properties, and their relationships from these articles, creating a benchmark dataset containing 253 compositions and 1685 properties.

Testing employed a local Qwen-3-14B model with chain-of-thought reasoning enabled. The model was deployed on a system with 1× Tesla A100 80GB GPU, 8 CPU cores, and 243 GB RAM. Standard information extraction metrics were used for robust and standardized evaluation: precision, recall and F1-score.

– Precision reflects the proportion of relevant information among all extracted data. High precision indicates most extracted information is correct (few false positives).

– Recall indicates the proportion of available relevant information successfully extracted. High recall signifies minimal omission of relevant information (few false negatives).

– F1-score is the harmonic mean of precision and recall. This metric provides a balanced performance assessment by considering both extraction accuracy and completeness.

Results (Table 1) confirm the method's high reliability:

- Composition identification precision: 100% (all extracted compositions/properties matched the benchmark)
- Only 1% of glass compositions were omitted.
- Correct relationships were established for 99% of accurately extracted compositions and properties (pair F1-score = 0.99).

This relational accuracy is critical for building trustworthy knowledge bases.

Table 1

Information extraction quality assessment

Metric	Compositions	Properties	Composition/Property
Precision	1	1	0.99
Recall	0.99	0.99	0.99
F1	0.99	0.99	0.99

The high F1-score underscores the potential for automating glass data extraction, reducing manual processing time. These results demonstrate that modern LLMs can significantly accelerate expert knowledge extraction while maintaining high reliability.

Conclusion

This study developed a method that integrates automated full-text article extraction from ScienceDirect with the application LLM and engineered domain-specific prompt. The developed method demonstrated high accuracy in extracting structured data on glass composition and properties from scientific publications. Testing on a separate set of 50 articles yielded an F1-score of 0.99, demonstrating its high efficacy. The capabilities of the method significantly facilitate the transition from manual data collection to automated knowledge extraction. This achievement paves the way for creating self-learning systems capable of accelerating the development of glasses with targeted properties.

REFERENCES

1. **Krallinger M., Rabal O., Lourenço A., Oyarzabal J., Valencia A.** Information retrieval and text mining technologies for chemistry. *Chemical Reviews*, 2017, Vol. 117, No. 12, Pp. 7673–7761. DOI: 10.1021/acs.chemrev.6b00851

2. **Hill J., Mulholland G., Persson K., Seshadri R., Wolverton C., Meredig B.** Materials science with large-scale data and informatics: Unlocking new opportunities. *MRS Bulletin*, 2016, Vol. 41, Pp. 399–409. DOI: 10.1557/mrs.2016.93

3. **Mazurin O.V., Leko V.K., Streltsina M.V., Shvayko-Shvaykovskaya T.P.** Sovremennoe sostoianie bazy dannykh informatsionnoi sistemy SciGlass v oblasti opticheskikh kharakteristik stekol [The current state of the SciGlass information system database in the field of optical characteristics of glasses]. *Nauchno-tekhnicheskii vestnik sankt-peterburgskogo gosudarstvennogo universiteta informatsionnykh tekhnologii, mekhaniki i optiki* [*Scientific and Technical Bulletin of Information Technologies, Mechanics and Optics*], 2004, No. 13, P. 253.

4. **Foppiano L., Lambard G., Amagasa T., Ishii M.** Mining experimental data from materials science literature with large language models: an evaluation study. *Science and Technology of Advanced Materials: Methods*, 2024, Vol. 4, No. 1, Art. no. 2356506. DOI: 10.1080/27660400.2024.2356506

5. **Peng R., Liu K., Yang P., Yuan Z., Li S.** Embedding-based retrieval with LLM for Effective Agriculture Information Extracting from Unstructured Data. *arXiv:2308.03107*, 2023. DOI: 10.48550/arXiv.2308.03107

6. **Patiny L., Godin G.** Automatic extraction of FAIR data from publications using LLM. *ChemRxiv*, 2023. DOI: 10.26434/chemrxiv-2023-05v1b-v2

7. **Biswas A., Talukdar W.** Robustness of structured data extraction from in-plane rotated documents using multi-modal Large Language Models (LLM). *arXiv:2406.10295*, 2024. DOI: 10.48550/arXiv.2406.10295

8. **Birhane A., Kasirzadeh A., Leslie D., Wachter S.** Science in the age of large language models. *Nature Reviews: Physics*, 2023, Vol. 5, Pp. 277–280. DOI: 10.1038/s42254-023-00581-4

9. **Manjotho A.A., Tewolde T.T., Duma R.A., Niu Z.** LLM-guided fuzzy kinematic modeling for resolving kinematic uncertainties and linguistic ambiguities in text-to-motion generation. *Expert Systems with Applications*, 2025, Vol. 279, Art. no. 127283. DOI: 10.1016/j.eswa.2025.127283

10. **Chen X., Huang X., Gao Q., Huang L., Liu G.** Enhancing text-centric fake news detection via external knowledge distillation from LLMs. *Neural Networks*, 2025, Vol. 187, Ar. no. 107377. DOI: 10.1016/j.neunet.2025.107377

11. **Achaim J., Adler S., Agarwal S. et al.** GPT-4 Technical Report. *arXiv:2303.08774*, 2023. DOI: 10.48550/arXiv.2303.08774

12. **Devlin J., Chang M.-W., Lee K., Toutanova K.** BERT: Pre-training of deep bidirectional transformers for language understanding. *arXiv:1810.04805*, 2018. DOI: 10.48550/arXiv.1810.04805

13. **Ebtesam A., Hamza A., Abdulaziz Al. et al.** The Falcon Series of Open Language Models. *arXiv:2311.16867*, 2023. DOI: 10.48550/arXiv.2311.16867

14. **Jessop D.M., Adams S.E., Willighagen E.L., Hawizy L., Murray-Rust P.** OSCAR4: a flexible architecture for chemical text-mining. *Journal of Cheminformatics*, 2011, Vol. 3., Art no. 41. DOI: 10.1186/1758-2946-3-41

15. **Hawizy L., Jessop D.M., Adams N., Murray-Rust P.** ChemicalTagger: A tool for semantic text-mining in chemistry. *Journal of Cheminformatics*, 2011, Vol. 3, Art. no. 17. DOI: 10.1186/1758-2946-3-17

16. **Mavračić J., Court C.J., Isazawa T., Elliott S.R., Cole J.M.** ChemDataExtractor 2.0: Autopopulated ontologies for materials science. *Journal of Chemical Information and Modeling*, 2021, Vol. 61, No. 9, pp. 4280–4289. DOI: 10.1021/acs.jcim.1c00446

17. **Chen B., Zhang Z., Langrené N., Zhu S.** Unleashing the potential of prompt engineering for large language models. *Patterns*, 2025, Vol. 6, No. 6, Art. no. 101260. DOI: 10.1016/j.patter.2025.101260

18. **Knoth N., Tolzin A., Janson A., Leimeister J.M.** AI literacy and its implications for prompt engineering strategies. *Computers and Education: Artificial Intelligence*, 2024, Vol. 6, Art. no. 100225. DOI: 10.1016/j.caeai.2024.100225

19. **Huang S., Cole J.M.** BatteryBERT: A pretrained language model for battery database enhancement. *Journal of Chemical Information and Modeling*, 2022, Vol. 62, No. 24, Pp. 6365–6377. DOI: 10.1021/acs.jcim.2c00035

20. **Zheng Z., Zhang O., Borgs C., Chayes J.T., Yaghi O.M.** ChatGPT Chemistry assistant for text mining and the prediction of MOF synthesis. *Journal of the American Chemical Society*, 2023, Vol. 145, No. 32, Pp. 18048–18062. DOI: 10.1021/jacs.3c05819

21. **Da Silva V.T., Rademaker A., Lionti K., Giro R., Lima G., Fiorini S., Archanjo M., Carvalho B.W., Neumann R., Souza A., Souza J.P., de Valnisio G., Paz C.N., Cerqueira R., Steiner M.** Automated, LLM enabled extraction of synthesis details for reticular materials from scientific literature. *arXiv.2411.03484*, 2024. DOI: 10.48550/arXiv.2411.03484

22. **Ji Z., Lee N., Frieske R., Yu T., Su D., Xu Y., Ishii E., Bang Y.J., Madotto A., Fung P.** Survey of hallucination in natural language generation. *arXiv.2202.03629*, 2022. DOI: 10.48550/arXiv.2202.03629

INFORMATION ABOUT AUTHORS / СВЕДЕНИЯ ОБ АВТОРАХ

Evgeniy A. Pavlov
Павлов Евгений Алексеевич
E-mail: pavlov_ea@spbstu.ru
ORCID: https://orcid.org/0000-0002-7437-6153

Pavel D. Drobintsev
Дробинцев Павел Дмитриевич
E-mail: drob@ics2.ecd.spbstu.ru

Victor A. Klinkov
Клинков Виктор Артемович
E-mail: klinkovvictor@yandex.ru

Alexander V. Semencha
Семенча Александр Вячеславович
E-mail: asemencha@spbstu.ru

Igor G. Chernorutskiy
Черноруцкий Игорь Георгиевич
E-mail: igcher1946@mail.ru

Submitted: 01.07.2025; Approved: 12.09.2025; Accepted: 19.09.2025.
Поступила: 01.07.2025; Одобрена: 12.09.2025; Принята: 19.09.2025.

Computing, Telecommunication and Control, 2025, Vol. 18, No. 3, Pp. 68–79.
Информатика, телекоммуникации и управление. 2025. Том 18, № 3. С. 68–79.

Circuits and Systems for Receiving, Transmitting and Signal Processing

Устройства и системы передачи, приема и обработки сигналов

Research article
DOI: https://doi.org/10.18721/JCSTCS.18306
UDC 621.3.049.774.2

DESIGN OF FILTERS USING PSEUDO RESISTORS FOR BIOMEDICAL DEVICES

A.A. Pyatlin ✉ , *D.V. Morozov*

Peter the Great St. Petersburg Polytechnic University,
St. Petersburg, Russian Federation

✉ poccomaxa@cave3d.com

Abstract. The current state of the scientific and technical problem being solved is assessed, initial data are obtained, a 4th-order Sallen–Key filter and a filter with imitation of the inductors by an active circuit based on impedance converters using operational amplifiers and pseudo resistors are developed, and the results are compared. It is recommended to use the low-pass filter with imitation of the inductors in an electronic stethoscope, since it has, compared to the Sallen–Key filter, a less roll-off in the passband, greater attenuation in the stopband, a sharper drop in the frequency response in the transition region, better noise characteristics and a larger dynamic range. The filter with imitation of the inductors has lower nonlinear distortions and demonstrates operability with a spread of temperatures and element ratings, especially in the frequency range containing the main peaks of heartbeat and lung sounds, and the power consumption and hardware costs of such a filter are comparable to similar characteristics of the Sallen–Key filter.

Keywords: electronic stethoscope, Sallen–Key filter, imitation of inductance, pseudo resistor, negative impedance converter, noise characteristic, integrated circuit layout

Acknowledgements: The production of the integrated microcircuit was carried out at the expense of the Ministry of Education and Science of Russia within the framework of the federal project "Training of personnel and scientific foundation for the electronic industry" under the state assignment for the implementation of research work "Development of a methodology for prototyping an electronic component base in domestic microelectronic production based on the MPW service".

Citation: Pyatlin A.A., Morozov D.V. Design of filters using pseudo resistors for biomedical devices. Computing, Telecommunications and Control, 2025, Vol. 18, No. 3, Pp. 68–79. DOI: 10.18721/JCSTCS.18306

Научная статья
DOI: https://doi.org/10.18721/JCSTCS.18306
УДК 621.3.049.774.2

РАЗРАБОТКА ФИЛЬТРОВ С ИСПОЛЬЗОВАНИЕМ ПСЕВДОРЕЗИСТОРОВ ДЛЯ БИОМЕДИЦИНСКОГО ОБОРУДОВАНИЯ

А.А. Пятлин ✉ , *Д.В. Морозов*

Санкт-Петербургский политехнический университет Петра Великого,
Санкт-Петербург, Российская Федерация

✉ poccomaxa@cave3d.com

Аннотация. Проведена оценка современного состояния решаемой научно-технической проблемы, получены исходные данные, разработаны фильтр Саллена–Ки 4-го порядка и фильтр с имитацией катушек индуктивности активной схемой на основе конверторов импеданса с использованием операционных усилителей и псевдорезисторов, проведено сравнение результатов. В электронном стетоскопе рекомендуется использовать фильтр нижних частот с имитацией катушек индуктивности, т.к. он имеет по сравнению с фильтром Саллена–Ки меньшую неравномерность в полосе пропускания, большее подавление, более резкий спад амплитудно-частотной характеристики в переходной области, лучшую шумовую характеристику и больший динамический диапазон. Фильтр с имитацией катушек индуктивности имеет меньшие нелинейные искажения и показывает работоспособность при разбросе температур и номиналов элементов, особенно в диапазоне частот, содержащих основные пики звуков биения сердца и легких, а потребляемая мощность и аппаратные затраты такого фильтра сравнимы с аналогичными характеристиками фильтра Саллена–Ки.

Ключевые слова: электронный стетоскоп, фильтр Саллена–Ки, имитация индуктивности, псевдорезистор, отрицательный конвертор импеданса, шумовая характеристика, топология интегральной схемы

Финансирование: Изготовление интегральной микросхемы осуществлялось за счет средств Минобрнауки России в рамках федерального проекта «Подготовка кадров и научное обеспечение электронной промышленности» по государственному заданию на выполнение научно-исследовательской работы «Разработка методики прототипирования электронной компонентной базы в отечественном микроэлектронном производстве на основе сервиса МКР».

Для цитирования: Pyatlin A.A., Morozov D.V. Design of filters using pseudo resistors for biomedical devices // Computing, Telecommunications and Control. 2025. Т. 18, № 3. С. 68–79. DOI: 10.18721/JCSTCS.18306

Introduction

It is often necessary to solve the problem of eliminating unwanted noises and to allocate the desired frequency range during signal processing. For this purpose, filters are used. Low-pass filters (LPFs) are used in the portable electronic stethoscope, because the main frequency peaks of the input signals are in the range up to 1 kHz. At the same time, it is necessary to take into account that the electronic stethoscope should be compact, and therefore minimize its sizes.

The chip area reduction can be realized by applying several methods. Firstly, it is necessary to get rid of inductors, because they have rather large dimensions, dissipate a lot of energy, and cause deviations from calculated characteristics. Inductors can be eliminated by replacing them with negative impedance converters (NICs) based on operational amplifiers (OpAmps), which significantly reduces the chip area. That is, ARC filter implementation should be applied to solve this problem.

Secondly, it is necessary to reduce the values of capacitances in the chip, because after replacing the inductors, the size of the chip depends on the size of the capacitors it contains, the area of which is directly proportional to their nominal values.

Thirdly, resistors should be replaced by pseudo resistors. The nominal values and, accordingly, the sizes of resistors increase significantly after reducing the nominal values of the capacitors in the filter. The problem of minimizing the resistor area while maintaining its characteristics can be solved by using the pseudo resistor, which is a special MOS transistor(s) connection circuit. Such a pseudo resistor has linear resistance and small area at the same time.

In [1], the stethoscope has been created that has the input range of heartbeat frequencies from 10 to 500 Hz. The amplitude of the input signal is 35–50 mV, and power supply is 3.7 V. A 2nd-order Sallen–Key filter is used.

In [2], a portable electronic stethoscope with a wireless data transmission module is described. The device is equipped with 2nd-order Sallen–Key filter with the cut-off frequency of 1 kHz.

In [3], an electronic stethoscope is developed, which uses a 4th-order Butterworth bandpass filter. The main heart sound oscillations are in the frequency range of 44 to 54 Hz. The filter has the passband gain of 40 dB and the cut-off frequency of 150 Hz.

The research reported in [4] examines the performance of the 3M Littmann 3200 electronic stethoscope, which amplifies the signal by 24 times and operates in the frequency range from 20 to 2000 Hz, covering the most part of the acoustic energy of respiratory and cardiac tones.

The study [5] deals with the development of an electronic stethoscope using MEMS technology. In this case, it is found that the frequencies of heartbeat tones mainly lie in the range from 20 to 600 Hz.

Authors of [6] analyzed the design of a wireless electronic stethoscope that allows the user to remotely record and transmit listening results. The research shows that the fundamental heart rate ranges between 30 and 70 Hz, while the heart tones and faint noises are distributed between 100 and 1000 Hz. The device uses 2nd-order Butterworth filter with the cut-off frequency of 1 kHz and attenuation of 2 dB at it.

In [7], the use of digital stethoscope for diagnosis of heart disease is discussed. It was found that the basic frequency range of a healthy heartbeat is between 20 and 400 Hz. However, frequencies up to 1 kHz should be considered for a complete analysis of the heart, as they include sounds indicative of heart pathologies.

The basic circuits and applications of pseudo resistors have also been considered, for example, the paper [8] analyzed the design of pseudo resistors to be used in LPFs. The study states that a transistor biased in the weak inversion region can act as a linear resistor in a circuit where the resistance is controlled by the gate voltage of the MOS transistor and can provide very high resistance values including hundreds of mega ohms.

The work [9] describes adjustable pseudo resistors in the design of analog LPFs for implantable biomedical devices. The adjustable pseudo resistors, according to this work, are MOS transistors put into the inversion mode in which large resistance values can be achieved. The resistance is regulated by the transistor gate voltage. The transistors are connected in series to increase the resistance.

In [10], the effect of nonlinearity of pseudo resistors on the performance of biomedical equipment is studied. The work uses a double pseudo resistor switching scheme without a control voltage on the transistor gate.

As can be seen, 4th-order Sallen–Key filter is used most often in the electronic stethoscope, which is a realization of Butterworth approximation. Butterworth approximation has a flat frequency response in the passband. However, we would like to achieve a sharper transition between the passband and the stopband in order to more effectively attenuate signals at frequencies of the stopband. Zolotarev approximation helps to accomplish this task. Although Zolotarev approximation has ripples in the passband in comparison with Butterworth approximation, it is possible to significantly reduce attenuation at the cut-off frequency in comparison with Butterworth approximation. We choose for implementation of Zolotarev approximation an active LPF based on NICs with the cut-off frequency of 1 kHz.

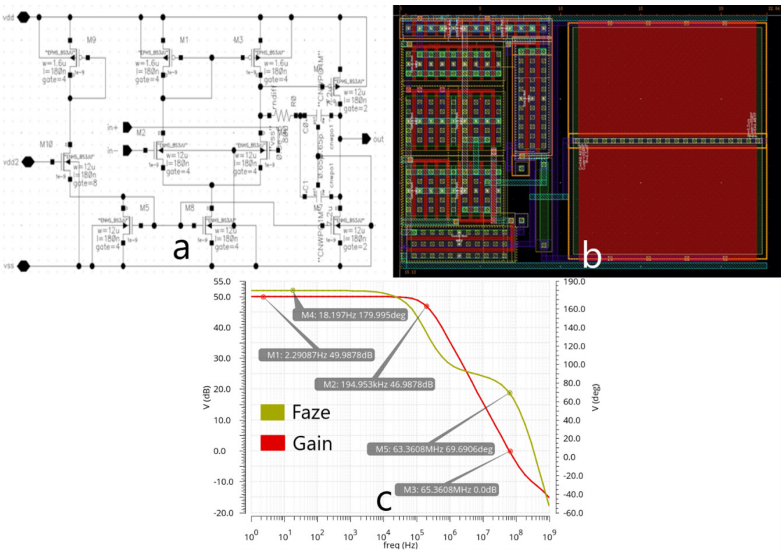


Fig. 1. Schematic (*a*), layout (*b*), the amplitude-frequency and the phase-frequency responses (*c*) of the two-stage OpAmp

Thus, the aim of the work is to design the LPF for the electronic stethoscope. The JSC "Mikron" 180 nm technology is used.

Design of the OpAmp

Firstly, a two-stage OpAmp on MOSFETs was designed, the circuit and layout are shown in Fig. 1a and b. Then a simulation in the frequency domain was conducted, the amplitude-frequency and the phase-frequency responses are shown in Fig. 1c. The OpAmp occupies an area of 345 μm^2 as shown in Fig. 1b. In addition, the sources and drains of some transistors are united to reduce the chip-on area.

The gain of the OpAmp is 50 dB, the 3dB frequency is 195 kHz and the phase margin is 70°, according to Fig. 1, c. The nonlinear distortion of the OpAmp is 93 dB, and the power consumption is 135 μW. This characteristic of the OpAmp is obtained at a load of 10 MOhms, which corresponds to the typical loads of the analyzed filters. The characteristics at loads of 5 MOhms, 200 MOhms, 10 pF, 25 pF were also obtained, which correlates with the loads of the filters. These characteristics are sufficiently close to the shown above. This OpAmp is quite simple to implement, but it provides the necessary characteristics and additionally attenuates the signal at high frequencies.

Research of pseudo resistors

Since one of objectives of this work is to minimize the area occupied by the filter, in the proposed filter, first, the large area inductors were replaced by NICs, and then the capacitor values and therefore their area were reduced.

As a result, the resistor nominal values in the filter became large, as well as the area occupied by the resistors. The resistance values of such resistors are to several hundreds of mega ohms. As a result, it is necessary to solve the problem of realization of large resistance values with much smaller on-chip area than ordinary resistors. This problem can be solved by using pseudo resistors in the filter.

A pseudo resistor is a MOS transistor biased to the weak inversion region, which has a near linear and large resistance up to hundreds of mega ohms. However, in this work pseudo resistors based on p- and n-transistors are in saturation. We similarly call these circuits a pseudo resistor because it also provides a linear and large resistance, smaller than transistors in the weak inversion region, but sufficient for the purposes of this work. There are several circuits for connecting MOS transistors into pseudo resistors. For instance, tunable pseudo resistors whose resistance is controlled by the gate voltage, but then each pseudo

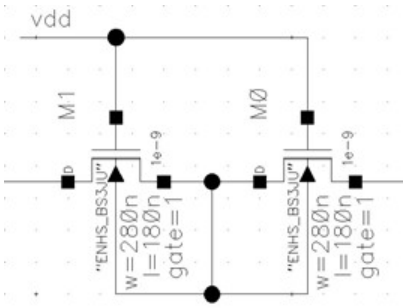


Fig. 2. Pseudo resistor circuit with two transistors

resistor would require a separate voltage source. In this case, it is necessary to create several highly stable voltage sources, which, firstly, is hard, and secondly, requires additional hardware costs and on-chip area.

Therefore, it is much more suitable to include a pseudo resistor with the gate connected to one of the voltage source already used in the filter circuit, thus avoiding the need to create an additional power supply that also consumes additional power. Such a pseudo resistor will have a constant resistance, which can be adjusted by the length and width of the MOS transistor gate. In addition, the resistance value can be increased by connecting several such pseudo resistors in series. Another way to increase the resistance of a pseudo resistor is to use two MOS transistors connected in a certain way, the general circuit of such a pseudo resistor used in the filter is shown in Fig. 2. The transistor substrate is connected to the transistor drain or to the node between the transistors drain and source. Thus, such a pseudo resistor is in a separate area relative to the rest of the chip and has its own potential. Table 1 shows the pseudo resistor resistance depending on the transistor channel length L, the width W and connection type.

Table 1

Resistance of pseudo resistors

W, μm	L, μm	Resistance of pseudo resistor, kOhm			
		One n-MOS transistor	One p-MOS transistor	Two n-MOS transistors	Two p-MOS transistors
0.28	0.18	21	13	37	23
0.28	1	113	68	195	118
0.28	10	975	747	1700	1270
0.28	100	9600	7460	17000	13200
1	0.18	6.8	4	12	7
10	0.18	0.66	0.36	1.15	0.64
100	0.18	0.065	0.038	0.11	0.064

As we can see from Table 1, the resistance of the pseudo resistor increases with increasing length and decreasing width of the MOS transistor. The two-transistors circuit gives almost twice the resistance of the single transistor. Circuits with n-MOS transistors provide more resistance than those with p-MOS transistors.

Considering the fact that a large resistance has to be realized and creating a separate well for the transistor requires additional area on the chip, two-transistor circuits are chosen for use in the filter. The transistor width is minimum and equals 280 nm and the length is 10 μm, because the properties of the transistor deteriorate if too large a value of the length is taken. To obtain the required resistance

value, such pseudo resistors are placed in series. For instance, in Mikron HCMOS8D technology, the "rndiff" resistor of 1 MOhm has length 5820 um and width 1.1 um, and has an area of 6400 um^2, while a pseudo resistor with the same value has an area of 130 μm^2 and 25 μm^2 for n- and p-transistors, respectively. The total value of all resistors in the filter with imitation of the inductors is 250 MOhms and it occupies an area of 1.6×10^6 μm^2, while for pseudo resistors the area would be 2.3×10^4 μm^2, which reduces the area by a factor of about 70. Similarly, the total value of all resistors in the Sallen–Key filter is 47 MOhms and it occupies an area of 3×10^5 μm^2, while for pseudo resistors the area would be 6.1×10^2 μm^2, which reduces the area by a factor of about 50.

Design of 4th-order Sallen–Key filter

The circuit of 4th-order Sallen–Key filter was designed based on [11]. Design of the filter consists in successive complication of the initial circuit and simulation of intermediate filters. Thus, first, a prototype of the filter is assembled, then it is recalculated for the desired frequency, one stage of the Sallen–Key filter is calculated, then two stages using ideal elements. Next, one by one, ideal OpAmps and capacitors are replaced by real ones, and resistors are replaced by pseudo resistors. Calculation of resistor and capacitor values, as well as sequential simulation of the filter circuits are given in [11] as well. Thus, we obtain a filter with real elements, small capacitances in value and area, as well as with pseudo resistors with small sizes and large resistance values. We used the voltage follower based on the OpAmp and the capacitor with capacitance of 10 pF as a load of the filter, because after the filter usually use an analog-to-digital converter (ADC) with this value of input impedance (capacitance) in the electronic stethoscope, and the voltage follower is an element to connect the filter output with the input of the ADC. The supply voltage is 1.8 V, and the input sinusoidal signal with DC value equal to half the supply voltage of 0.9 V is used. The filter cut-off frequency should be 1 kHz, and it is calculated by (1) for one filter stage according to [12]:

$$f = \frac{1}{2\pi\sqrt{R_1 R_2 C_1 C_2}}, \tag{1}$$

where f is the filter cut-off frequency; R_1, R_2 are resistance values; C_1, C_2 are capacitance values.

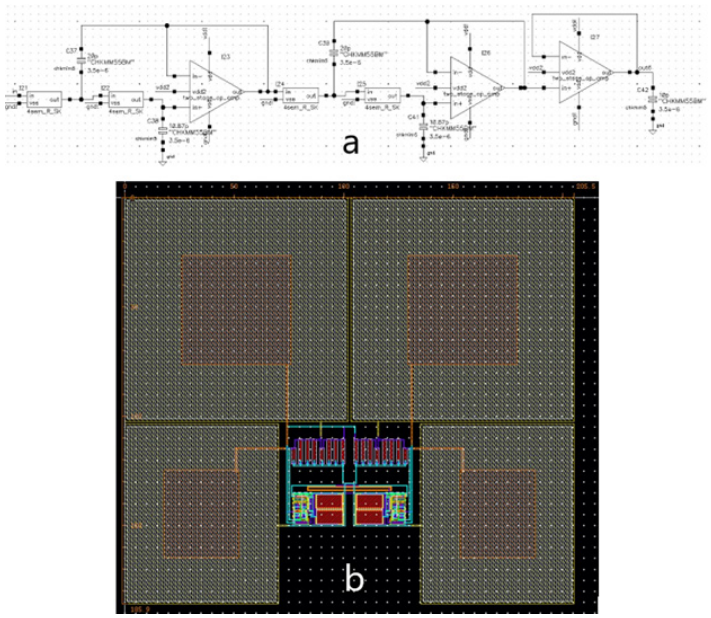


Fig. 3. Circuit (*a*) and layout (*b*) of the 4th-order Sallen–Key filter

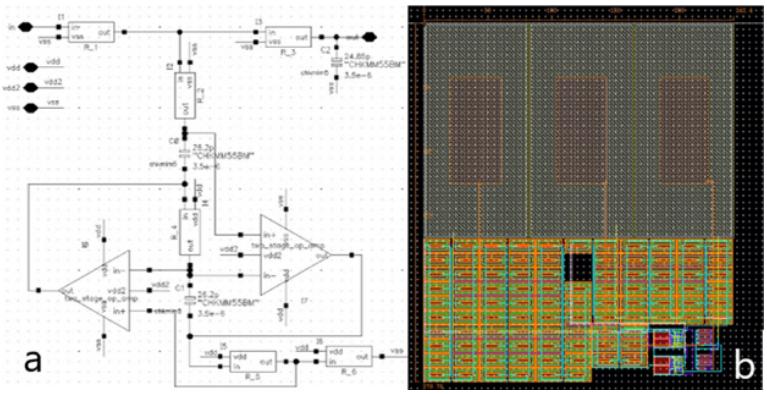


Fig. 4. Circuit (*a*) and layout (*b*) of the filter with imitation of the inductors

The final step of design is layout working out and simulation. Fig. 3 shows the circuit and layout of the 4th-order Sallen–Key filter. The area occupied by the filter is 38202 μm^2. As can be seen from Fig. 3, capacitors take up most of the chip area, however it was possible to significantly reduce the chip area by using pseudo resistors.

Design of the filter with imitation of the inductors

To implement Zolotarev approximation, the filter circuit with imitation of the inductors by an active circuit based on NICs using pseudo resistors is chosen, which allows to achieve less ripples in dB in the passband compared to the monotonic decreasing of the Sallen–Key filter characteristic and greater attenuation. Moreover, it allows to significantly reduce the area occupied by the final filter circuit in comparison with the area of the initial circuit on the chip.

First, a normalized prototype of the 3rd order elliptic filter C0350 is taken from [13]. This filter is then recalculated to the required frequency of 1 kHz, according to the formulas given in [11]. It is possible to replace the inductors with NICs according to [14]. In order to realize the filter with the smallest number of NICs, Bruton transformation of the LCR-prototype should be used, after which a single NIC is required to realize the grounded inductance according to [15]. Therefore, Bruton transformation is applied to the recalculated filter according to the formulas given in [11]. Bruton transformation performs an equivalent transformation of the LCR-circuit of the prototype filter, after which all resistors are replaced by capacitors, inductors – by resistors, and capacitors – by D-elements, supercapacitances. Each supercapacitance, in turn, can be replaced by a single NIC.

Next, similarly to the previous filter, a circuit with ideal elements is created, which, in turn, are replaced by real ones, and resistors are replaced by pseudo resistors. The stages of circuit design, simulation and calculations are given in detail in [16]. It is worth noting that when replacing resistors with pseudo resistors, the choice of the latter with p- or n-MOS transistors depends on the nonlinear distortion they introduce at a given location on the circuit. The circuit and layout of the filter with imitation of the inductors are shown in Fig. 4. The area occupied by the filter is 67819 μm^2. As can be seen from Fig. 4, with the capacitors occupying more than half of the chip area and the pseudo resistors just under half of the chip area. It was possible to significantly reduce the chip area by using pseudo resistors.

Simulation and comparison of the characteristics of the 4th-order Sallen–Key filter and the filter with imitation of the inductors

Firstly, let us analyze the frequency response of the filters, the simulation of which is shown in Fig. 5.

As can be seen from Fig. 5, the cut-off frequency of both filters is 1 kHz, while ripples in the passband of the filter with imitation of the inductors are less and equal to 0.89 dB, while in the Sallen–Key filter the attenuation is 3 dB. In addition, the filter with imitation of the inductors has more rapid

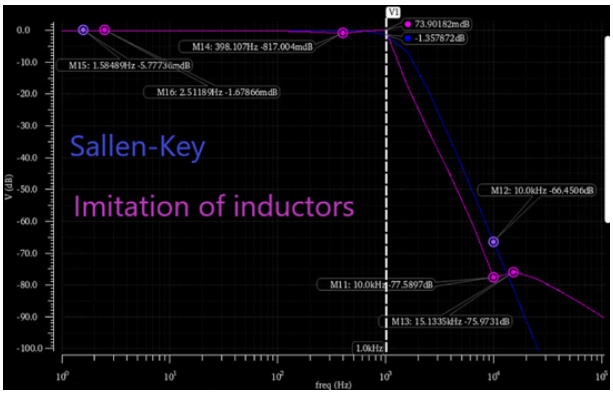


Fig. 5. Frequency responses of Sallen–Key filter and the filter with imitation of the inductors

decline of the frequency response in the transition band, and the stopband starts from 10 kHz with greater attenuation more than 76 dB, while in the Sallen–Key filter attenuation in the stopband is no more than 69 dB. According to these parameters, the filter with imitation of the inductors has some advantages over the Sallen–Key filter.

It is also worth noting that the frequency responses of the filters display the real attenuation of the input signal with the working amplitude. To verify this statement, the simulation in the time domain was performed with the maximum possible amplitude of the input signal of 50 mV and characteristic frequencies of heartbeat and lung sounds, the obtained data were recalculated in dB. As a result, the attenuation values in the time and frequency domains are close.

Next, the filters have been simulated at different temperatures and technology corner parameters. For such simulation, we have considered all possible cases, and then we have selected two worst cases of frequency response up and down relative to the frequency response for typical elements and room temperature. The temperature values are –40, 27 and 85°C. Since these temperatures allow us to consider the operation of the device at negative temperatures, room temperature as well as at elevated temperatures, this temperature range is much larger than the range of temperatures at which the filter is used. Table 2 shows simulation results of the frequency response of the filters.

Table 2

Comparison of frequency response deviations of the filters

Frequency response type	Cut-off frequency, Hz	Cut-off frequency deviation, Hz	Pass-band ripples, dB	Deviation of the passband ripples, dB	Stopband frequency, Hz	Deviation of the stopband frequency, Hz
Sallen–Key filter						
–40 ss	1636	636	3	0	16873	6873
+27 tt	1000	0	3	0	10000	0
+85 ff	638	362	3	0	6392	4392
Filter with imitation of the inductors						
–40 ss	1720	720	3	2	22165	12165
+27 tt	1000	0	1	0	10000	0
+85 ff	500	500	2	1	5100	4900

As can be seen from Table 2, the Sallen–Key filter shows better stability of characteristics when changing temperatures and corner parameters than the filter with imitation of the inductors. However, it should be noted that the ripples of the frequency response of the filter with imitation of the inductors are concentrated at the end of the passband, that is, it does not affect the signals at the characteristic frequencies of heartbeat and lung beats. Moreover, the filter with imitation of the inductors provides more attenuation in the stopband, and the required attenuation is achieved at lower frequencies than for the Sallen–Key filter.

Next, we evaluate the nonlinear distortion of the filters at characteristic heart and lung beat frequencies with the maximum amplitude of the input signal equal to 50 mV. Nonlinear distortions are estimated from the spectrum as the difference of values between the main tone and the next largest harmonics, the results for two filters are given in Table 3.

Table 3

Nonlinear distortions of the filters

Frequency, Hz	50	200	500	800	1000
Nonlinear distortion of the Sallen–Key filter, dB	67	56	53	51	50.9
Nonlinear distortion of filter with imitation of the inductors, dB	70	56	54	50.9	50.2

The nonlinear distortion is slightly less for the Sallen–Key filter at the end of the passband. In addition, it is less for the filter with imitation of the inductors at the start of the passband, where the main peaks of heartbeat sounds are concentrated. At the same time, both filters show acceptable nonlinear distortions, because after the filter it is supposed ADC with N = 8 digit capacity, the dynamic range of which is estimated at 49 dB, and therefore nonlinear distortions of the filter should not exceed 49 dB. A larger ADC bit rate is not required, as it is necessary to catch only the main peaks of sounds, but not to get the sound of high quality.

Now let us estimate the dynamic range of the filters. First, we will find the upper limit of the dynamic range. To do this, we will gradually increase the amplitude of the input signal until the nonlinear distortion reaches 49 dB. Thus, we have found the maximum allowable amplitude of the input signal for the Sallen–Key filter equal to 86 mV and for the filter with imitation of the inductors equal to 90 mV. The filter with imitation of the inductors has a slightly larger upper limit of dynamic range.

Then let us consider the noise characteristics of the filters, which are shown in Fig. 6.

The noise response values are smaller for the filter with imitation of the inductors than for the Sallen–Key filter in the whole frequency range except for the end of the passband. The lower limit of the dynamic range of the filters was also determined. The minimum possible amplitude of the input signal for the Sallen–Key filter is 0.6 mV, and for the filter with imitation of the inductors is 0.56 mV.

The dynamic range of the Sallen–Key filter is $20\log_{10}\left(\frac{86}{0.6}\right) = 43$ dB, and for the filter with imitation of the inductors – $20\log_{10}\left(\frac{90}{0.56}\right) = 44$ dB. Thus, the dynamic range of the filter with imitation of the inductors is 1 dB larger than the dynamic range of the Sallen–Key filter.

The power consumption of the filters is simulated, which is mainly determined by the power consumption of OpAmps. The power consumption of the filters is similar, since the number of OpAmps in the filters is the same, the power consumption of the Sallen–Key filter is 287 μW, and that of the filter with imitation of the inductors is 293 μW.

The areas occupied by the filters on the crystal are also found. For the Sallen–Key filter, the area is 38202 μm^2, and for the filter with imitation of the inductors it is almost twice as large and equals 67819 μm^2. Although for the filter with imitation of the inductors the area is larger due to the large

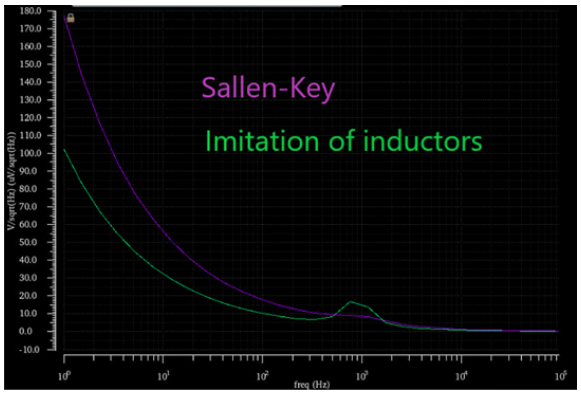


Fig. 6. Noise characteristics of the filters

size of the pseudo resistors, the use of pseudo resistors allowed to significantly reduce the chip area. It should be taken into account that in the electronic stethoscope the ADC is used, the area of which is comparable with filter areas, and the difference in filter areas is not critical in order to be able to make a choice in favor of using the filter with imitation of the inductors in the electronic stethoscope for the rest of the considered characteristics of the filters.

Table 4 summarizes the characteristics of the filter with imitation of the inductors and compares them with similar characteristic of the filters from published works.

Table 4

Characteristics of the filter with imitation of the inductors and the filters from published works

Work	This work	[1]	[17]	[18]	[19]	[20]	[21]	[22]	[22]
Technology, μm	0.18	–	0.35	0.18	0.18	0.18	0.8	0.18	0.18
Filter	II[1]	Bw[2]	–	Bq[3]	OTA	Bw[2]	–	SK[4] (rc)	SK[4] (sc)
Order	3	4	4	4	9	5	2	2	2
Power supply, V	1.8	9	0.6	1	1.8	1	1.25	–	–
Cut-off frequency, Hz	1000	150	200	500–1000	5400	250	750	610	600
Ripples, dB	0.89	3	1.8	–	–	3.5	–	3	3
Stopband frequency, kHz	10	1	0.6	–	–	1	–	–	–
Attenuation at the stopband frequency, dB	77	40	70	–	–	65	–	–	–
Non-linear distortion, dB	50.2	–	60	40	38	49	49	26	26
Noise, $\mu V/\sqrt{Hz}$	8–17	–	–	–	–	20	–	32	21
Dynamic range, dB	44	–	47	55	34	50	62	–	–
Power consumption, μW	293	–	0.9 n[5]	14 n[5]	360 n[5]	450 n[5]	2.5 n[5]	413	296
Area, mm^2	0.067	–	0.17	0.13	0.03	0.13	0.23	–	–

[1] Filter with imitation of inductors.
[2] Butterworth filter.
[3] Biquadratic filter.
[4] Sallen–Key filter.
[5] Very low power consumption technology has been applied in the development of the filter; the values are given in nW.

Conclusion

The filter with imitation of the inductors by an active circuit based at the NICs with the use of pseudo resistors is proposed for the electronic stethoscope. The problem of area minimization of filters with preservation of their characteristics was solved by successive exclusion of inductors from the filter circuit, reduction of nominal values and areas of capacitances, replacement of resistors by pseudo resistors. The use of pseudo resistors instead of resistors allows to reduce the area occupied by the chip, while keeping the necessary values of resistances.

Mikron 180 nm technology is used. Simulation results for the Sallen–Key filter of the 4th order, as well as for the filter with imitation of the inductors by an active circuit based on the NICs using pseudo resistor were presented based on the filter topologies. Comparison of the filters is carried out and it is revealed that the filter with imitation of the inductors has some advantages over the 4th-order Sallen–Key filter in terms of the obtained characteristics.

Thus, for electronic stethoscope it is recommended to use the LPF with imitation of the inductors, as it provides better or at least the same characteristics as the filters used for the electronic stethoscope in other works.

REFERENCES

1. **Waqar M., Inam S., ur Rehman M.A., Ishaq M., Afzal M., Tariq N., Amin F., Qurat-ul-Ain.** Arduino based cost-effective design and development of a digital stethoscope. *2019 15th International Conference on Emerging Technologies* (*ICET*), 2019, pp. 1–6. DOI: 10.1109/ICET48972.2019.8994674

2. **Frank P.-W.L., Meng M.Q.-H.** A low cost Bluetooth powered wearable digital stethoscope for cardiac murmur. *2016 IEEE International Conference on Information and Automation* (*ICIA*), 2016, pp. 1179–1182. DOI: 10.1109/ICInfA.2016.7831998

3. **Malik B., Eya N., Migdadi H., Ngala M.J., Abd-Alhameed R.A., Noras J.M.** Design and development of an electronic stethoscope. *2017 Internet Technologies and Applications* (*ITA*), 2017, pp. 324–328. DOI: 10.1109/ITECHA.2017.8101963

4. **Oliynik V.** On potential effectiveness of integration of 3M Littmann 3200 electronic stethoscopes into the third-party diagnostic systems with auscultation signal processing. *2015 IEEE 35th International Conference on Electronics and Nanotechnology* (*ELNANO*), 2015, pp. 417–421. DOI: 10.1109/ELNANO.2015.7146923

5. **Shi P., Li Y., Zhang W., Zhang G., Cui J., Wang S., Wang B.** Design and implementation of bionic MEMS electronic heart sound stethoscope. *IEEE Sensors Journal*, 2022, Vol. 22, no. 2, pp. 1163–1172. DOI: 10.1109/JSEN.2021.3131001

6. **Joshi S.S., Patil M.R., Kanawade N.P., More A.P.** Bluetooth-based wireless digital stethoscope. *2021 International Conference on Emerging Smart Computing and Informatics* (*ESCI*), 2021, pp. 197–202. DOI: 10.1109/ESCI50559.2021.9396835

7. **Suseno J.E., Burhanudin M.** The signal processing of heart sound from digital stethoscope for identification of heart condition using wavelet transform and neural network. *2017 1st International Conference on Informatics and Computational Sciences* (*ICICoS*), 2017, pp. 153–158. DOI: 10.1109/ICICOS.2017.8276354

8. **Neshatvar N., Nashash H.A., Basha L.A.** Design of low frequency highpass filter using pseudo resistors. *2011 1st Middle East Conference on Biomedical Engineering*, 2011, pp. 407–410. DOI: 10.1109/MECBME.2011.5752152

9. **Neshatvar N., Nashash H.A., Basha L.A.** Design of low frequency analog low pass filter using tunable pseudo resistors. *2nd Middle East Conference on Biomedical Engineering*, 2014, pp. 39–42. DOI: 10.1109/MECBME.2014.6783202

10. **AbuShawish I.Y.I., Mahmoud S.A., Majzoub S., Hussain A.J.** Biomedical amplifiers design based on pseudo-resistors: A review. *IEEE Sensors Journal*, 2023, Vol. 23, no. 14, pp. 15225–15238. DOI: 10.1109/JSEN.2023.3280668

11. **Pyatlin A.A.** Chastotno-izbiratel'noe ustroistvo dlia elektretnogo kondensatornogo mikrofona. Vyp. kvalif. rab. [Frequency-selective device for an electret condenser microphone. Final qualifying work], St. Petersburg, 2022, 62 p. DOI: 10.18720/SPBPU/3/2022/vr/vr22-2641

12. **Piatlin A.A., Morozov D.V.** Chastotno-izbiratel'noe ustroistvo dlia elektretnogo kondensatornogo mikrofona [Frequency-selective device for an electret condenser microphone. Final qualifying work]. *Nedelia Nauki IEiT. Materialy Vserossiiskoi konferentsii* [*Week of Science of the Institute of Economics and Technology. Proceedings of the All-Russian Conference*], 2022, pp. 112–115.

13. **Zverev A.I.** Handbook of filter synthesis. NY, London, Sydney: John Wiley and Sons, Inc. 1967. 576 p.

14. **Piatlin A.A., Morozov D.V.** Chastotno-izbiratel'noe ustroistvo dlia elektretnogo kondensatornogo mikrofona [Frequency-selective device for an electret condenser microphone. Final qualifying work]. *Nedelia Nauki IEiT. Materialy Vserossiiskoi konferentsii* [*Week of Science of the Institute of Economics and Technology. Proceedings of the All-Russian Conference*], 2023, pp. 202–205.

15. **Korotkov A.S.** Switched-capacitor filter designs: tutorial. *St. Petersburg: Polytechnic University Publishing House*, 2014, 191 p. DOI: 10.18720/SPBPU/2/si21-236

16. **Piatlin A.A., Morozov D.V.** Low-pass filter for an electret condenser microphone. *2023 International Conference on Electrical Engineering and Photonics* (*EExPolytech*), 2023, pp. 53–56. DOI: 10.1109/EExPolytech58658.2023.10318556

17. **Sawigun C., Thanapitak S.** A 0.9-nW, 101-Hz, and 46.3-μVrms IRN low-pass filter for ECG acquisition using FVF biquads. *IEEE Transactions on Very Large Scale Integration* (*VLSI*) *Systems*, 2018, Vol. 26, no. 11, pp. 2290–2298. DOI: 10.1109/TVLSI.2018.2863706

18. **Yang M., Liu J., Xiao Y., Liao H.** 14.4 nW fourth-order bandpass filter for biomedical applications. *Electronics Letters*, 2010, Vol. 46, no. 14, pp. 973–974. DOI: 10.1049/el.2010.1520

19. **Gosselin B., Sawan M., Kerherve E.** Linear-phase delay filters for ultra-low-power signal processing in neural recording implants. IEEE Transactions on Biomedical Circuits and Systems, 2010, Vol. 4, no. 3, pp. 171–180. DOI: 10.1109/TBCAS.2010.2045756

20. **Lee S.-Y., Cheng C.-J.** Systematic design and modeling of a OTA-C filter for portable ECG detection. *IEEE Transactions on Biomedical Circuits and Systems*, 2009, Vol. 3, no. 1, pp. 53–64. DOI: 10.1109/TBCAS.2008.2007423

21. **Rodriguez-Villegas E., Yufera A., Rueda A.** A 1.25-V micropower Gm-C filter based on FGMOS transistors operating in weak inversion. *IEEE Journal of Solid-State Circuits*, 2004, Vol. 39, no. 1, pp. 100–111. DOI: 10.1109/JSSC.2003.820848

22. **Chapagai K., Bahubalindruni P., Nishtha.** 2nd Order Sallen Key switched capacitor LPF with N-type transistors. *2018 31st International Conference on VLSI Design and 2018 17th International Conference on Embedded Systems* (*VLSID*), 2018, pp. 319–324. DOI: 10.1109/VLSID.2018.83

INFORMATION ABOUT AUTHORS / СВЕДЕНИЯ ОБ АВТОРАХ

Artem A. Pyatlin
Пятлин Артем Андреевич
E-mail: poccomaxa@cave3d.com

Dmitry V. Morozov
Морозов Дмитрий Валерьевич
E-mail: morozov_dv@spbstu.ru
ORCID: https://orcid.org/0000-0003-3403-0120

Submitted: 04.10.2024; Approved: 11.09.2025; Accepted: 15.09.2025.
Поступила: 04.10.2024; Одобрена: 11.09.2025; Принята: 15.09.2025.

Computing, Telecommunication and Control, 2025, Vol. 18, No. 3, Pp. 80–88.
Информатика, телекоммуникации и управление. 2025. Том 18, № 3. С. 80–88.

Research article
DOI: https://doi.org/10.18721/JCSTCS.18307
UDC 621.396.96

ALGORITHM FOR AUTOMATIC RECOGNIZING OF RADAR SCAN TYPE, BASED ON THE EXTRACTION OF STATICAL FEATURES FROM INPUT ANALYZED PROCESS

V.A. Ivannikova ✉ , *V.F. Korotkov*

"Special Technological Center" JSC, St. Petersburg, Russian Federation

✉ ivannikova.vicky@yandex.ru

Abstract. An algorithm for recognizing the radar antenna scan type by radio monitoring station is proposed. It is based on the extraction of a set of statistical features from the input radio signal, characterizing the signal amplitude envelope, modulated according to the scan type of radar antenna pattern. For feature extraction, mechanical, one-dimensional electronic and two-dimensional electronic scanning are considered, along with the beam steering schemes of radar antenna pattern used in practice. The essence of the proposed algorithm is the sequential comparison of the features with certain thresholds. The proposed recognition scheme is relevant for tasks of classifying the radio electronic equipment signals in a complex radio electronic environment and resolving ambiguity in recognizing radio signal sources with overlapping regions in the feature space, which characterizes time-frequency parameters of the radio signal. Experimental results are given, showing that the proposed algorithm ensures high recognition quality and is promising for improving existing and developing new radio monitoring equipment.

Keywords: radio emission source recognition, radar systems, antenna pattern, radar scan type, electronic scanning, mechanical scanning

Citation: Ivannikova V.A., Korotkov V.F. Algorithm for automatic recognizing of radar scan type, based on the extraction of statical features from input analyzed process. Computing, Telecommunications and Control, 2025, Vol. 18, No. 3, Pp. 80–88. DOI: 10.18721/JCSTCS.18307

Научная статья
DOI: https://doi.org/10.18721/JCSTCS.18307
УДК 621.396.96

АЛГОРИТМ ДЛЯ АВТОМАТИЧЕСКОГО РАСПОЗНАВАНИЯ ВИДА РАДИОЛОКАЦИОННОГО ОБЗОРА, ОСНОВАННЫЙ НА ВЫДЕЛЕНИИ СТАТИСТИЧЕСКИХ ПРИЗНАКОВ ИЗ ВХОДНОГО АНАЛИЗИРУЕМОГО ПРОЦЕССА

В.А. Иванникова ✉ , *В.Ф. Коротков*

ООО «Специальный Технологический Центр»,
Санкт-Петербург, Российская Федерация

✉ ivannikova.vicky@yandex.ru

Аннотация. Предложен алгоритм распознавания вида радиолокационного обзора станцией радиотехнического контроля, основанный на выделении из входного радиосигнала набора статистических признаков, характеризующих огибающую амплитуду сигнала, модулированную в соответствии с видом сканирования диаграммы направленности антенны радиолокационной станции. Для выделения признаков распознавания рассмотрены механическое, одномерное электронное и двумерное электронное сканирование, а также применяемые на практике схемы перемещения луча диаграммы направленности антенны радиолокационных средств. Суть предложенного алгоритма заключается в последовательном сравнении признаков с определенными порогами. Предложенная схема распознавания актуальна для задач классификации сигналов радиоэлектронных средств в сложной радиоэлектронной обстановке и разрешения неоднозначности распознавания источников радиосигналов с пересекающимися областями в пространстве признаков, характеризующих частотно-временные параметры радиосигнала. Приводятся результаты экспериментов, показывающие, что предлагаемый алгоритм обеспечивает высокое качество распознавания и перспективен для применения в разработке новых и усовершенствовании существующих средств радиотехнического контроля.

Ключевые слова: распознавание источников радиоизлучения, радиолокационные системы, диаграмма направленности антенны, вид обзора радиолокационных систем, электронное сканирование, механическое сканирование

Для цитирования: Ivannikova V.A., Korotkov V.F. Algorithm for automatic recognizing of radar scan type, based on the extraction of statical features from input analyzed process // Computing, Telecommunications and Control. 2025. Т. 18, № 3. С. 80–88. DOI: 10.18721/JCSTCS.18307

Introduction

Modern radio monitoring stations are capable of providing prompt and reliable information on the location and parameters of radar system (RS) signals. That is why the improvement of their tactical and technical characteristics is an urgent task.

Among a wide range of tasks in the development of monitoring stations means, the key task is to recognize the type and mode of radar operation. This task is solved by measuring the time-frequency parameters of the radar signal. As a rule, such parameters as carrier frequency, pulse duration, pulse repetition interval, type of pulse modulation are used.

However, many radars performing the same functions, such as space surveillance or weapon control, may have overlapping fields in the feature space. In this case, it is important to have additional information about the radar to resolve the recognition ambiguity. It is known that the received radar series pulses also contain information about the scan type (ST) of radar space scanning [1]. This information instructs on each radar's individual features, so it can be used to make a decision about the radar operating type and

mode, linking them to specific positions, monitoring the moving sources. In practice, the visual-manual method is used to solve the problem of recognizing the type of radar. Under these conditions, analyzing the possibility of automating this procedure is of particular importance.

This issue is poorly covered in literature, only some approaches to the formalized description of some types of ST are described in a number of foreign sources [2–10].

The purpose of this paper is to develop an algorithm for recognizing radar ST with a recognition probability of 80% or more based on the systematization of available information. The output data of the proposed algorithm will be used in a multi-level recognition system, where the first level is recognition based on the frequency and time parameters of the radar signal, the second level is classification of the radar by ST, and finally, the third level is decision making based on a set of features.

Brief characterization of radar scan methods

A radar scan is the movement of the antenna pattern beam to view a given (monitored) area of space.

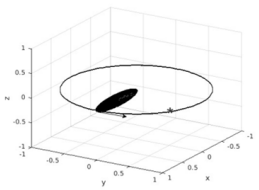
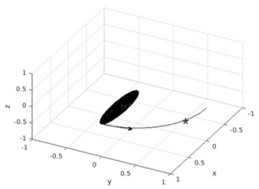
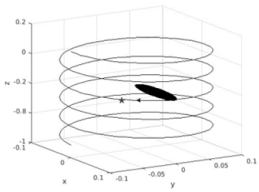
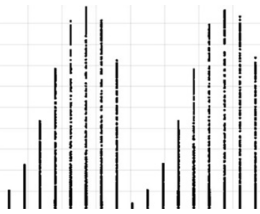
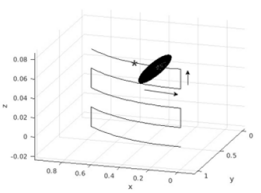
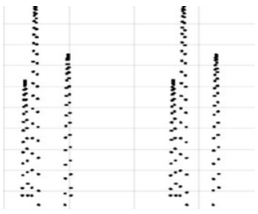
At the receiving and analyzing equipment of the radio monitoring station, the view of ST can be represented in the form of amplitude-time representation (ATR) (Fig. 1).

The ATR is the amplitude envelope of a series of pulses modulated according to the radar ST. Knowledge of this feature can be useful in deciding on the type of radar and its mode of radar operation.

In radiolocation, the following ST are the most popular in practice: circular, sector, helical and raster. The corresponding ATR views are shown in Table 1.

Table 1

ST and their ATRs

ST	Antenna pattern beam movement	ATR
Circular		
Sector		
Helical		
Raster		

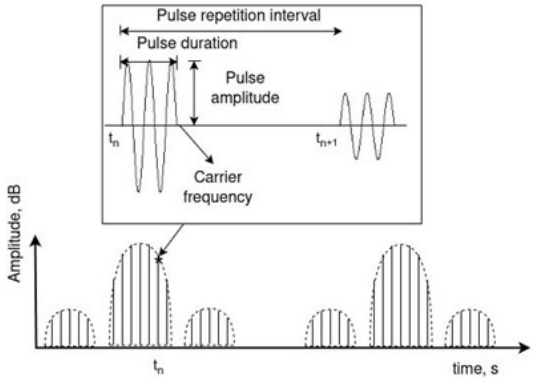


Fig. 1. ATR of the radar signal (the input of radio monitoring station)

In the circular scan, the radar makes a circular rotation in the azimuth plane with a constant velocity Ωα. The ATR of this type of view is characterized by the antenna pattern (AP) width measured at the 3 dB level and the scanning period Ta.

The sector scan of space, in contrast to the circular one, is limited by a certain azimuth angle. The time interval between the main beams of the AP at the receiving point of the monitoring station has two values, except for the case when the AP will be located in the middle of the scanning sector.

In a helical scan, the spatial motion of the AP is a combination of circular rotation of the diagram in the horizontal plane with velocity Ωα and its gradual movement in the vertical plane with velocity Ωβ. In this case, each point of the AP moves along a line close to a helical line.

Raster scan is a type of helical scan with limitation of the scanning sector in the azimuthal plane.

The ST discussed above belong to the class of mechanically scanned radars.

In addition to mechanical scanning (MS), modern radars use electronic scanning (ES) with implemented phased arrays (IPA) to change the spatial orientation of the main beam of the AP. A distinction is made between IPA with one-dimensional and two-dimensional scanning or, in other words, antennas with beam motion in one plane and antennas with beam motion in two planes.

The main difference between MS and ES is that the latter allows the beam direction of the radar antenna to be changed rapidly, almost inertia-free.

Features for distinguishing radar stations by ST

Let us represent the pulse series amplitude envelope (ATR) at the input of the monitor station as an N-dimensional sample:

$$\mathbf{Y} = \left(\mathbf{y}_1,\ \mathbf{y}_2,\ \ldots,\ \mathbf{y}_n,\ \ldots,\ \mathbf{y}_N\right),$$

where $\mathbf{y}_n = \left(\tilde{a}\left[n\right],\ t\left[n\right]\right)^{\tilde{a}[n]=\frac{a[n]}{\max(a[n])}}$ is the normalized amplitude of the n-th pulse $a\left[n\right]$, t_n is the arrival time of the nth pulse, N is the number of pulses (samples) in the sample Y, $n = \overline{1, N}$.

In this paper, two features are used to classify radars according to the way they scan space: F_1 – the similarity coefficient and F_2 – the amplitude differences wobble $\left\{\tilde{a}\left[n\right]\right\}$.

The feature F_1, which allows for distinguishing radar with MS from radar with ES, is calculated by the formula:

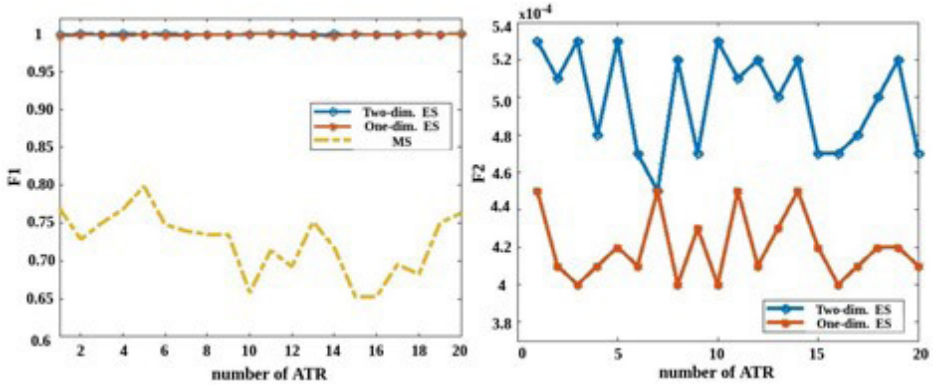


Fig. 2. Result of estimation of informativeness of features F_1 and F_2, allowing for recognizing radars by the method of space scanning into three classes: MS, one-dimensional ES and two-dimensional ES

$$F_1 = \frac{\sum_{j=1}^{N-2} S_1[j] S_2[j]}{\left(\sqrt{\sum_{j=1}^{N-2} S_1^2[j]} \times \sqrt{\sum_{j=1}^{N-2} S_2^2[j]} \right)}, \quad (1)$$

where $S_1[j] = \max\left(\left|d_1[j+1]\right|, \left|d_2[j]\right|\right)$, $S_2[j] = \left|d_2[j]\right|$, …, $j = 1,2,\ldots,N-2$, $d_1[n] = \tilde{a}[n+1] - \tilde{a}[n]$, $n = 1,2,\ldots,N-1$ are the amplitude differences of the first order, $d_2[n] = d_1[n+1] - d_1[n]$, $n = 1,2,\ldots,N-2$ are the second-order amplitude differences $S_1[i] > 0$, $S_2[i] > 0$, $0 \le F_1 \le 1$.

The feature F_2, which allows for distinguishing a one-dimensional ES from a two-dimensional one, is calculated in two steps. First, the sequence $\{d_1[n]\}$ is used to form a series of elements whose values are less than the set value (in this paper, it is assumed to be 0.0003), and then the variance of the series obtained in this way is calculated.

The result of evaluating the informativeness of features F_1 and F_2 by simulation modelling method is illustrated in Fig. 2.

Fig. 2 shows that the feature F_1 allows for distinguishing ES from MS by comparing its value with a certain threshold λ_1. The value of feature F_2 for one-dimensional ES is much larger than for two-dimensional ES, which proves the possibility of its use in classification by comparing it with a certain threshold λ_2.

Decision making on the ST of the MS radar is based on the application of the following three features: F_3 is the number of main AP beams, F_4 is the magnitude of the envelope amplitude wobble of the pulse series and F_5 is the time interval between the main AP beams.

The feature F_3, which allows for distinguishing circular ST from sector ST, is calculated using the normalized mutual correlation function between the signal $\{\tilde{a}[m]\}$ and the signal $\{b[m]\}$, representing the samples of the main AP beam with maximum amplitude:

$$r[k] = \frac{\sum_{n=1}^{V} \tilde{a}[n+k] b[n]}{\sqrt{\sum_{n=1}^{V} \tilde{a}^2[n+k] \cdot \sqrt{\sum_{n=1}^{V} b^2[n]}}}, \quad k = 1,2,\ldots,N-V, \quad (2)$$

where V is the number of samples of the signal $\{b[n]\}$.

Selection of the main AP beam $\{b[n]\}$ on the signal $\{\tilde{a}[n]\}$ is performed by finding the reference $\tilde{a}[n]$ with the maximum value and the nearest samples (left and right), whose values are greater than the set threshold λ_3.

The feature F_4 is calculated as the difference between the maximum and minimum value of the amplitude of the main beams on the signal $\{\tilde{a}[n]\}$:

$$F_4 = \max(\tilde{a}[n]) - \min(\tilde{a}[n]). \quad (3)$$

The feature F_5 is found as the ratio of the maximum time interval between the maxima of the main beams on the signal $\{\tilde{a}[n]\}$ to the minimum one:

$$F_3 = \Delta T_{\max} / \Delta T_{\min},$$

where $\Delta T_{\max} \Delta T_{\min}$ are the maximum and minimum interval between the maxima of the main beams.

The result of evaluating the informativeness of features F_3–F_5 conducted by simulation modelling method is illustrated in Fig. 3–5.

The figure above shows that the radar with circular ST is characterized by only one beam per period, whereas with the sector ST two beams are observed (Fig. 3). The value of the feature F_4 for sector ST fluctuates around the value equal to zero (Fig. 4). This allows for distinguishing sector and circular scan from raster and helical ones by this feature by comparing it with the threshold λ_4. The feature F_5 is constant for circular and helical ST and varies for raster ST (Fig. 5). Therefore, it can be used to distinguish raster ST from helical ST by comparing it with the threshold λ_5.

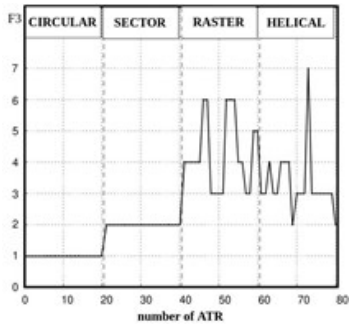


Fig. 3. Result of estimating the informativeness of feature F_3 to distinguish circular ST from sectoral ST

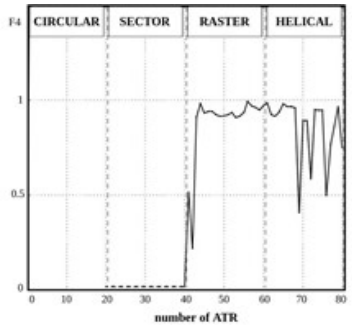
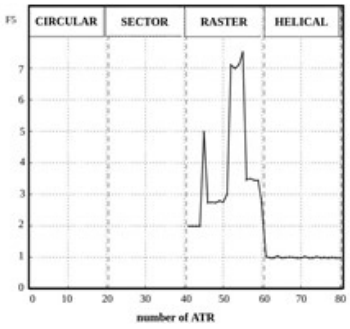


Fig. 4. Result of estimating the informativeness of feature F_4 to distinguish sectoral and circular ST from helical and raster ST



Result of estimating the informativeness of feature F_5 to distinguish raster ST from helical ST

Algorithm for recognizing radar stations by ST

The proposed algorithm includes three main steps: calculating the antenna scan period T_a, extracting the recognition features and deciding on the space scanning method and ST.

Stage 1. The estimation of the ST period is carried out using the samples of the normalized autocorrelation function, which are calculated by the formula:

$$r[k] = \frac{\sum_{n=1}^{W} a[n]a[n+k]}{\sqrt{\sum_{n=1}^{W} a^2[n] \cdot \sqrt{\sum_{n=1}^{W} a^2[n+k]}}}, \quad k = 1,2,\ldots,N-W, \tag{4}$$

where $a[n]$ is the envelope amplitude of a pulse series at the moment of arrival of the n-th pulse, k is the serial number of the shift (delay) in time, W is the length of the window, which is chosen based on the requirements of accuracy and computational complexity (in modelling W is taken equal to N/2). As an estimate of T_a (in counts) is taken the value of k at which $r[k]$ is greater than the set value of 0.98.

Stage 2. The features are calculated using formulas (1–3).

Stage 3. Classification of radars by the method of space scanning and recognition of the ST is carried out using the decision tree method. The structure of the decision tree using the features discussed above is shown in Fig. 6.

The scanning method and the ST is represented in a tree structure by finite nodes.

The traversal of the decision tree starts from the root of the tree (feature F_1 is checked), then a sequential traversal of child nodes is performed (features F_2–F_5 are checked) until an appropriate solution is obtained for a given node.

Simulation

To illustrate described processing algorithm, simulation was carried out. As input data, 320 ATR signals of radars with different types of scanning and space scanning methods were modelled. From them, to check the quality of radar classification by scanning method, the following signals were selected: 40 – with MS and 80 (by 40 for each method of electronic scanning) – with ES, and 200 signals (50 for each type of scanning) were selected to check the quality of recognition of ST. The F_1–F_5 features were compared with pre-calculated thresholds, $\lambda_1 = 0.8$, $\lambda_2 = 0.0005$, $\lambda_3 = 1.06$, $\lambda_4 = 1.1$, $\lambda_5 = 1.3$.

Tables 2 and 3 show the results of the simulations.

A value equal to the ratio of the number of trials m in which a given event (radar scan method and ST) appeared to the total number n of trials actually conducted was used as an indicator of decision quality.

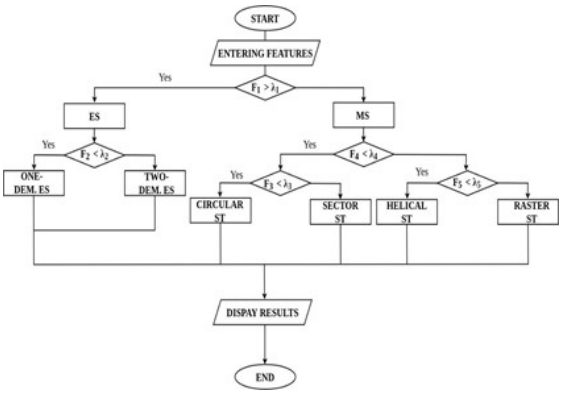


Fig. 6. Algorithm for recognizing the ST

Table 2

Result of the recognition quality evaluation of the space scanning method

Scanning method	One-dimensional ES	Two-dimensional ES	MS	Probability, %
One-dimensional ES	80	18	2	80
Two-dimensional ES	1	95	4	95
MS	1	1	98	98

Table 3

Result of evaluation of the quality of recognition of the ST

Reliability, % / The type of RO is	Circular	Sectoral	Helical	Raster
Circular	97	3	0	0
Sectoral	3	97	0	0
Helical	0	0	85	15
Raster		0	14	86

The recognition quality was assessed at a signal-to-noise ratio R equal to 10 dB. As can be seen from Tables 2 and 3, the percentage of correct decisions in assigning the radar to one of the three classes (Table 2) was at least 80%, and the quality of recognition of the ST was at least 85%.

Conclusion

Circular and sector ST recognition showed a high probability of recognition equal to 97%. The validity for the helical and raster type was 85% and 86%, respectively, provided that there was no false recognition as a sector or circular ST. Since raster scanning is a kind of helical scanning, such results can be justified by the position of the monitoring station in the middle of the radar scanning sector.

The analysis of the available sources of information on the issue of evaluation and selection of features for recognition of the type of scanning can be used in the modernization of existing and development of prospective monitor stations.

The obtained data show that the considered features and recognition algorithm allows to classify radar stations by the method of space scanning and determine their type of radar stations. Further research should be focused on testing the algorithm in real conditions of radio monitoring station operation.

REFERENCES

1. **Wiley R.G.** *ELINT: The Interception and Analysis of Radar Signals*. Norwood, MA: Artech House, Inc., 2006.

2. **Dogancay K.** Online optimization of receiver trajectories for scan-based emitter localization. *IEEE Transactions on Aerospace and Electronic Systems*, 2007, Vol. 43, No. 3, Pp. 1117–1125. DOI: 10.1109/TAES.2007.4383601

3. **Greer T.H.** *Automatic recognition of radar scan type*, US Patent, No. 6697007 B2 (2004).

4. **Kim Y.-H., Kim W.-J., Song K.-H., Han J.-W., Kim H.-N.** Modeling of a radar signal for scan pattern. *MILCOM 2009 – 2009 IEEE Military Communications Conference*, 2009, Pp. 1–6. DOI: 10.1109/MILCOM.2009.5379909

5. **Çolakoğlu B., Eravcı B., Bayri A.** Emitter localization based on antenna scan properties. *2011 IEEE 19th Signal Processing and Communications Applications Conference* (*SIU*), 2011, Pp. 554–557. DOI: 10.1109/SIU.2011.5929710

6. **Hman H.** Scan-based emitter passive localization. *IEEE Transactions on Aerospace and Electronic Systems*, 2007, Vol. 43, No. 1, Pp. 36–54. DOI: 10.1109/TAES.2007.357153

7. **Hman H., Dogancay K.** Passive localization of scanning emitters. *IEEE Transactions on Aerospace and Electronic Systems*, 2010, Vol. 46, No. 2, Pp. 944–951. DOI: 10.1109/TAES.2010.5461671

8. **Lunden J., Koivunen V.** Automatic radar waveform recognition. *IEEE Journal of Selected Topics in Signal Processing*, 2007, Vol. 1, No. 1, Pp. 124–136. DOI: 10.1109/JSTSP.2007.897055

9. **Zaczek T.E.** *Antenna scan pattern generator*, US Patent, No. 4327417A (1980).

10. **Barshan B., Eravci B.** Automatic radar antenna scan type recognition in electronic warfare. *IEEE Transactions on Aerospace and Electronic Systems*, 2012, Vol. 48, No. 4, Pp. 2908–2931. DOI: 10.1109/TAES.2012.6324669

INFORMATION ABOUT AUTHORS / СВЕДЕНИЯ ОБ АВТОРАХ

Victoria A. Ivannikova
Иванникова Виктория Александровна
E-mail: ivannikova.vicky@yandex.ru
ORCID: https://orcid.org/0009-0006-6987-7144

Vladimir F. Korotkov
Коротков Владимир Фёдорович
E-mail: diofant2912@gmail.com

Submitted: 28.02.2025; Approved: 30.06.2025; Accepted: 04.08.2025.
Поступила: 28.02.2025; Одобрена: 30.06.2025; Принята: 04.08.2025.

Computing, Telecommunication and Control, 2025, Vol. 18, No. 3, Pp. 89–101.
Информатика, телекоммуникации и управление. 2025. Том 18, № 3. С. 89–101.

Research article
DOI: https://doi.org/10.18721/JCSTCS.18308
UDC 621.37

REDUCED VOLTAGE STRESS CLASS E POWER AMPLIFIER OPERATING A COMPLEX IMPEDANCE LOAD: A PERFORMANCE ANALYSIS

H.D. Pham ✉ , *V.A. Sorotsky, R.I. Zudov*

Peter the Great St. Petersburg Polytechnic University,
St. Petersburg, Russian Federation

✉ phamduc2511997@gmail.com

Abstract. Class E power amplifiers (PAs) attract the interest of experts involved in the development of communication and telecommunications equipment due to their high efficiency. However, the high voltage stress across transistors, which exceeds the supply voltage by 3.6–4 times, limits the output power of such amplifiers. An alternative to solve this issue could be a PA in which the peak voltage across transistor is reduced by 2 times, but still maintains the main advantages of traditional Class E, such as zero-voltage switching (ZVS) and zero-derivative voltage switching (ZDVS). In well-known publications, the study of the characteristics of PAs with lower voltage across transistors is limited to the particular case of a real impedance load. However, this condition may not be true when the PA operates in a frequency band, which will inevitably lead to errors in calculating their characteristics. The objective of the paper is to develop an analytical model of Class E PA with reduced voltage stress when operating with complex impedance load. The adequacy of the analytical model is confirmed by simulation, which shows that the relative error in the calculation of the main characteristics of the PA does not exceed 6.5%. The issues of synthesizing a filtering and matching circuit have been considered that ensures expansion of the frequency band at specified rated values in output power and voltage stress in transistor turned-on moment.

Keywords: power amplifier, class E, switching power losses, complex impedance load, harmonic balance method

Citation: Pham H.D., Sorotsky V.A., Zudov R.I. Reduced voltage stress Class E power amplifier operating a complex impedance load: A performance analysis. Computing, Telecommunications and Control, 2025, Vol. 18, No. 3, Pp. 89–101. DOI: 10.18721/JCSTCS.18308

Научная статья
DOI: https://doi.org/10.18721/JCSTCS.18308
УДК 621.37

ХАРАКТЕРИСТИКИ УСИЛИТЕЛЯ МОЩНОСТИ КЛАССА Е С ПОНИЖЕННЫМ НАПРЯЖЕНИЕМ НА ТРАНЗИСТОРАХ ПРИ РАБОТЕ НА КОМПЛЕКСНУЮ НАГРУЗКУ

Х.Д. Фам ✉, *В.А. Сороцкий*, *Р.И. Зудов*

Санкт-Петербургский политехнический университет Петра Великого,
Санкт-Петербург, Российская Федерация

✉ phamduc2511997@gmail.com

Аннотация. Усилители мощности (УМ) класса Е вызывают интерес у специалистов, занимающихся разработкой аппаратуры связи и телекоммуникаций благодаря высокому КПД. Однако высокое пиковое напряжение на транзисторах, превышающее напряжение питания в 3,6...4 раза, ограничивает выходную мощность таких усилителей. Альтернативой этим устройствам могут стать УМ класса Е с пониженным напряжением на транзисторах, у которых максимальное напряжение на транзисторах снижено в 2 раза, но при сохраняются основные достоинства режима класса Е, такие как ZVS и ZVDS. В известных публикациях исследование характеристик УМ с пониженным напряжением на транзисторах ограничивается случаем нагрузки с вещественным импедансом. Однако данное условие нарушается при работе УМ в полосе частот, что неизбежно приведет к погрешности при расчете характеристик УМ. Целью настоящей публикации является разработка аналитической модели, позволяющей определить характеристики УМ в общем случае при работе на нагрузку с комплексным импедансом. В работе представлены результаты проверки достоверности аналитической модели, которая позволила установить, что ее погрешность не превышает 6,5%. Рассмотрены вопросы синтеза согласующей цепи, обеспечивающей расширение полосы частот УМ при заданных допустимых значениях непостоянства выходной мощности и скачка напряжения на транзисторах в момент коммутации.

Ключевые слова: усилитель мощности, класс Е, коммутационные потери мощности, нагрузка с комплексным импедансом, метод гармонического баланса

Для цитирования: Pham H.D., Sorotsky V.A., Zudov R.I. Reduced voltage stress Class E power amplifier operating a complex impedance load: A performance analysis // Computing, Telecommunications and Control. 2025. Т. 18, № 3. С. 89–101. DOI: 10.18721/JCSTCS.18308

Introduction

Improving power efficiency and weight and size parameters of radio transmitters can be achieved by using switched-mode power amplifiers (PAs). However, without a significant decrease in switching power losses, it doesn't guarantee achieving high efficiency. Therefore, in recent years there has been great interest in the study of PAs operating in classes DE and E where, under certain conditions, it is possible to ensure *zero-voltage-switching* (*ZVS*) and *zero-derivative-voltage-switching* (*ZDVS*), thus eliminating switching losses and insuring high efficiency [1–3].

Compared to Class DE, Class E PAs demonstrates distinct advantages:

1) higher operating frequency;
2) noticeably easier formation of signals to drive transistors.

Class E PAs lose to their competitors is in more intense operating conditions of the transistors: if in Class DE PAs the peak voltage across transistors is equal to the supply voltage, then in Class E PAs it increases by 3.6–4 times.

Proceeding from the above, when designing a PA with increased output power, the circuit considered in [4, 5] is of great interest. On the one hand, it has all the advantages of a Class E PA, and on the other

hand, it has a lower transistor voltage stress at least as much as 2 times. When analyzing the characteristics of this PA, the authors limited themselves to considering only a particular case of a resistive load. However, when the transmitter operates in a frequency band, this condition is violated, which will inevitably lead to errors in the obtained results.

Taking into account the said above, the objective of the research presented in the paper was to develop an analytical model of a Class E PA with reduced voltage stress on transistors (for the sake of abbreviation, we will further designate this Class as EV), while operating on a complex impedance load and to determine its main characteristics.

Analytical model of Class EV PA with a complex impedance load

The schematic of Class E_V PA is shown in Fig 1, *a*, which consists of transistors $VT1$, $VT2$; inductors L_1, L_2 and a complex impedance load $Z = R + jX$. To simplify a theoretical analysis, the following assumptions were introduced:

1. The transistors' on-state resistance is close to zero, and their switching time is negligible compared to the output oscillation period.
2. The output transistors' capacitance does not depend upon the drain-to-source voltage and has a constant value.
3. Losses in coils are considered negligible.
4. The inductances L_1, L_2 are the same and equal to $L/2$.
5. Given the use of a filter at the PA output for higher-harmonic suppression, the sinusoidal approximation for the load Z (Fig.1, *a*) current waveform can be applied [6–8]:

$$i_0 = I_m \sin\left(\omega t + \varphi\right),$$

with unknown amplitude I_m and initial pharse φ can be determined during analysis (Fig. 1, *b*).

Based on the above assumptions, the analysis of the processes in PA can be implemented using the equivalent circuit shown in Fig. 1, *b*, replacing transistors with switches $S1$, $S2$ with capacitors in parallel $C_1 = C_2 = C$.

Let's consider the steady-state operating mode of the circuit and assume that in the time interval $0 < \omega t \leq \pi$ the upper transistor is in the on-state (the switch $S1$ is closed), and the lower transistor, accordingly, is in the off-state. Then applying in the circuit in Fig. 1, *b* Kirchhoff's current and voltage laws, one can compose a system of equations:

a) *b*)

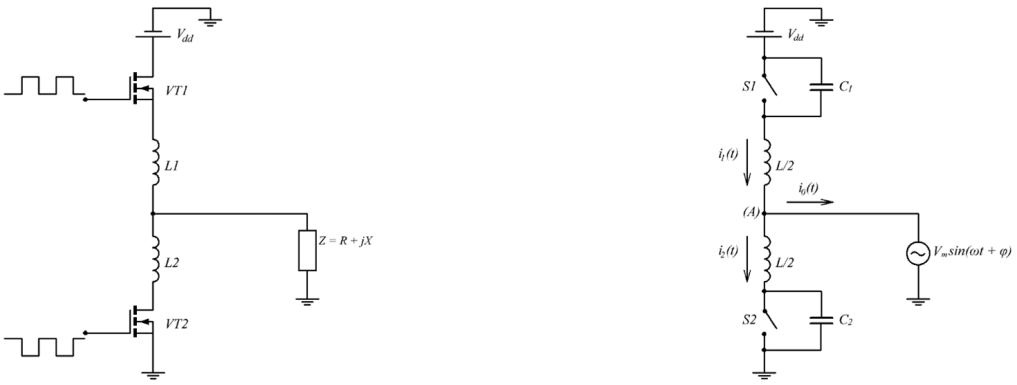


Fig. 1. PA Class E_V: functional circuit (*a*); equivalent circuit (*b*)

$$\begin{cases} V_{dd} = \upsilon_{L1}(t) + \upsilon_{L2}(t) + \upsilon_{C2}(t) & (1) \\ i_1(t) = i_2(t) + i_0(t) & (2) \end{cases},$$

where V_{dd} – supply voltage; $\upsilon_{L1}(t)$, $\upsilon_{L2}(t)$, $\upsilon_{C2}(t)$, – voltage across inductors L_1, L_2 and output capacitor transistor C_2.

Solving (1), (2) together, after necessary transformations we get the voltage across transistor $VT2$ during the time interval $0 < \omega t \le \pi$:

$$\begin{gathered} \upsilon_2(\theta) = \pi\nu\sin(\nu\theta) - \nu I_0 \sin\varphi\sin(\nu\theta) + 1 - \cos(\nu\theta) + \\ + \frac{\nu I_0 \sin\varphi}{2(1-\nu^2)}\sin(\nu\theta) - \frac{\upsilon^2 I_0 \sin\varphi\sin\theta}{2(1-\nu^2)} - \frac{\nu^2 I_0 \cos\varphi}{2(1-\nu^2)}\left[\cos(\nu\theta) - \cos\theta\right], \end{gathered} \tag{3}$$

where

$$\begin{cases} \nu = \dfrac{\omega_0}{\omega} = \dfrac{1}{\omega\sqrt{LC}}, \\ \theta = \omega t, \\ I_0 = \dfrac{\omega L I_m}{V_{dd}} \\ \upsilon_2(\theta) = \dfrac{\upsilon_{C2}(\theta)}{V_{dd}}. \end{cases} \tag{4}$$

The current through $VT1$ at the time interval $0 < \omega t \le \pi$, when $S1$ is in the on-state, is given by:

$$\begin{gathered} i_1(\theta) = \frac{\pi\cos(\nu\theta)}{I_0} + \frac{\sin(\nu\theta)}{I_0\nu} - \sin\varphi\cos(\nu\theta) + \frac{\nu\cos\varphi\sin(\nu\theta)}{2(1-\nu^2)} - \\ - \frac{\cos\varphi\sin\theta}{2(1-\nu^2)} + \frac{\sin\varphi}{2(1-\nu^2)}\left[\cos(\nu\theta) - \cos\theta\right] + \sin(\theta+\varphi). \end{gathered} \tag{5}$$

Similarly, during the time interval $\pi < \theta \le 2\pi$ when $S2$ is off-state, the voltage and the current through the lower transistor $VT2$ satisfy relations:

$$\begin{gathered} \upsilon_1(\theta) = \pi\nu\sin(\nu\theta) + I_0\sin\varphi\sin(\nu\theta) + 1 - \cos(\nu\theta) - \\ - \frac{\nu I_0\sin\varphi}{2(1-\nu^2)}\sin(\nu\theta) + \frac{\nu^2 I_0\sin\varphi\sin\theta}{2(1-\nu^2)} + \frac{\nu^2 I_0\cos\varphi}{2(1-\nu^2)}\left[\cos(\nu\theta) - \cos\theta\right]; \end{gathered} \tag{6}$$

$$\begin{gathered} i_2(\theta) = \frac{\pi\cos(\nu\theta)}{I_0} + \frac{\sin(\nu\theta)}{I_0\nu} + \sin\varphi\cos(\nu\theta) - \frac{\nu\cos\varphi\sin(\nu\theta)}{2(1-\nu^2)} + \\ + \frac{\cos\varphi\sin\theta}{2(1-\nu^2)} - \frac{\sin\varphi}{2(1-\nu^2)}\left[\cos(\nu\theta) - \cos\theta\right] - \sin(\theta+\varphi). \end{gathered} \tag{7}$$

The unknown parameters φ and I_0 can be determined using relations (3) and (6). The investigation concentrates on the practically important "soft- switching", where transistor switching occurs under both *ZVS* (*zero voltage switching*) and *ZVDS* (*zero voltage derivative switching*) [7, 9–11]. The first condition ensures that the output capacitance is completely discharged before turn-on (typical for Class E PA), resulting in zero switching losses. The second condition assumes that the derivative of the voltage on the transistor output capacitance is equal to zero, which means that there is no current when the switching starts. This condition is necessary to eliminate voltage surges in parasitic inductances and due to that to reduce high-frequency oscillations.

The analytical solutions for φ(ν) and $I_0(\nu)$ are represented in Table 1.

Table 1

The analytical solutions

$\nu = \frac{\omega_0}{\omega}$	0.8	0.9	1.0	1.1	1.2	1.3	1.4
φ, deg	–35.8	–36.9	–38.2	–39.7	–41.5	–43.8	–49.7
$I_0 = \frac{\omega L I_m}{V_{dd}}$	3.7	2.7	2.0	1.6	1.2	0.9	0.7

Fig. 2 displays the normalized voltage waveforms across transistor capacitances $\left(\upsilon_1(\theta),\ \upsilon_2(\theta)\right)$ and corresponding current waveforms $\left(i_1(\theta),\ i_2(\theta)\right)$ as functions of phase angle θ, calculated for various frequency parameter ν values.

As shown in Fig. 2, the capacitor voltage waveforms and transistor currents both satisfy the "*soft-switching*" conditions. At the same time, the voltage stress on transistor is approximately 2 times less than that of PA Class E. This allows increasing the output power compared to the Class E PA while maintaining high efficiency.

Calculation of load impedance to provide transistor "*soft-switching*"

The voltage at node A (Fig. 1, *b*) is equal to:

$$U(\theta) = \begin{cases} \frac{1}{2}\pi\nu\sin(\nu\theta) - \frac{1}{2}\nu I_0\sin\varphi\sin(\nu\theta) + 1 - \frac{1}{2}\cos(\nu\theta) - \frac{\nu^2 I_0\cos\varphi\cos(\nu\theta)}{4(1-\nu^2)} + \frac{\nu^2 I_0\cos\varphi\cos\theta}{2(1-\nu^2)} - \\ -\frac{I_0\cos\varphi\cos\theta}{4(1-\nu^2)} + \frac{\nu^2 I_0\sin\varphi}{4(1-\nu^2)}\cdot\sin(\nu\theta) + \frac{I_0\sin\varphi\sin\theta}{4(1-\nu^2)} - \frac{\nu^2 I_0\sin\varphi\sin\theta}{2(1-\nu^2)}, \quad 0 < \theta \le \pi; \\ -\frac{1}{2}\pi\nu\sin(\nu\theta) - \frac{1}{2}\nu I_0\sin\varphi\sin(\nu\theta) + \frac{1}{2}\cos(\nu\theta) - \frac{\nu^2 I_0\cos\varphi\cos(\nu\theta)}{4(1-\nu^2)} - \frac{I_0\cos\varphi\cos\theta}{4(1-\nu^2)} + \\ +\frac{\nu^2 I_0\cos\varphi\cos\theta}{2(1-\nu^2)} + \frac{\nu^2 I_0\sin\varphi}{4(1-\nu^2)}\cdot\sin(\nu\theta) + \frac{I_0\sin\varphi\sin\theta}{4(1-\nu^2)} - \frac{\nu^2 I_0\sin\varphi\sin\theta}{2(1-\nu^2)}, \quad \pi < \theta \le 2\pi. \end{cases} \tag{8}$$

In order to compose the harmonic balance equation, the voltage $U(\theta)$ amplitude and phase of the first harmonic are needed:

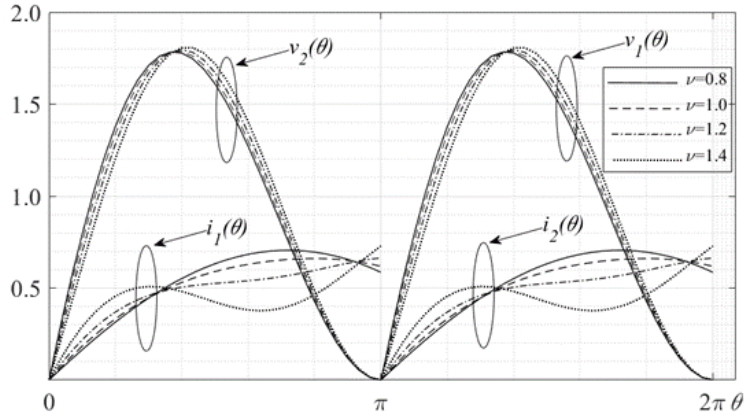


Fig. 2. Time-domain diagrams of the voltage drop and current in transistor at different ν values

$$U_1 = \sqrt{a_1^2 + b_1^2}, \quad \varphi_U = \text{arctg}\frac{b_1}{a_1}, \tag{9}$$

where a_1 and b_1 are the coefficients of the Fourier series:

$$a_1 = \frac{1}{\pi}\int_0^{2\pi} U(\theta)\cos\theta d\theta, \tag{10}$$

$$b_1 = \frac{1}{\pi}\int_0^{2\pi} U(\theta)\sin\theta d\theta. \tag{11}$$

The modulus and phase of the load impedance are equal to:

$$|Z| = \frac{|U_1|}{|I_m|} = \frac{\sqrt{a_1^2 + b_1^2}}{I_m}; \tag{12}$$

$$\varphi_z = \varphi_U - \varphi. \tag{13}$$

Fig. 3 shows the normalized load impedance (real $R_N = ReZ_L / ReZ_0$ and imaginary $X_N = ImZ_L / ReZ_0$ parts) in the frequency band $[f_L, f_U]$ with overlap coefficient $K_\omega = f_L, f_U$, where f_U and f_L are the upper and the lower bounds of the operational frequency band, respectively; ReZ_0 is the real part of load impedance at the upper bound f_U.

The derived dependencies enable parametric synthesis of the load circuit satisfying the required input impedance variation in the frequency band and attenuation above it. Based on the results of solving this synthesis problem it becomes possible to determine the frequency overlap factor K_ω.

Validation of the analytical model

The analytical expressions were validated through a comprehensive simulation using the model shown in Fig. 4, where $S1$, $S2$ – active devices (switches); $V1$ – supply voltage; $V2$, $V3$ – pulse generators; $C1$, $C1$ – capacitances that consider transistors' output ones; $L1$, $L1$ –inductors; Lx, Rl – imaginary and real parts of impedance load, respectively. Fig. 5 plots voltage and current time diagrams.

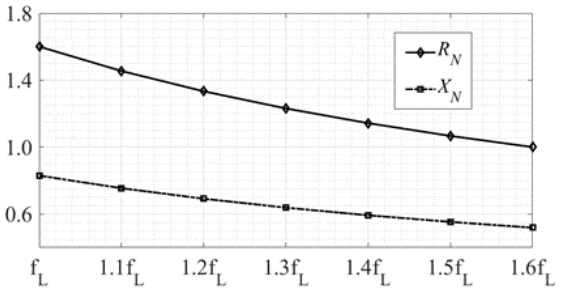


Fig. 3. Normalized impedance load versus frequency

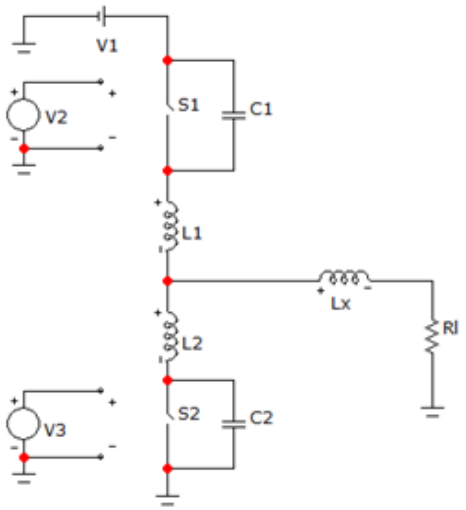


Fig. 4. Class E_V PA simulation model

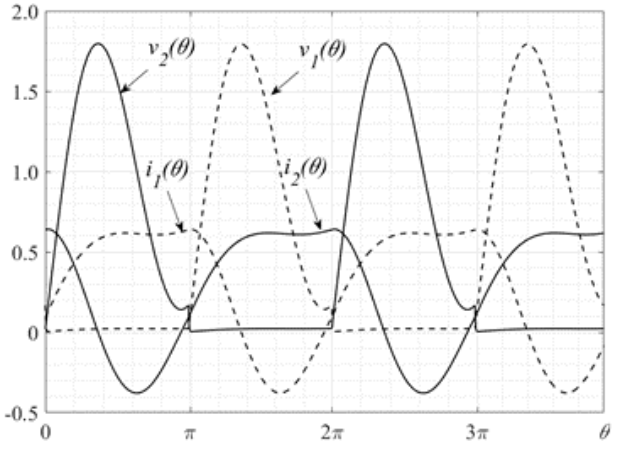


Fig. 5. Simulated curves of voltage and current across switches *S*1, *S*2

Table 2 presents the results obtained using simulation modeling in the frequency band $[fL, f_U]$, where Z_L – normalized complex impedance load; *3f*, *5f* – relative level of 3rd and 5th harmonics; Π – transistors' voltage stress; Δυ – voltage jump on transistors, normalized by supply voltage.

Table 2

Simulation results

F	f_{H}	$1.2f_{\text{H}}$	$1.4f_{\text{H}}$	$1.6f_{\text{H}}$
Z_L	$1.60 + j0.83$	$1.33 + j0.69$	$1.14 + j0.59$	$1.00 + j0.52$
3f, dBc	–24.5	–24.5	–24.5	–24.5
5f, dBc	–40.4	–40.3	–40.4	–40.5
П	1.83	1.82	1.81	1.81
Δυ, %	6.1	6.5	6.4	4.9

As evidenced by the comparative analysis of data in Table 2, the analytical model demonstrates excellent agreement with simulation results, exhibiting: 2.0% relative error in voltage stress (П), 6.5% deviation in voltage jumps (Δυ). However, the harmonic distortion levels remain non-negligible, with the 3rd harmonic at –24 dBc and the 5th harmonic at –40 dBc. This means that the synthesis of the PA filtering and matching circuit must be carried out not only based on the condition of implementing the required law of behavior of the real and imaginary parts of its input impedance, but, in addition to this, considering the required attenuation of a given set of higher harmonics.

Synthesis of a filtering and matching circuit

The requirement to suppress higher harmonics to compliant levels (e.g., < –40 dBc for $n \geq 3$) inherently constrains possible structure options of filtering and matching circuit (FMC) in favor of low-pass filters (LPF) [12, 13]. Taking this constraint into account, the task of FMC synthesis can be formulated as the parametric optimization problem of a number LPF variants of Elliptic and Chebyshev types of diffirent orders. The structures of the Elliptic (E) LPF of the 3rd and 5th orders are shown in Fig. 6, *a*, *b* as well as the Chebyshev LPF of the 5th order is shown in Fig. 7.

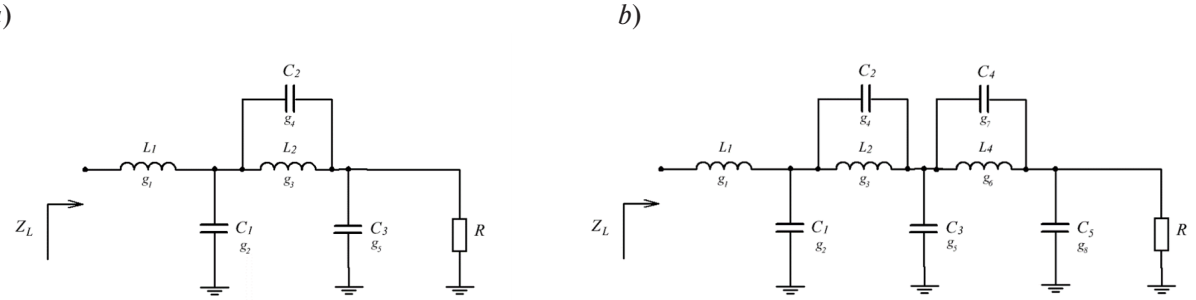


Fig. 6. The structures of 3rd and 5th Order Elliptic LPF

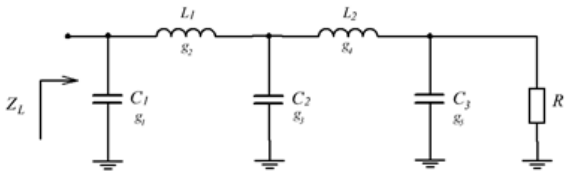


Fig. 7. The structure of 5th Order Chebyshev LPF

The parametric optimization problem can be represented as:

$$\vec{x}^* = \arg\min_{\vec{x}\in D} F\left(\vec{x}, f\right), \tag{14}$$

where $D = \left\{ \begin{array}{c} x_i > 0,\ i = 1 \ldots n \\ \vec{x} \in R^n \left| -2\ dB < 20\log\left|H\left(\vec{x}, f\right)\right| < 2\ dB,\ \text{when}\ f < f_c \right. \\ 20\log\left|H\left(\vec{x}, f\right)\right| \le -40\ dB,\ \text{when}\ f \ge f_s \end{array} \right\}$, $\vec{x}$ is the vector of the LPF elements parameters, f_c is the LPF passband frequency, f_s is the LPF stopband frequency. In this case, the objective function will have the form:

$$F\left(\vec{x}, f\right) = \sum_{k=1}^{m} \sqrt{\frac{\left(ReZ_L\left(f_k\right) - ReZ_F\left(\vec{x}, f_k\right)\right)^2}{ReZ_L\left(f_i\right)^2} + \frac{\left(ImZ_L\left(f_k\right) - ImZ_F\left(\vec{x}, f_k\right)\right)^2}{ImZ_L\left(f_i\right)^2}}, \tag{15}$$

where the set $Z_L(f_k)$ corresponds to the required law of change of the FMC input impedance and $Z_F(f_k)$ – the calculated values of the FMC input within the passband.

The normalized LPF elements values obtained in optimization are given in Table 3, and the load voltage spectrum diagrams are shown in Fig. 8–10.

From the analysis of the diagrams in Fig. 8–10 it follows that the output voltage spectrum contains only odd harmonics, the relative amplitudes of which are given in Tables 4–6. The values of the normalized transistor voltage stress (П) are also presented here. A voltage stress deviation when using different types of LPF from the analytical calculation is shown in Fig. 11.

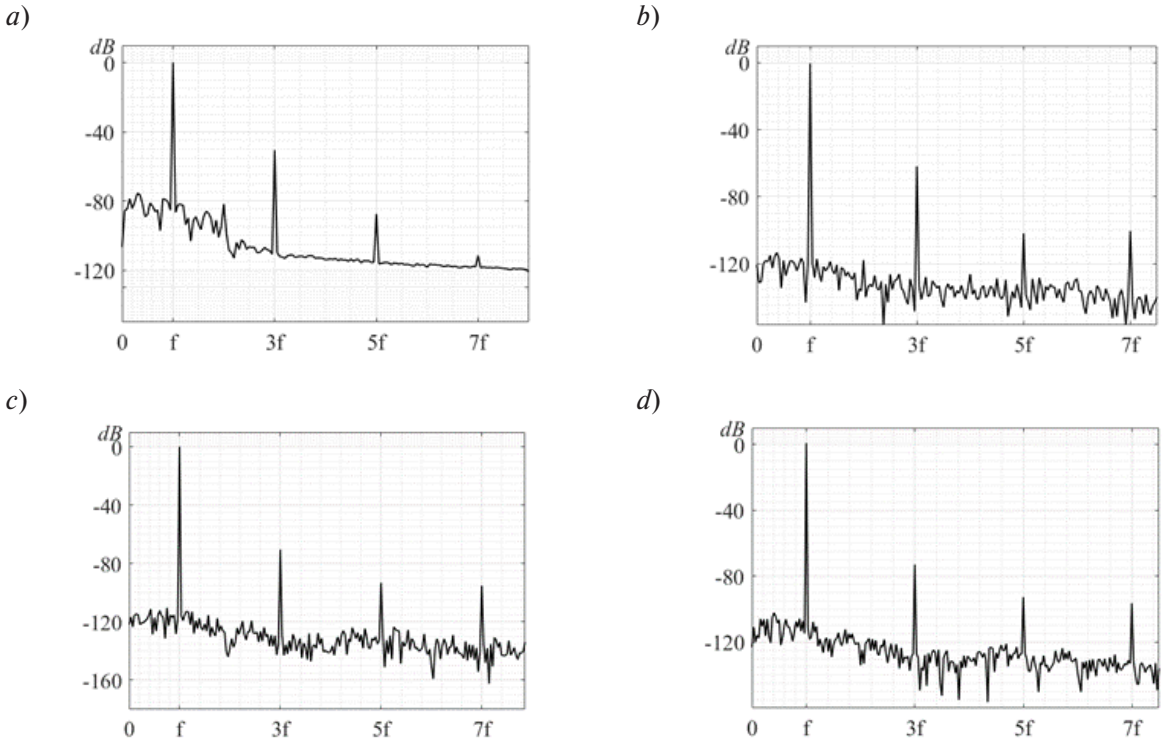


Fig. 8. Load voltage spectrums for a 3rd order elliptic LPF at different operating frequencies: f_L (*a*), $1.2f_L$ (*b*), $1.4f_L$ (*c*), $1.6f_L$ (*d*)

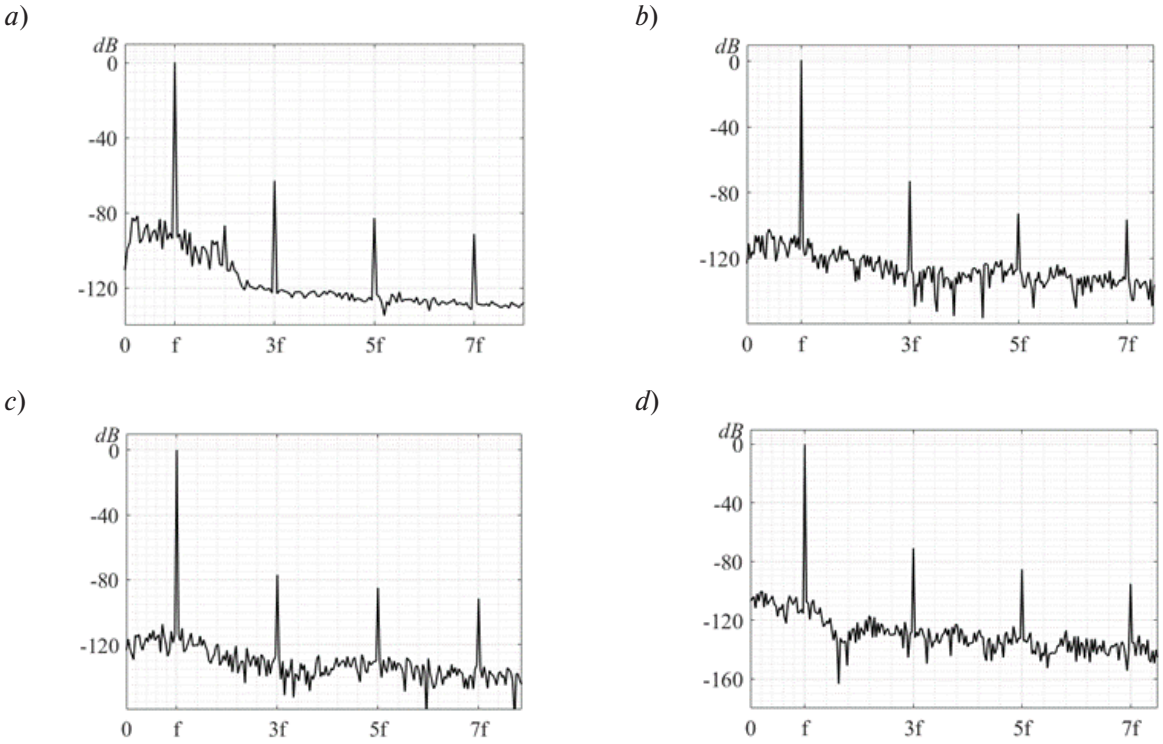


Fig. 9. Load voltage spectrums for a 5th order elliptic LPF
at different operating frequencies: f_L (*a*), $1.2f_L$ (*b*), $1.4f_L$ (*c*), $1.6f_L$ (*d*)

Table 3

Normalized LPF elements values

LPF	Nominal values							
	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8
3rd Order Elliptic	0.704	2.806	1.022	0.061	1.386	–	–	
5th Order Elliptic	0.722	2.734	0.866	0.217	1.039	0.197	1.974	0.489
5th Order Chebyshev	0.051	0.684	2.988	1.034	1.503	–	–	–

Table 4

Characteristics of the PA when implementing FMC on the basis of a 3rd order elliptic LPF

Frequency	Higher-harmonic level, dBc			П
	3	5	7	
f_L	–51	–77	–111	1.92
$1.2f_L$	–54	–91	–100	1.71
$1.4f_L$	–60	–92	–104	1.62
$1.6f_L$	–73	–93	–97	1.86

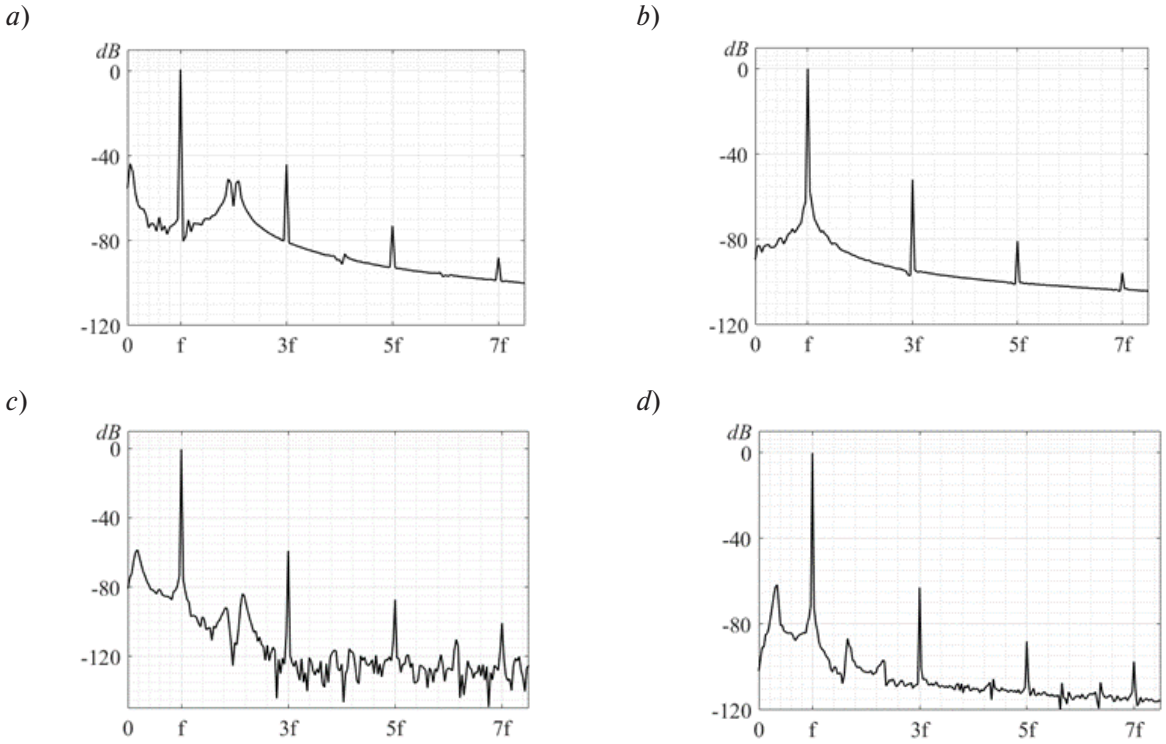


Fig. 10. Load voltage spectrums for a a 5th order Chebyshev LPF at different operating frequencies: f_L (*a*), $1.2f_L$ (*b*), $1.4f_L$ (*c*), $1.6f_L$ (*d*)

A comparative analysis presented in Tables 4–6 and Fig. 11 shows that the FMC based on the 5th order elliptic LPF attenuates higher harmonics better, with the level of the 3rd harmonic being no more than –65 dBc. In the case of 3rd order elliptic LPF, the level of the third harmonic is –51 dBc, and for the 5th order Chebyshev LPF, it is –44 dBc. Using the 5th order elliptic filter also shows results with a smaller theoretical deviation compared to other LPFs. Therefore, it can be concluded that using the 5th order elliptic LPF provides better performance.

Table 5

Characteristics of the PA when implementing FMC on the basis of a 5th order elliptic LPF

Frequency	Higher-harmonic level, dBc			П
	3	5	7	
f_L	–65	–85	–92	1.88
$1.2f_L$	–89	–86	–92	1.74
$1.4f_L$	–73	–84	–90	1.72
$1.6f_L$	–70	–84	–94	1.89

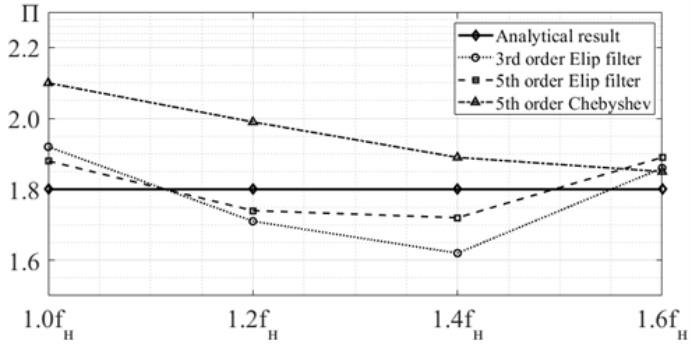


Fig. 11. Comparison of voltage stress deviation when using different types of LPF

Table 6

Characteristics of the PA when implementing FMC on the basis of a 5th order Chebyshev LPF

Frequency	Higher-harmonic level, dBc			П
	3	5	7	
f_L	–44	–72	–87	2.10
$1.2f_L$	–52	–76	–83	1.99
$1.4f_L$	–57	–85	–99	1.89
$1.6f_L$	–63	–88	–97	1.85

Conclusion

Summarizing the results presented in this paper, we highlight the followings:

1) Based on the harmonic balance method, the analytical model of a Class E PA with reduced voltage stress across transistors (Class EV) has been developed which allows for determination of its characteristics when operating a complex impedance load.

2) Using the obtained analytical relations with a known law of load impedance change in the frequency band, the currents and voltages on the elements of the PA can be calculated.

3) The assessment of the adequacy of the proposed analytical model has been carried out which confirmed 2.0% relative error in voltage stress estimation and 6.5% deviation while calculating voltage jump across the transistor at turn-on.

4) A method for synthesizing a filtering and matching circuit is proposed, which allows to expand the operating frequency band of the amplifier with given restrictions on the output power change as well as on the level of switching losses in the transistors, and the permissible level of higher harmonics in the load.

5) The normalized element values of the filtering and matching circuit were optimized, permitting PA operation over a frequency band with an overlap factor of $K_\omega = 1.6$ and decreased switching losses.

6) It has been shown that the relative level of higher harmonics in the output voltage in the case of a 5th-Order Elliptic low-pass filter can be reduce to a value of –65 dB.

REFERENCES

1. **Acar M., Annema A.J., Nauta B.** Analytical design equations for class-E power amplifiers. *IEEE Transactions on Circuits and Systems I: Regular Papers*, 2007, Vol. 54, Pp. 2706–2717. DOI: 10.1109/TCSI.2007.910544

2. **Eroglu A.** *Introduction to RF Power Amplifier Design and Simulation*, 1st ed. Boca Raton: CRC Press, 2016. DOI: 10.1201/9781315215297

3. **Zudov R.I.** Influence of rotational position error stage voltage on the output spectrum in switch power amplifier. *2014 International Conference on Actual Problems of Electron Devices Engineering*, 2014, Pp. 241–243.

4. **Singh G.D., Nallam N.** An RF choke-less Class E power amplifier. *IEEE Transactions on Circuits and Systems II: Express Briefs*, 2020, Vol. 67, No. 11, Pp. 2422–2426. DOI: 10.1109/TCSII.2020.2966552

5. **Yuta I., Hirotaka K.** Choke-less Class -E Oscillator Using p-MOSFET and n-MOSFET. *2023 IEEE 32nd International Symposium on Industrial Electronics* (*ISIE*), June 2023, DOI: 10.1109/ISIE51358.2023.10228146.

6. **Pham H.D., Sorotsky V.A.** Characteristics of Class E power amplifier with complex impedance load. *Computing, Telecommunications and Control*, 2025, Vol. 18, No. 1, Pp. 72–84. DOI: 10.18721/JCSTCS.18106

7. **Sorotsky V., Pham H.D., Zudov R.** Design of a Class E Power amplifier with complex impedance load. *2024 International Conference on Electrical Engineering and Photonics* (*EExPolytech*), 2024, Pp. 76–78. DOI: 10.1109/EExPolytech62224.2024.10755740

8. **Zudov R.** Efficiency of a class DE power amplifier for RF signals with high peak-to-average power ratio. *2019 International Conference on Electrical Engineering and Photonics* (*EExPolytech*), 2024, Pp. 28–30. DOI: 10.1109/EExPolytech.2019.8906856

9. **Grebennikov A.** *RF and Microwave Power Amplifier Design*, 2nd ed. NY: McGraw-Hill, 2015.

10. **Kazimierczuk M.K.** *RF Power Amplifiers*, 2nd ed. Bognor Regis (US): John Wiley & Sons, Ltd, 2015. DOI: 10.1002/9781118844373

11. **Sokal N.O., Grebennikov A.** *Switchmode RF Power Amplifiers*, 1st ed. USA: Elsevier Inc., 2007. DOI: 10.1016/B978-0-7506-7962-6.X5028-X

12. **Chen K., Peroulis D.** Design of highly efficiency broadband class-E power amplifier using synthesized low-pass matching networks. *IEEE Transactions on Microwave Theory and Techniques*, 2011, Vol. 59, No. 12, Pp. 3162–3173. DOI: 10.1109/TMTT.2011.2169080

13. **Kizilbey O.** Design of class-E GaN HEMT power amplifier using elliptic low pass matching network with 86% efficiency. *IEICE Electronics Express*, 2013, Vol. 10, No. 2, Art. no. 20120960. DOI: 10.1587/elex.10.20120960

INFORMATION ABOUT AUTHORS / СВЕДЕНИЯ ОБ АВТОРАХ

Huu Duc Pham
Фам Хыу Дык
E-mail: phamduc2511997@gmail.com
ORCID: https://orcid.org/0009-0004-1628-1772

Vladimir A. Sorotsky
Сороцкий Владимир Александрович
E-mail: sorotsky@mail.spbstu.ru

Roman I. Zudov
Зудов Роман Игоревич
E-mail: rzudov@spbstu.ru
ORCID: https://orcid.org/0000-0001-9738-9291

Submitted: 17.04.2025; Approved: 29.06.2025; Accepted: 03.09.2025.
Поступила: 17.04.2025; Одобрена: 29.06.2025; Принята: 03.09.2025.

Computing, Telecommunication and Control, 2025, Vol. 18, No. 3, Pp. 102–110.
Информатика, телекоммуникации и управление. 2025. Том 18, № 3. С. 102–110.

Research article
DOI: https://doi.org/10.18721/JCSTCS.18309
UDC 621.375.026

SELECTION OF THE OPTIMAL STRUCTURE OF A TRANSFORMER BASED ON SINGLE-TURN ELEMENTS FOR HIGH-POWER SWITCHING TRANSISTOR HARMONIC OSCILLATORS

V.A. Filin[1], *Kh.A. Sattarov*[2], *V.A. Yurova*[1,3] ✉, *A.N. Golovin*[1]

[1] Bonch-Bruevich St. Petersburg State University of Telecommunications, St. Petersburg, Russian Federation;

[2] Tashkent Universirty of information technology named after Muhammad al-Khwarizmi, Samarkand branch, Samarkand, Uzbekistan;

[3] North-Western State Medical University named after I.I. Mechnikov, St. Petersburg, Russian Federation

✉ va-yurova@mail.ru

Abstract. Most electronic devices use power supplies and signal generators. A rationally designed high-power switching radio frequency (RF) generator usually contains one oscillatory circuit, a transformer and a grounded load. This work presents the rationale and calculations for the proposed structure of an RF transformer for high-power switching harmonic oscillators, based on the connection of unit elements. An analysis of the obtained calculation results was carried out to select the most optimal structure for most practical applications. It has been established that using N single-turn transformers achieves an N-fold gain in the volume and weight of ferrite and copper tubing. However, this gain is not always convenient to implement due to design considerations. The presented calculations confirm that the proposed structure is the most optimal for most practical applications, as it is free from many drawbacks of the classical high-power transformer scheme with multi-turn windings.

Keywords: semiconductor electronics, transformer, switching HF generator, power supply, electronic components

Citation: Filin V.A., Sattarov Kh.A., Yurova V.A., Golovin A.N. Selection of the optimal structure of a transformer based on single-turn elements for high-power switching transistor harmonic oscillators. Computing, Telecommunications and Control, 2025, Vol. 18, No. 3, Pp. 102–110. DOI: 10.18721/JCSTCS.18309

Научная статья
DOI: https://doi.org/10.18721/JCSTCS.18309
УДК 621.375.026

ВЫБОР ОПТИМАЛЬНОЙ СТРУКТУРЫ ТРАНСФОРМАТОРА НА ОСНОВЕ ОДНОВИТКОВЫХ ЭЛЕМЕНТОВ ДЛЯ МОЩНЫХ КЛЮЧЕВЫХ ТРАНЗИСТОРНЫХ ГЕНЕРАТОРОВ ГАРМОНИЧЕСКИХ КОЛЕБАНИЙ

В.А. Филин[1], *Х.А. Саттаров*[2], *В.А. Юрова*[1,3] ✉, *А.Н. Головин*[1]

[1] Санкт-Петербургский государственный университет телекоммуникаций им. проф. М.А. Бонч-Бруевича, Санкт-Петербург, Российская Федерация;

[2] Ташкентский университет информационных технологий им. Мухаммада ал-Хоразмий, Самаркандский филиал, Самарканд, Узбекистан;

[3] Северо-Западный государственный медицинский университет им. И.И. Мечникова, Санкт-Петербург, Российская Федерация

✉ va-yurova@mail.ru

Аннотация. Для большинства устройств электронной техники используются источники питания и генераторы сигналов. Рационально построенный мощный ключевой высокочастотный (ВЧ) генератор содержит обычно один колебательный контур, трансформатор и заземленную нагрузку. В работе проведены обоснования и расчет предложенной структуры ВЧ трансформатора для мощных ключевых генераторов гармонических колебаний, основанной на соединении единичных элементов. Проведен анализ полученных результатов вычислений для выбора наиболее оптимальной для большинства практических применений структуры. Установлено, что при использовании N одновитковых трансформаторов достигается выигрыш в объеме и весе феррита и медной трубки в N раз. Однако этот выигрыш не всегда удобно реализовать из конструктивных соображений. Представленные расчеты подтверждают, что предложенная структура является наиболее оптимальной для большинства практических применений, поскольку лишена многих недостатков классической схемы мощного трансформатора с многовитковыми обмотками.

Ключевые слова: полупроводниковая электроника, трансформатор, ключевой генератор ВЧ, источник питания, электронные компоненты

Для цитирования: Filin V.A., Sattarov Kh.A., Yurova V.A., Golovin A.N. Selection of the optimal structure of a transformer based on single-turn elements for high-power switching transistor harmonic oscillators // Computing, Telecommunications and Control. 2025. Т. 18, № 3. С. 102–110. DOI: 10.18721/JCSTCS.18309

Introduction

A rationally designed high-power switching radio frequency (RF) generator usually contains one oscillatory circuit, a transformer and a grounded load. In the generator circuit, the transformer performs two main functions: it converts the voltage (current) to a value determined by the load and isolates the generator from the load by DC voltage or mains voltage (50 Hz). A transformer designed for a given power in the load at a high operating frequency f is an essential element of the circuit, largely determining the performance characteristics, weight-dimention parameters and cost of the generator. The significant value of the parasitic leakage inductance of a classic multi-turn RF transformer for many generator topologies is a factor limiting the operating frequency and energy parameters. For these reasons, improving the electrical characteristics of the transformer, reducing its weight and dimensions is of great practical importance [1–2].

The equivalent circuit of the industrial generator load in the form of an inductor can be represented by a parallel connection of the inductance L_l and the resistive conductance G_l (Fig. 1) with a quality factor $Q_1 = 1/(\omega L_l G_l)$.

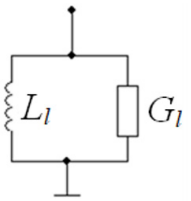


Fig. 1. Generator load equivalent circuit

Tuning the secondary winding circuit of a transformer to resonance allows compensating for the reactive component of conductivity, eliminating reactive currents in the secondary winding and, thus, unloading the transformer. However, in some cases, such a solution may prove impractical, for example, due to the lack of suitable capacitors with a given operating voltage and reactive power, inconveniences in their installation etc. In this case, the secondary winding of the transformer is loaded not on the resistive conductance G_l, but on the complex conductance $Y = G_l + 1/j\omega L_l$ (Fig. 1), i.e., when $G_l << 1/\omega L_l$ the load increases almost by Q_l times. Accordingly, the current and oscillation power in the secondary winding increase. The effect of the parasitic leakage inductance $L_{S.C.} = L_2(1 - k^2)$ also increases significantly. The volumes and weights of ferrite and copper will increase by Q_l times, and the geometric dimensions of the transformer – by $\sqrt[3]{Q_l}$ times.

Tuning the entire circuit to resonance is carried out using capacitors connected to the primary winding. For example, the weight of a transformer of an industrial generator with a load capacity of $P_l = 60$ kW in resonant mode at an operating frequency of $f = 66$ kHz is approximately 10–15 kg. When the load quality is $Q_l = 10$ and the capacitors are excluded, the weight of the transformer actually increases to 100 kg. In this case, as in other cases, when the transformer power is increased due to current while maintaining the voltage level or when increasing the operating frequency, difficulties may arise with the implementation of the transformer, since the estimated number of turns of any of the windings may be less than one. There are difficulties with the winding wire, the thickness of which may be too large, phenomena such as skin effect, displacement of current to the boundaries of the wires under the influence of a magnetic field etc. occur.

The goal of this work is an attempt to justify and calculate a new structure of an RF transformer for high-power switching generators of harmonic oscillations, based on the connection of single elements, i.e., single-turn transformers of a “cable” design. Such a structure, in the author’s opinion, is optimal for most practical applications, since it is free from many drawbacks of the classical high-power transformer scheme with multi-turn windings.

Single-turn transformers and their connections

When constructing high-power thyristor or transistor RF generators and secondary power sources, in order to simplify and reduce the cost of the design, transformerless primary (mains) power sources are usually used, giving 300, 500 or 600 V of direct voltage at the output. In this case, the increase in power is achieved solely by increasing the direct current, which entails a decrease in the inductance and the number of turns of the primary winding of the RF transformer. At the same time, the dimensions and thickness of the winding wire increase. Under otherwise equal conditions, the influence of the magnetic field on the distribution of current in the magnetic circuit and in the wire (skin effect, eddy currents) increases, which leads to increased losses of electrical energy in the transformer and, accordingly, to its heating.

Similar phenomena occur in the secondary winding of the transformer, especially when a low-resistance inductor consuming high current is connected to it. The skin effect and current displacement to the edges of the winding, caused by the magnetic field, increase as the transformer’s operating

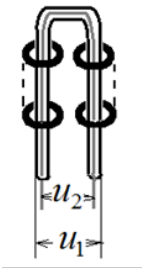


Fig. 2. Single-turn transformer type

frequency rises. If the dimensions of the transformer decrease proportionally to the decrease in the wavelength, then, due to the similarity principle, the relative influence of the skin effect does not change.

Problems with a small (less than one) number of turns and the skin effect can be solved to a large extent by replacing one high-power transformer with a combination of low-power ones, splitting the magnetic field into a number of unrelated magnetic fluxes, i.e., applying the litz wire principle to the magnetic circuit. At first glance, this will lead to a more complex design. However, this complication is insignificant or may even turn out to be a simplification, if single-turn transformers of extremely simple design are used. A single-turn transformer (Fig. 2) is a U-shaped copper tube onto which ferrite rings are threaded. This tube is the primary (secondary) winding, and the secondary (primary) winding is a litz wire bundle (or a second copper tube) passed inside the copper tube. This "cable" design of the windings ensures, among other things, a sharp decrease in the leakage inductance of the windings and obtaining their coupling coefficient k, amounting as several nines after the decimal point. An important advantage of single-turn transformers is the presence of only one turn in the primary and secondary windings with an air gap between them, ensuring good heat dissipation.

Series-parallel connections of the primary and secondary windings of transformers allow obtaining integer or fractional transformation coefficients. Of course, such a method of voltage transformation is acceptable if the transformation coefficient is not too large and lies within several units; which is typical for generators with inductors and switching sources of secondary power supply with a small transformation ratio. The winding terminals (copper tubes) can be used for mechanical mounting of transformers on strip lines serving as input or output buses.

Fig. 3 shows an example of a connection of four single-turn transformers – series-parallel for primary windings and parallel for secondary windings, and Fig. 4 shows a standard equivalent circuit for such transformer connections.

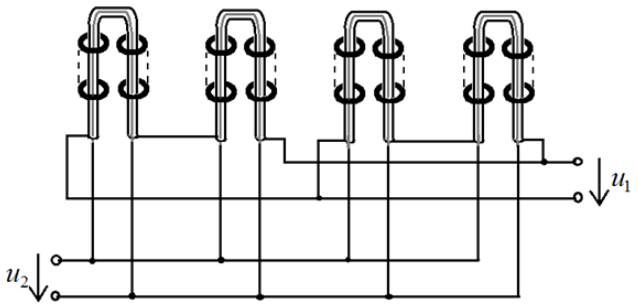


Fig. 3. Series-parallel connection of single-turn transformers

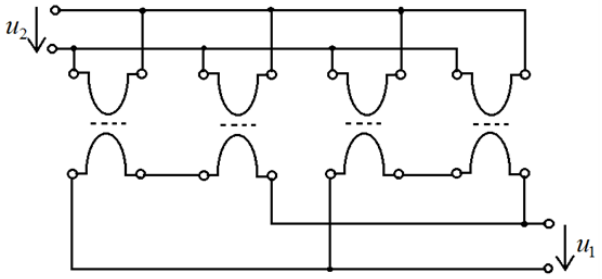


Fig. 4. Equivalent circuit of series-parallel connection

In addition to the very high coefficient of magnetic coupling of the windings they provide, the design methodology for single-turn transformers should be based on the requirement to obtain the necessary transformation coefficient and sufficiently small losses in the magnetic circuit and windings (turns). The transformation coefficient is ensured by selecting the required number of single-turn transformers connected in series-parallel.

For example, the circuit in Fig. 3 provides a transformation coefficient of $n = 0.5$ or $n = 2$, where two transformers can be used instead of four. If we take six transformers and connect three primary windings in series and threes in parallel, and the secondary windings with two in series and three twos in parallel, we can obtain a transformation ratio of 2:3 (or 3:2), etc. In general, for $m{:}l$ transformation, ml transformers are required. In some cases, this is not a significant limiting factor, since the transformers are extremely simple (Fig. 2) and, in addition, if there are power control units, there is no need to strictly maintain the specified (calculated) transformation coefficient.

Methodology for calculating elements of a single-turn transformer and completing a circuit based on it

Here is a method for the calculating the elements of a single-turn transformer. No-load losses in the magnetic circuit

$$P_{nl} = \hat{P}_{nl} V_m, \quad \hat{P}_{nl} = P_{m0} \left(f / f_{m0} \right)^{\alpha} \left(B_m / B_{m0} \right)^{\beta}, \tag{1a, b}$$

where $\hat{P}_{nl}$ is the specific losses per unit volume V_m of the magnetic circuit, $f_{m0} = 1$ kHz, $B_{m0} = 1$ T, $P_{m0} = \hat{P}_{nl}$ at $f = f_{m0}$, $B_m = B_{m0}$. The values P_{m0}, α, β are given in [3–6], for example, for ferrite 2000NMZ: $P_{m0} = 0.178$ W/cm^3; $\alpha = 1.3$; $\beta = 1.7$. To ensure acceptable values of P_{nl}, it is necessary to select a sufficiently small value of induction B_m. After selection B_m the cross-sectional area S_m of the magnetic circuit is determined for one turn ($w_1 = 1$) from the following formula:

$$S_m \left(m^2 \right) = {U_{m1} \left(\mathrm{V} \right)} \Big/ {2\pi f \left(\mathrm{kHz} \right) B_m \left(\mathrm{T} \right)}. \tag{2}$$

To ensure sufficiently small losses in a tubular loop, it is necessary that the diameter of the tube be large enough and the ohmic resistance $R_{\sim}$ at the operating frequency f be small enough. This resistance depends on the depth of the skin layer:

$$\lambda(m) = 1 \Big/ \sqrt{\pi f \mu_a / \rho}, \quad \lambda = 0.26 \text{ mm} \quad (\text{copper}, \ f = 66 \text{ kHz}), \tag{3a, b}$$

where $\mu_a = 4\pi \cdot 10^{-7}$ H/m is absolute magnetic constant, $\rho = 1.75 \cdot 10^{-8}$ Ohm·m is the specific resistance of copper.

For small diameter copper wire $d_0 \le \lambda$

$$R_{\sim} = K_{\text{ad}} \cdot R_{=}, \;\; R_{=} = \rho l_1 / S_1, \;\; K_{\text{ad}} \cong 1 + \left(w^2/15\right)\left(d_0/\lambda\right)^4, \qquad (4a, b)$$

where $R_{=}$ is DC resistance, l_1, S_1 are the length and the cross-sectional area of the wire in meters, K_{ad} is the coefficient of resistance increase $R_{=}$ due to the skin effect, w is the number of turns of the winding.

The thickness of the copper tube forming the turn, practically for design reasons (mechanical strength) at high frequencies is significantly greater than the skin layer equal to, for example, 0.26 mm (see formula 3b). In this case, the equivalent thickness of the tube, determining its conductivity, can be considered approximately equal to the skin layer depth λ in accordance with formulas (3a, b). We obtain the tube diameter d equal to

$$d\left(\text{mm}\right) = \frac{S\left(\text{mm}^2\right)}{\pi\lambda\left(\text{mm}\right)} = \frac{I_{m1}\left(A\right)}{\sigma\left(\dfrac{A}{\text{mm}^2}\right) \cdot \pi\lambda\left(\text{mm}\right)}, \qquad (5)$$

where the current density σ = (5–10)(A/mm^2) is taken to be several times greater than when calculating a conventional transformer, since the cooling of the tube turn in the presence of air gaps with ferrite and the second turn is significantly better than with a multi-turn winding. In addition, the mass of the tube is increased due to part of the thickness lying outside the skin layer. For a wire turn in the form of a bundle of litz wire placed inside the tube, formula (5) is also valid at σ = (1...2) A/mm^2.

The assembly of single-turn transformers is as follows. If we proceed from classical single-turn transformers (one turn in the primary and secondary windings), then with their help it is possible to realize transformation coefficients characterized by an integer number of times. That is 1:n (n = 1, 2, 3, 4, etc.), then n transformers are required for this (or 2n, as in Fig. 3). Other combinations are also possible, for example 2:3 (six transformers), 3:4 (twelve transformers) etc. By selecting a transformation ratio close to the required (calculated) one, the output power can be brought to the required value using the control unit (if available).

If these options are unacceptable, then a non-classical transformer can be used, in which only one winding (primary or secondary) is implemented as a single turn in the form of a copper tube, and the other winding (windings) is located inside this tube. For convenience of winding turns, the tube can be cut along its internal diameter. In this case, the current through the joint of the tube halves will be practically absent, since the lines of the electric field and the tube current run along it, and not in the transverse direction. In this case, U-shaped ferrites should be used, covering the copper tube in pairs.

This method provides great opportunities for varying the transformation coefficient and the number of windings. At the same time, the advantages of single-turn transformers are basically preserved – a very high winding coupling coefficient, the magnetic litz wire effect (due to the number of connected transformers) and the principle of mechanical fastening of the copper tube (turn) to the strip lines connecting to the generator and load.

Calculation example

In all cases, it is convenient to calculate a single-turn transformer based on a winding realized by a copper tube. As an example, we will calculate a set of transformers in a single-turn version for a power of P_l = 60 kW, a frequency of f = 66 kHz, an input voltage amplitude of U_{m1} = 500 V and a transformation coefficient of n = 0.5. Therefore, one of the two pairs of transformers shown in Fig. 4 can be used, with a series connection of the primary and parallel connection of the secondary windings. In this case, on any turn of each of the transformers, the voltage amplitude is U_{m1} = 250 В, the power is 0.5 P_l = 30 kW and current $I_{m1} = P_l/U_{m1}$ = 240 A.

The outer diameter of a copper tube implementing a turn at σ = 8 A/mm², λ =0.26 mm according to formula (5) will be equal to:

$$d = \frac{240A}{8\left(\frac{A}{\text{mm}^2}\right)\cdot\pi\cdot 0.26\ \text{mm}} = 34\ \text{mm}.$$

For reasons of mechanical strength of the transformer, the thickness of the tube wall can be taken equal to 1–2 mm (many times greater than λ = 0.26 mm) and the internal diameter is no less than 30 mm.

The cross-sectional area S and the diameter d_0 of the second turn – the litz wire bundle will be calculated according to (5) at σ = 1 A/mm² at

$$S = \frac{240A}{1\text{A}/\text{mm}^2} = 240\ \text{mm}^2, \quad d_0 = \sqrt{\frac{960\ \text{mm}^2}{\pi}} = 17.5.$$

A bundle with such a cross-section, taking into account the laying of internal cores and insulation, can be placed inside a tube with an internal diameter of 30 mm. Ferrite rings with external and internal diameters D_2 = 65 mm and D_1 = 40 mm are suitable for forming a magnetic circuit. The cross-sectional area of the magnetic circuit at B_m = 0.1 T according to expression (2)

$$S_{\text{m}}\left(\text{m}^2\right) = \frac{Um1\left(\text{V}\right)}{2\pi f\left(\text{Hz}\right)B_{\text{m}}\left(T\right)} = \frac{250\text{V}}{2\pi\cdot 66\cdot 10^3\,\text{Hz}\cdot 0.1\text{T}} = 6\cdot 10^{-3}\,\text{m}^2.$$

The height h of the ferrite cylinder (Fig. 1) will be equal to

$$h = \frac{S_{\text{m}}}{D_2 - D_1} = \frac{6\cdot 10^{-3}\,\text{m}^2}{25\cdot 10^{-3}\,\text{m}} = 0.24\ \text{m} = 24\ \text{cm}.$$

Ferrite volume of one transformer

$$V_{\text{m}} = \left(\frac{\pi D_2^2}{4} - \frac{\pi D_1^2}{4}\right)\cdot 2h = 990\ \text{cm}^3.$$

The volume of ferrite for the entire set (two transformers) will be 1980 cm², which, according to estimates, exceeds the optimal (minimum) volume, since in this case the optimality conditions are not met. This is the price for the design advantages of single-turn transformers.

The inductance of the winding (turn-tube) at w_1 = 1, μ = 2000 will be equal to

$$L_1 = \mu_{\text{a}}\mu\frac{S_{\text{m}}}{l_{\text{m}}} = 4\pi\cdot 10^{-7}\cdot 2\cdot 10^3\,\frac{6\cdot 10^{-3}}{0.5\left(65-40\right)\cdot 10^{-3}\pi} = 384\ \mu\text{H}$$

per set (2 transformers) $2L_1$ = 768 μH. We will get the calculated value of the resistance of the tube turn

$$R_1 = \rho\frac{l_1}{S_1} = \rho\frac{2\left(h+D_2\right)}{\pi d\lambda} = 1.75\cdot 10^{-8}\,\frac{0.24+0.065}{\pi\cdot 0.034\cdot 0.26\cdot 10^{-3}} = 1.2\cdot 10^{-3}\ \text{Ohm}.$$

Losses in a single transformer tube turn are

$$P_{s.c.1} = \frac{1}{2} I_{m1}^2 R_1 = \frac{1}{2} \cdot 240^2 \cdot 1.2 \cdot 10^{-3} = 34.5 \text{ W}.$$

These losses are about 0.1% of the power of one transformer (30 kW), therefore they are significant only from the point of view of the transformer quality. However, since the turn is made in the form of a sufficiently massive copper tube, open for cooling, the heating factor is also insignificant. This opens the possibility of making the second turn of the transformer not as a litz wire bundle, but as a second U-shaped copper tube located concentrically inside the first tube. In this case, the first tube should be cut longitudinally along the diameter into two U-shaped halves, inside which the second tube with an insulating gap should be placed. After placement, the joint of the U-shaped halves can be soldered, but there should be no significant currents in the joint, since the current lines run along the tube.

In the case under consideration, the transformation coefficient was 1:0.5, which made it possible to use only two single-turn transformers. For other transformation ratios, the number of transformers may be larger. It is important that with an increase in the number of single-turn transformers, their total volume and weight or losses decrease if the operating frequency is high enough and the wall thickness of the tube-turn is significantly greater than the skin layer. This is explained by the fact that with the skin layer thickness remaining unchanged, the equivalent conductive cross-section of the copper tube decreases proportionally to its diameter, while the volume and weight of copper (and ferrite) decreases proportionally to the square of the tube (ferrite ring) diameter.

Thus, when using N single-turn transformers, an N-fold gain in the volume and weight of the ferrite and copper tube is achieved. However, this gain is not always convenient to implement due to design considerations. For example, the calculated set of two transformers could be implemented according to the circuit shown in Fig. 4, i.e., from four transformers. In this case, the height of the transformers h would increase from 24 cm to 48 cm, the outer diameter D_2 of the ferrite rings would decrease from 6.5 cm to 3.25 cm, and the diameter of the copper tube d – from 3 cm to 1.5 cm. Such a ratio of h, D_2 and d is hardly convenient from a design standpoint, although it is fundamentally feasible.

Naturally, increasing the number of single-turn transformers (without changing their total power) makes sense only as long as the thickness of the copper tube-turn remains significantly greater than the thickness of the skin layer.

Conclusion

The paper presents the substantiation and calculation of the proposed structure of the RF transformer for high-power switching generators of harmonic oscillations [7–8], based on the connection of single elements, i.e., single-turn transformers of the “cable” design. The presented calculations confirm that such a structure is optimal for most practical applications, since it is devoid of many of the disadvantages of the classical scheme of a high-power transformer with multi-turn windings. The proposed structure is part of a project to create a device based on *UBPM*-1 [9–10] with the possibility of using it in solar *MPPT DC*/*DC* converters, solar inverters, uninterruptible power supplies etc.

REFERENCES

1. **Xin A., Zhang B., Zhu C., Huang L., Fang Z., Liu Y.** Design of a high-frequency irreversible electroporation bipolar pulse generator based on H-bridge inverter with an isolating transformer. *IEEE Transactions on Applied Superconductivity*, 2024, Vol. 34, No. 8, Art no. 9502605. DOI: 10.1109/TASC.2024.3441607

2. **Badrzadeh B., Gustavsen B.** High-frequency modeling and simulation of wind turbine transformer with doubly fed asynchronous generator. *IEEE Transactions on Power Delivery*, 2012, Vol. 27, No. 2, Pp. 746–756. DOI: 10.1109/TPWRD.2011.2181429

3. **Zlobin V.A., Andreev V.A., Zvorono Yu.S.** *Ferritovye materialy* [*Ferrite materials*]. Leningrad: Energiya, 1970.

4. **Korickij Yu.V., Pasynkov V.V., Tareeva B.M.** *Spravochnik po elektrotekhnicheskim materialam* [*Handbook of electrical engineering materials*]. Vol. 3. Leningrad: Energoatomizdat, 1988.

5. **Mihajlova M.M., Filippov V.V., Muslanov V.P.** *Magnitomyagkie ferrity dlya radioelektronnoj apparatury* [*Soft magnetic ferrites for electronic equipment*]. Moscow: Radio i svyaz', 1983.

6. **Kunevich A.V., Podol'skij A.V., Sidorov I.N.** *Ferrity, enciklopedicheskij spravochnik* [*Ferrites, encyclopedic reference*], in 5 vol. Vol. 1. St. Petersburg: LIK, 2004.

7. **Dmitrikov V.F., Shushpanov D.V.** On the issue of capacitively coupled circuits using to explain the mutual influence of wires. *Journal Radioengineering*, 2024, Vol. 88, No. 1, Pp. 15–30. DOI: 10.18127/j00338486-202401-03.

8. **Holmes D.G., Lipo T.A.** *Pulse width modulation for power converters: Principles and practice*. NY: Wiley-IEEE Press, 2003.

9. **Filin V.A., Golovin A.N., Smirnov V.S., Yurova V.A.** Circuit design solution and algorithm of digital control of processes for a universal bidirectional transistor module of switching mode voltage converter. *Journal Radioengineering*, 2024, Vol. 88, No. 6, Pp. 100–109. DOI: 10.18127/j00338486-202406-13

10. **Golovin A., Filin V.** Problems of construction of power stages of VLF transistors radio transmitters. *Actual Problems of Infotelecommunications in Science and Education*, 2023, Pp. 604–607.

INFORMATION ABOUT AUTHORS / СВЕДЕНИЯ ОБ АВТОРАХ

Vladimir A. Filin
Филин Владимир Алексеевич
E-mail: filin_vladimir@mail.ru

Khurshid A. Sattarov
Саттаров Хуршид Абдишукурович
E-mail: s.xurshid@tuit.ru
ORCID: https://orcid.org/0000-0001-6942-0618

Valentina A. Yurova
Юрова Валентина Александровна
E-mail: va-yurova@mail.ru
ORCID: https://orcid.org/0000-0001-6317-8007

Alexei N. Golovin
Головин Алексей Николаевич
E-mail: cathseugut@yandex.ru

Submitted: 30.04.2025; Approved: 14.07.2025; Accepted: 11.09.2025.
Поступила: 30.04.2025; Одобрена: 14.07.2025; Принята: 11.09.2025.

Computing, Telecommunication and Control, 2025, Vol. 18, No. 3, Pp. 111–122.
Информатика, телекоммуникации и управление. 2025. Том 18, № 3. С. 111–122.

Research article
DOI: https://doi.org/10.18721/JCSTCS.18310
UDC 621.372.8

TOPOLOGY AND DESIGN METHODOLOGY OF A QUADRATURE WAVEGUIDE POWER DIVIDER FOR AN ARBITRARY FREQUENCY RANGE

D.V. Lomsadze, N.G. Kolmakova, S.V. Volvenko ✉

Peter the Great St. Petersburg Polytechnic University,
St. Petersburg, Russian Federation

✉ volk@cee.spbstu.ru

Abstract. This article focuses on the modeling of a quadrature waveguide power divider, taking into account technological requirements and manufacturing feasibility. A relatively simple yet effective device topology is proposed, along with a method for calculating its geometric parameters to operate reliably within a predefined frequency range. The study includes a comprehensive analysis of existing power divider configurations, provides a detailed overview of the theoretical foundations for designing a quadrature waveguide-slot bridge and an in-depth examination of all design stages of the power divider, using the 27–35 GHz frequency range as an example. Based on the developed model, a working prototype was fabricated. Measurements of its frequency characteristics confirmed the validity and practical applicability of the proposed approach. The design methodology developed in this work can be effectively applied to create a quadrature waveguide-slot bridge in any given frequency range, making it a versatile tool for engineers and researchers in the field of microwave technologies.

Keywords: quadrature power divider, waveguide-slot bridge, rectangular waveguide, calculation method, device topology

Citation: Lomsadze D.V., Kolmakova N.G., Volvenko S.V. Topology and design methodology of a quadrature waveguide power divider for an arbitrary frequency range. Computing, Telecommunications and Control, 2025, Vol. 18, No. 3, Pp. 111–122. DOI: 10.18721/JCSTCS.18310

Научная статья
DOI: https://doi.org/10.18721/JCSTCS.18310
УДК 621.372.8

ТОПОЛОГИЯ И МЕТОДИКА ПРОЕКТИРОВАНИЯ КВАДРАТУРНОГО ВОЛНОВОДНОГО ДЕЛИТЕЛЯ МОЩНОСТИ ДЛЯ ПРОИЗВОЛЬНОГО ДИАПАЗОНА ЧАСТОТ

Д.В. Ломсадзе, Н.Г. Колмакова, С.В. Волвенко ✉

Санкт-Петербургский политехнический университет Петра Великого,
Санкт-Петербург, Российская Федерация

✉ volk@cee.spbstu.ru

Аннотация. Данная статья посвящена моделированию квадратурного волноводного делителя мощности с учетом технологических требований и возможностей его изготовления. Предложена довольно простая, но эффективная топология устройства, а также методика расчета его геометрических параметров для обеспечения надежной работы в предварительно заданном частотном диапазоне. В работе представлен комплексный анализ существующих конфигураций делителей мощности, дан подробный обзор теоретических основ проектирования квадратурного волноводно-щелевого моста, а также детально рассмотрены все этапы проектирования делителя мощности на примере частотного диапазона 27–35 ГГц. На основе разработанной методики получена модель, благодаря которой был изготовлен рабочий опытный образец квадратурного волноводно-щелевого моста. Результаты измерения его частотных характеристик подтвердили корректность и обоснованность предложенного подхода, а также его практическую применимость. Разработанная в рамках данной работы методика проектирования может быть эффективно применена для создания квадратурного волноводно-щелевого моста в любом заданном диапазоне частот, что делает ее универсальным инструментом для инженеров и исследователей в области СВЧ-технологий.

Ключевые слова: квадратурный делитель мощности, волноводно-щелевой мост, прямоугольный волновод, методика расчета, топология устройства

Для цитирования: Lomsadze D.V., Kolmakova N.G., Volvenko S.V. Topology and design methodology of a quadrature waveguide power divider for an arbitrary frequency range // Computing, Telecommunications and Control. 2025. Т. 18, № 3. С. 111–122. DOI: 10.18721/JCSTCS.18310

Introduction

Power dividers (PD) are essential components of modern microwave technology. They are widely used in power amplifiers [1], antenna arrays [2] and other transceivers [3]. Advances in technology and production capabilities, along with the demand for wider communication channels and the increasing number of electronic devices, have driven interest in PDs operating at frequencies above 10 GHz. This trend requires efficient broadband signal processing and transmission, which can be achieved using various types of transmission lines. These include fiber-optic links [4], dielectric waveguides based on low-density polyethylene [5] and traditional hollow metallic waveguides. Each of these technologies offers distinct advantages in terms of size, weight, loss characteristics and ease of integration. However, hollow metallic waveguides possess the unique ability to handle high microwave power levels without significant losses or electrical breakdown, making them the primary choice for high-power applications, particularly in radar and satellite systems. Another option is to use microstrip and other planar structures; however, they suffer from increased dielectric and conductor losses at high frequencies [6], which

further reinforces the need for low-loss, high-power-capable waveguide solutions[1]. The most common PDs are those that split the input power equally between two output channels. These are T-, E- and Y-junctions [7, 8], double waveguide tees or magic tee [9], waveguide-slot bridges [10] and others. Depending on the type of junction, the output ports may be in-phase, out-of-phase or exhibit a 90° phase difference. Cascading of such PD [11] makes it possible to design a divider with a large number of output channels. Quadrature PDs stand out among these structures, with quadrature waveguide-slot bridges (QWSB) being one of their representatives. QWSB are four-channel devices that provide an equal power division between two output channels with a 90° phase shift. Their key advantage is the ability to terminate the remaining output with a load, ensuring isolation between the two power-divided output channels. The optimal behavior of QWSB is achieved either by selecting a special complex geometry of the matching area [10] or by using various matching elements such as inductive posts[2] and resistive baffles [12], which further complicate the configuration of the device.

The literature describes several techniques for designing waveguide-slot bridges by selecting a special geometry. For example, [13] considers a step transition as the matching section. However, this approach is highly sensitive to manufacturing tolerances and complicates fabrication, particularly at high frequencies. A solution to this issue was proposed in [10], where the step transition in the matching region is replaced with a smooth bend, reducing manufacturing complexity. However, the authors of this article do not pay enough attention to explaining the synthesis of the geometry of the PD.

To fill this gap, this paper provides a detailed method for designing a quadrature PD within a predefined frequency range without using complex geometries or extra elements. The capabilities of the proposed topology and its design methodology are demonstrated through the development of a broadband quadrature PD in the 27–35 GHz frequency range.

Concept and design of a quadrature PD for a predefined frequency range

This paper presents a design technique for an H-plane QWSB. It demonstrates how power division can be achieved within a given frequency range by selecting appropriate geometric parameters of the proposed topology. This approach is applicable to various frequency ranges and is limited only by the allowable range of changes in the geometric parameters of the structure. Fig. 1 shows the geometry of the PD being developed along with the key geometric parameters.

To design a quadrature PD in the frequency range (f_0, f_1) the following steps are necessary:

1. Select the appropriate cross-sectional dimensions of the rectangular waveguides a and b.
2. Use a simplified theoretical model to estimate the initial approximations of the main parameters of the splitting region d and L.
3. Based on the technological requirements and manufacturing capabilities, determine the minimum allowable value of e.
4. Set initial approximations for the parameters s and n.
5. Use an optimization process to determine the geometric parameters that provide the best matching and require division within the frequency band.

Let us look at these steps in more detail.

1. First, the dimensions of the rectangular waveguide must be selected such that the specified frequency range (f_0, f_1) lies within the operating range of this waveguide. From a practical point of view, rectangular waveguides of standard dimensions are the most convenient. However, a choice of non-standard waveguide sizes is also possible.

2. To estimate the initial approximations of parameters d and L, consider an ideal model of the waveguide-slot bridge[3], which consists of two waveguides connected along a narrow wall (shaded area

[1] Murav'ev V.V., Korenevskii S.A., Mishchenko V.N. Sverkhvysokochastotnye tekhnologii v sistemakh telekommunikatsii [Ultra-high-frequency technologies in telecommunication systems]. Minsk: BGUIR, 2007.
[2] Vol'man V.I., Pimenov Iu.V. Tekhnicheskaia elektrodinamika [Technical electrodynamics]. Moscow: Sviaz', 1971.
[3] Vol'man V.I., Pimenov Iu.V. Tekhnicheskaia elektrodinamika [Technical electrodynamics]. Moscow: Sviaz', 1971.

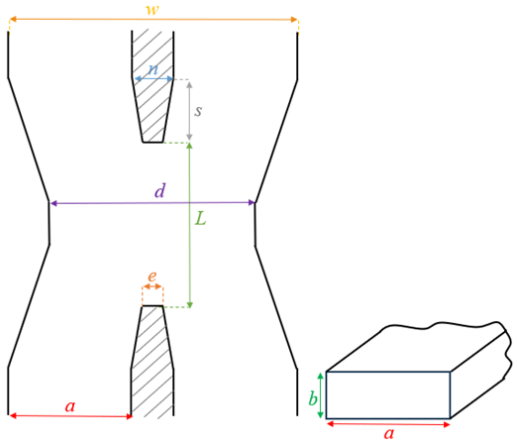


Fig. 1. Geometry of the H-plane PD with key parameters

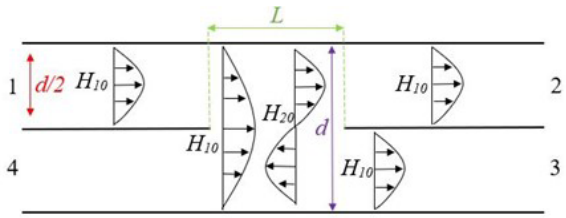


Fig. 2. The principle of operation of the waveguide-slot bridge

in Fig. 1). In this case, the width of the feeding waveguides is half the width of the splitting region, i.e., $d/2$. Channel 1 is excited by the H_{10} mode, while channels 2 and 3 are output channels where H_{10} modes with half amplitude and 90° phase shift propagate, and channel 4 has to be isolated. The absence of a propagating H_{10} mode in channel 4 is possible as a result of the combination of two H_{10} modes of equal amplitude and a phase difference of 180°, while all feeding waveguides must operate in the single-mode range. Thus, the problem can be considered as the sum of two cases: in the first case ports 1 and 4 are simultaneously excited in-phase, and in the second case they are excited in antiphase.

Assume that only H_{10} and H_{20} modes of the rectangular waveguide with width d can propagate in the splitting region, that is $f_1 < f_{H30}$, where f_{H30} is the cutoff frequency for H_{30} mode of the rectangular waveguide with width d. Since the feeding waveguides are single-mode, we can deduct from the definition of the cutoff frequency that

$$\frac{c}{f_0} < d < \frac{3c}{2f_1}, \tag{1}$$

where c is the speed of light in a vacuum.

Naturally, the inequality is true if $2f_1 < 3f_0$. Otherwise, it is necessary to focus only on the right side of the inequality and choose d close to $\frac{3c}{2f_1}$, ensuring the operation of the device in the low-frequency range during subsequent modeling steps.

Consider the problem of in-phase excitation of channels 1 and 4. In this case, two incident H_{10} modes of the waveguide with width $d/2$ combine at the slot to form one H_{10} mode of the waveguide with width d. As this mode propagates along the length L, it acquires a phase shift equal to

$$\varphi_1(\lambda) = \frac{2\pi L}{\lambda}\sqrt{1-\left(\frac{\lambda}{2d}\right)^2},$$

where λ is the wavelength in the waveguide.

Next, the resulting mode splits into H_{10} modes in channels 2 and 3. In the case of antiphase excitation of ports 1 and 4, the incident field forms a propagating H_{20} mode, which receives the phase shift equal to

$$\varphi_2(\lambda) = \frac{2\pi L}{\lambda}\sqrt{1-\left(\frac{\lambda}{d}\right)^2}.$$

Subsequently, it splits back into the main modes in channels 2 and 3. For the structure to exhibit quadrature bridge properties, it is necessary that[4] $\varphi_1(\lambda) - \varphi_2(\lambda) = \frac{\pi}{2}$, then

$$L = \frac{\lambda}{4}\frac{1}{\sqrt{1-\left(\frac{\lambda}{2d}\right)^2} - \sqrt{1-\left(\frac{\lambda}{d}\right)^2}}. \quad (2)$$

Thus, by selecting d to satisfy inequality (1) and using formula (2) to determine the value of L for the chosen d, we obtain initial approximations of these geometric parameters for subsequent modeling.

3. In the considered theoretical model, the distance between the waveguides was assumed to be 0, which is impossible from a practical point of view. At this stage of the design, it is necessary to account for the effect of a metal septum between the feeding waveguides (Fig. 1) and determine its minimum allowable width, i.e., the value of parameter e, based on the manufacturing capabilities.

4. In the previous steps 2 and 3, the widths of the feeding waveguides did not match the width a, defined in step 1. The goal of this stage is to integrate the resulting division area with the waveguides of the desired size. To achieve this, it is necessary to determine the initial approximations for the parameters n and s. As shown in Fig. 1, the region described by these parameters is essentially a smooth transition between the rectangular waveguide and the splitting region. The choice of these parameters will be determined not only by the amplitude-frequency characteristics, but also by the further manufacturing process for each specific product, as well as the requirements for the necessary structural strength. For example, the width n should not be too smaller than the diameter of the milling cutter used to fabricate the structure, otherwise the septum may break during manufacture or assembly.

5. At the final stage of modeling, the optimization process should be used to determine the values of the specified geometric parameters. It is convenient to use the following set as an optimization criterion:

$$S_{11},\ S_{41} < \beta;\quad -3dB - \alpha < S_{21},\ S_{31} < -3dB + \alpha,$$

where β is determined based on the requirement for matching the structure and isolation of port 4, while α defines the proximity to equal power division between ports 2 and 3. S_{11} is the amplitude of the reflection coefficient of the main mode in channel 1, and S_{21}, S_{31} and S_{41} are the amplitude of the transmission coefficients of the main mode into ports 2, 3 or 4 when port 1 is excited by the main mode.

It should be noted that the topology in Fig. 1 contains irregularities in the form of corners. Naturally, these irregularities will affect its electromagnetic characteristics. Moreover, in practice, they will be

[4] Vol'man V.I., Pimenov Iu.V. Tekhnicheskaia elektrodinamika [Technical electrodynamics]. Moscow: Sviaz', 1971.

rounded off. Therefore, the corner rounding radii must also be included as optimization parameters. The limiting values of these radii will be determined by the further technological process such as the diameters of the milling cutters used.

Application of the proposed topology and design methodology

The proposed approach to PD design is illustrated using the development of a PD for the 27–35 GHz frequency range.

1. To design a waveguide-slot bridge in the frequency range of 27–35 GHz, a standard WR28 rectangular waveguide with geometric dimensions of 7.11×3.56 mm^2 is used.

2. Using formulas (1) and (2), we obtain that 11.11 mm < d < 12.85 mm. Let us choose d = 11.2 mm, then the L values will be in the range from 9 to 12 mm.

Fig. 3 shows the reflection coefficient for the theoretical model (Fig. 2) at different values of L and the selected d. When choosing initial approximations of geometric parameters, the reflection coefficient S_{11} can be prioritized, as its behavior is generally used to assess the operating band of the device.

As shown in Fig. 3, the best matching from the point of view of the frequency range is achieved at L equal to 10 or 11 mm. Fig. 4 shows the amplitude-frequency characteristics of the transmission coefficients at L = 11 mm.

3. Having set the initial approximations for the parameters d = 11.2 mm and L = 11 m, let's consider how the amplitude-frequency response of the QWSB changes when a non-zero distance appears between the feeding waveguides. With the width of the waveguides fixed at $d/2$, we change e (Fig. 1). In this case, the width of the splitting region increases by the value of e. As seen in Fig. 5, the reflection

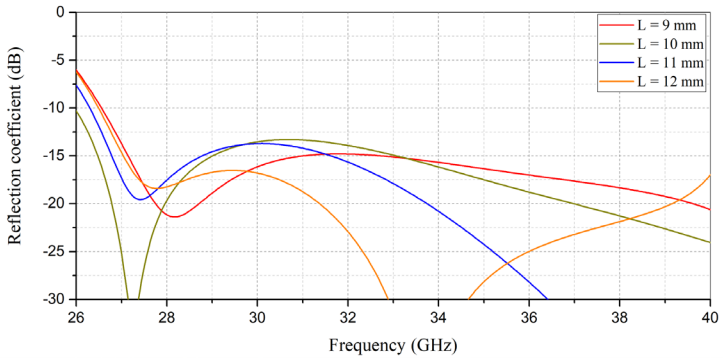


Fig. 3. Reflection coefficient versus frequency with varying L

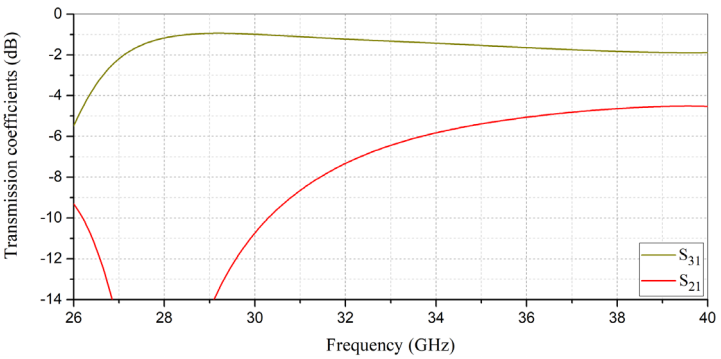


Fig. 4. Transmission coefficients versus frequency at L = 11 mm

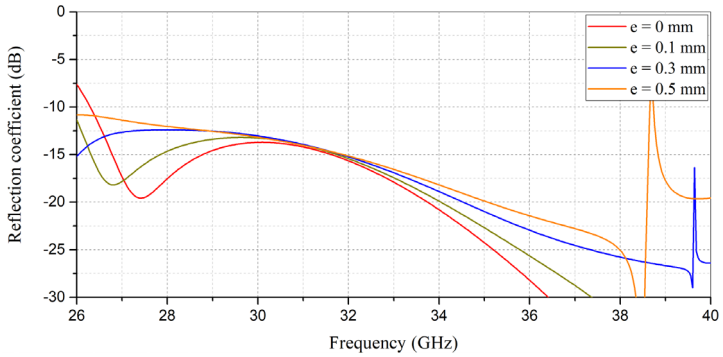


Fig. 5. Reflection coefficient versus frequency with varying e

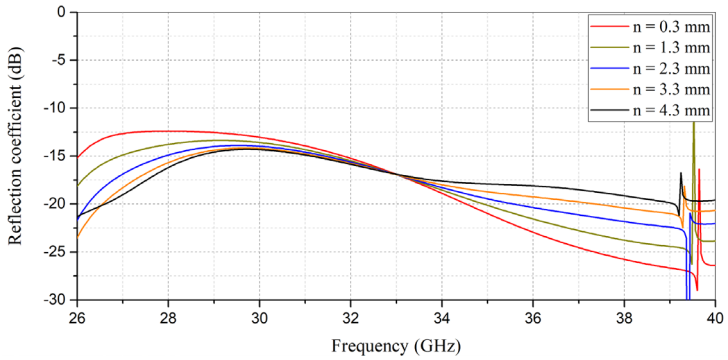


Fig. 6. Reflection coefficient versus frequency with varying n at e = 0.3 mm and s = 10 mm

coefficient S_{11} deteriorates significantly even with a small width of the septum that occurs between the feeding channels and the operating band of the PD shifts to the low-frequency region. This problem can be solved by smoothly increasing the distance between the channels. In this case, the parameter e can be selected relatively small, but not zero. Therefore, for further PD modeling, the value e = 0.3 mm was chosen, at which the obtained reflection coefficient lies in a desired frequency band.

4. Next, we will set the initial approximations for the parameters n and s, which determine smooth transitions between the projected division area and regular WR28 waveguides. Let s = 10 mm, which corresponds to the wavelength of the center of the studied range. Fig. 6 shows that as the parameter n increases, the matching improves and the working bandwidth expands. It should be borne in mind that further increases in n will result in larger weight and size parameters. Therefore, at this stage, we will fix n = 3.3 mm. Fig. 7 illustrates the effect of the transition length on the frequency response of the QWSB. On the one hand, an increase in s reduces the size of the QWSB, while on the other hand, the QWSB model becomes more consistent in the high-frequency range. For further design, the value s = = 5.5 mm was chosen.

The amplitude-frequency characteristics for the selected initial geometric parameters are shown in Fig. 8.

5. Having determined the initial parameters for the proposed topology, the final stage involved searching for optimal parameters. As a result of the optimization process in Ansys HFSS, taking into account the rounding of the corners of the topology, the quadrature PD configuration was obtained that ensures matching S_{11} at least –22 dB with a difference in the amplitudes of the transmission coefficients of the output ports S_{21} and S_{31} not exceeding 0.5 dB and the transmission coefficient S_{41} not exceeding

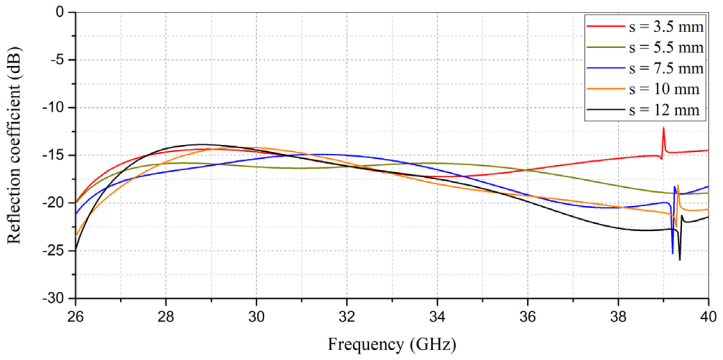


Fig. 7. Reflection coefficient versus frequency with varying *s*

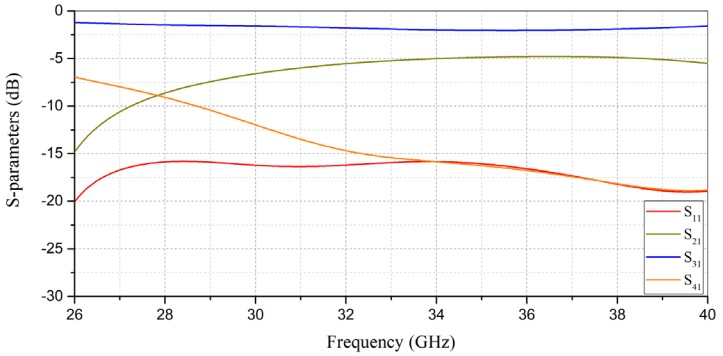


Fig. 8. Amplitude-frequency characteristics for the chosen initial parameters

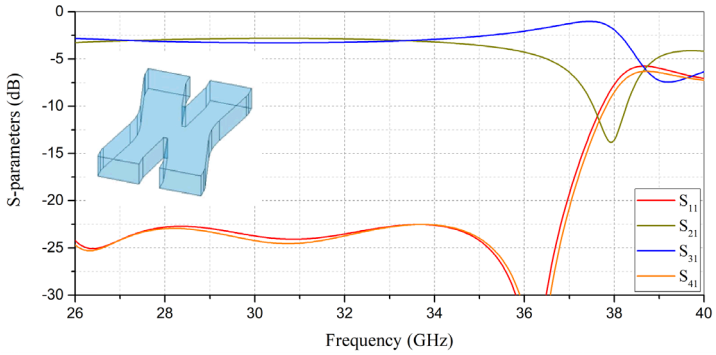


Fig. 9. Amplitude-frequency characteristics at optimal values of the main parameters

–22 dB in a given frequency range. According to the requirements for the quadrature PD, the phase difference between the output ports is 90°. The frequency characteristics of the developed model are shown in Figs. 9 and 10.

The optimal values of the main geometric parameters are shown in Table.

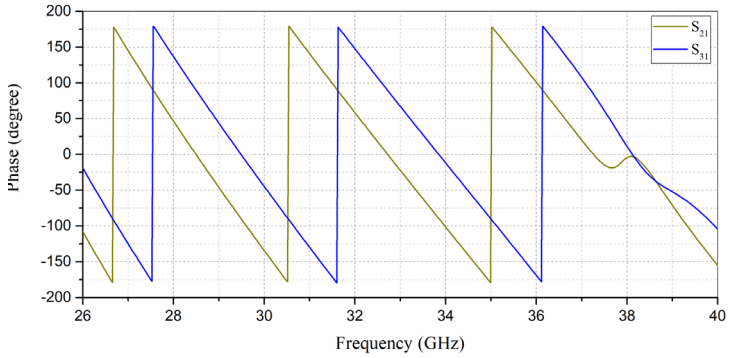


Fig. 10. Dependence of the phases of the output channels on the frequency

Table

Optimal values of the main parameters of the QWSB model

Parameter	Optimal value, mm
d	12.8
e	0.6
L	9.3
n	3.4
s	6.5

Experimental verification of the proposed approach

Based on the obtained model, an experimental sample of a quadrature PD was fabricated (Fig. 11). A matched load was installed in port 4 of the PD. A vector network analyzer was used to measure the frequency characteristics. The analyzer's measurement cables and the waveguide ports of the quadrature bridge were connected via pre-calibrated coaxial waveguide junctions. Each of the output ports was examined separately (Fig. 12).

A comparison of the experimental and calculated curves in Fig. 13 shows a high degree of correspondence in the required frequency range of 27–35 GHz. The shapes of the curves for the reflection coefficient (Fig. 13, *a*) and the transmission coefficients (Fig. 13, *b*) demonstrate similar characteristics and quantitative values, which confirms the accuracy of the simulation. Although the experimental

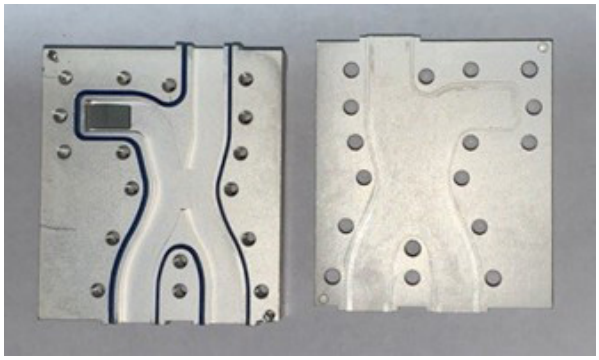


Fig. 11. Experimental sample of a quadrature PD

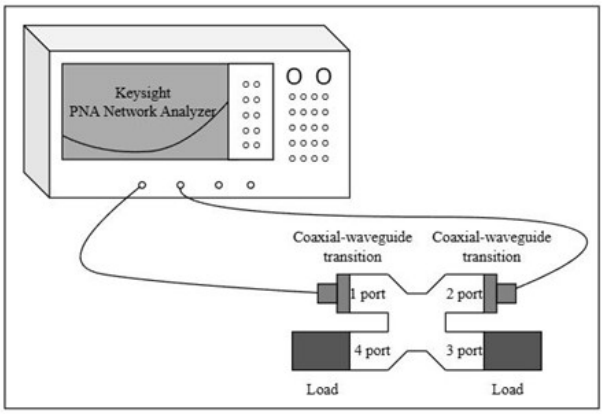


Fig. 12. Experimental setup for measuring a quadrature PD

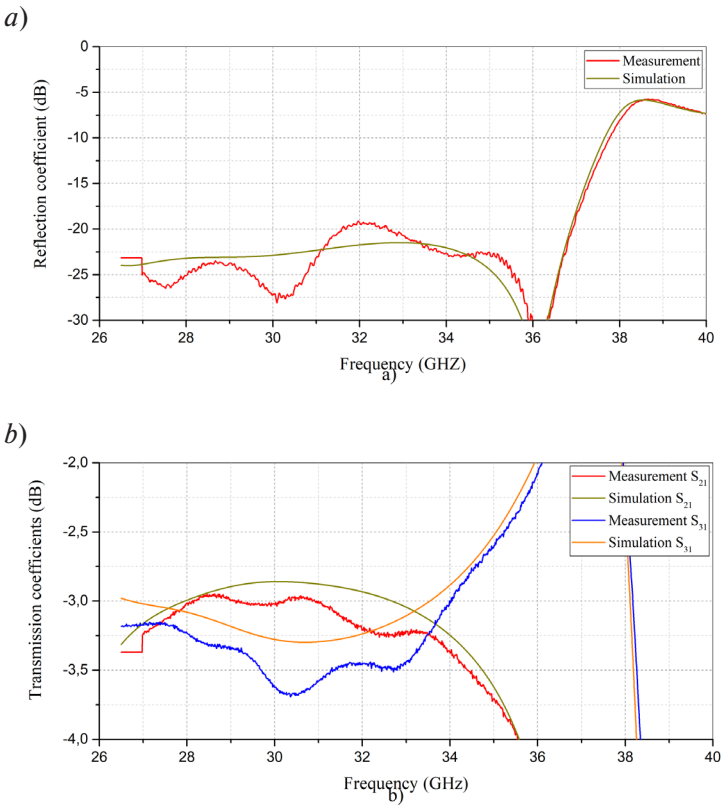


Fig. 13. Dependences of calculated and measured S_{11} on frequency (*a*); dependences of calculated and measured S_{21} and S_{31} on frequency (*b*)

curves have a more resonant character, this is explained by the measurement methodology and the use of matched loads in ports. Outside the division range, there is an almost complete coincidence of the curves, which indicates the reliability of the computational model in a wider frequency range. It should be noted that a discrepancy between the measured and simulated results can be also attributed to the roughness of the inner surfaces of the quadrature PD. Polishing these surfaces reduces losses, which would bring the experimental curves even closer to the calculated ones, as also discussed in [14, 15].

Conclusion

The topology of a quadrature waveguide PD and its design methodology have been developed, which makes it possible to create a PD for a predefined frequency range. The proposed approach is demonstrated by modeling an H-plane PD for operation in the frequency range of 27–35 GHz. The obtained model was verified by manufacturing a prototype and measuring its frequency characteristics. The measurement results confirmed the agreement between the calculated data and the real characteristics, thereby validating the proposed approach.

REFERENCES

1. **Sakata S., Komatsuzaki Y., Shinjo S.** Adaptive input-power distribution in Doherty power amplifier using modified Wilkinson power divider. *2020 IEEE Topical Conference on RF/Microwave Power Amplifiers for Radio and Wireless Applications* (*PAWR*), 2020, Pp. 34–37. DOI: 10.1109/PAWR46754.2020.9036005

2. **Romodin V.B., Kulik V.S.** Microstrip divider for compactsynthetic aperture radar antenna array. *Interekspo Geo-Sibir'*, 2017.

3. **Zhai G., Shi B.** Compact low loss millimeter wave 8-way radial waveguide power combiner. *TENCON 2017 – 2017 IEEE Region 10 Conference*, 2017, Pp. 1598–1601. DOI: 10.1109/TENCON.2017.8228112

4. **Abramson O., Gabbay D., Sharon A.** Applications of RF-over-fiber technology in the defense and civilian markets. *Microwave Journal, June Supplement*, 2025, P. 6.

5. **Park S.** Beyond copper and optical, a new interconnect eyes next gen data centers. *All About Circuits*, 2025.

6. **Mozharovskiy A.V., Soykin O.V., Artemenko A.A., Maslennikov R.O., Vendik I.B.** Wideband waveguide-to-microstrip transition for mm-wave applications. *Journal of the Russian Universities. Radioelectronics*, 2019, Vol. 22, No. 5, Pp. 17–32. DOI: 10.32603/1993-8985-2019-22-5-17-32

7. **Zhang L., Liu T.** A new H-plane T-junction waveguide power divider covering the full W-band. *2020 IEEE 3rd International Conference on Electronic Information and Communication Technology* (*ICEICT*), 2020, Pp. 803–805. DOI: 10.1109/ICEICT51264.2020.9334336

8. **Zhao P., Wang Q., Zhang F., He X.** A novel T-junction waveguide power divider with anti-phases and broad bandwidth. *2017 IEEE International Symposium on Antennas and Propagation & USNC/URSI National Radio Science Meeting*, 2017, Pp. 741–742. DOI: 10.1109/APUSNCURSINRSM.2017.8072413

9. **Deng J., Wang Q., Zhao P., Tian M., Li Q.** A quasi-planar H-Plane waveguide power divider with full bandwidth. *IEEE Microwave and Wireless Components Letters*, 2018, Vol. 28, No. 8, Pp. 645–647. DOI: 10.1109/LMWC.2018.2847028

10. **Zheng P., Zhou P., Yu W.-H., Sun H.-J., Lv X., Deng H.** W-band power divider based on H-plane slot waveguide bridge. *2012 International Conference on Microwave and Millimeter Wave Technology* (*ICMMT*), 2012, Pp. 1–4. DOI: 10.1109/ICMMT.2012.6230025

11. **Kang Y., Lin X.Q., Fan Y.** A THz one-to-four power divider with high isolation and arbitrary phase shift. *2020 International Conference on Microwave and Millimeter Wave Technology* (*ICMMT*), 2020, Pp. 1–3. DOI: 10.1109/ICMMT49418.2020.9386897

12. **Epp L.W., Hoppe D.J., Khan A.R., Stride S.L.** A high-power *Ka*-Band (31–36 GHz) solid-state amplifier based on low-loss corporate waveguide combining. *IEEE Transactions on Microwave Theory and Techniques*, 2008, Vol. 56, No. 8, Pp. 1899–1908. DOI: 10.1109/TMTT.2008.927299

13. **Hildebrand L.T.** Results for a simple compact narrow-wall directional coupler. *IEEE Microwave and Guided Wave Letters*, 2000, Vol. 10, No. 6, Pp. 231–232. DOI: 10.1109/75.852425

14. **Wang R., Cui W.** A rapid estimation of the conductor loss in the rectangular waveguide with rough surface. *2011 4th IEEE International Symposium on Microwave, Antenna, Propagation and EMC Technologies for Wireless Communications*, 2011, Pp. 498–500. DOI: 10.1109/MAPE.2011.6156144

15. **Hofmann A., Lomakin K., Kleinlein M., Bader T., Sippel M., Gold G.** SLS-printed E-band waveguides and the impact of surface roughness. *2023 53rd European Microwave Conference* (*EuMC*), 2023, Pp. 243–246. DOI: 10.23919/EuMC58039.2023.10290537

INFORMATION ABOUT AUTHORS / СВЕДЕНИЯ ОБ АВТОРАХ

Dzhuletta V. Lomsadze
Ломсадзе Джульетта Васильевна
E-mail: dlomsadze31@yandex.ru

Natalia G. Kolmakova
Колмакова Наталья Геннадьевна
E-mail: kolmakova.nataliya@gmail.com
ORCID: https://orcid.org/0000-0003-2595-4903

Sergey V. Volvenko
Волвенко Сергей Валентинович
E-mail: volk@cee.spbstu.ru
ORCID: https://orcid.org/0000-0001-7726-8492

Submitted: 05.06.2025; Approved: 11.09.2025; Accepted: 22.09.2025.
Поступила: 05.06.2025; Одобрена: 11.09.2025; Принята: 22.09.2025.

Computing, Telecommunication and Control, 2025, Vol. 18, No. 3, Pp. 123–130.
Информатика, телекоммуникации и управление. 2025. Том 18, № 3. С. 123–130.

Short message
DOI: https://doi.org/10.18721/JCSTCS.18311
UDC 621.3.049.774.2

AN 8-BIT WIDE INPUT SWING ANALOG-TO-DIGITAL CONVERTER BASED ON VOLTAGE-CONTROLLED OSCILLATOR

V.D. Bystrov ✉ , *D.V. Morozov* , *M.M. Pilipko*

Peter the Great St. Petersburg Polytechnic University,
St. Petersburg, Russian Federation

✉ bystrov.vd@edu.spbstu.ru

Abstract. An 8-bit analog-to-digital converter based on a voltage-controlled oscillator using 180 nm CMOS technology from Mikron JSC with a supply voltage of 3.3 V for the analog part and 1.8 V for the digital part is presented. The analog-to-digital converter has a wide range of input voltages from 0 to 3.3 V. The transistor-level simulation of the analog-to-digital converter in the time domain was performed in Cadence Virtuoso. The sampling rate was set to 1 MHz. The power consumption is about 1.8 mW. The dimensions of the designed layout are 103 μm by 109 μm. With an input frequency of 50 kHz and an amplitude of 1.55 V, the post-layout simulation shows an output SNDR of 36.48 dB (ENOB is 5.77 bits).

Keywords: analog-to-digital converter, voltage-controlled oscillator, linearity of the transfer characteristic, digital synthesis

Acknowledgements: The production of the integrated microcircuit was carried out at the expense of the Ministry of Education and Science of Russia within the framework of the federal project “Training of personnel and scientific foundation for the electronic industry” under the state assignment for the implementation of research work “Development of a methodology for prototyping an electronic component base in domestic microelectronic production based on the MPW service”.

Citation: Bystrov V.D., Morozov D.V., Pilipko M.M. An 8-bit wide input swing analog-to-digital converter based on voltage-controlled oscillator. Computing, Telecommunications and Control, 2025, Vol. 18, No. 3, Pp. 123–130. DOI: 10.18721/JCSTCS.18311

Краткое сообщение
DOI: https://doi.org/10.18721/JCSTCS.18311
УДК 621.3.049.774.2

ВОСЬМИРАЗРЯДНЫЙ АЦП С РАСШИРЕННЫМ ДИАПАЗОНОМ ВХОДНОГО СИГНАЛА НА ОСНОВЕ ГУН

В.Д. Быстров ✉ , *Д.В. Морозов* , *М.М. Пилипко*

Санкт-Петербургский политехнический университет Петра Великого,
Санкт-Петербург, Российская Федерация

✉ bystrov.vd@edu.spbstu.ru

Аннотация. Представлен восьмиразрядный аналого-цифровой преобразователь на основе генератора, управляемого напряжением, разработанный по 180 нм КМОП-технологии от АО «Микрон» с напряжением питания 3,3 В в аналоговой части и 1,8 В в цифровой части. Аналого-цифровой преобразователь имеет расширенный диапазон напряжений входного сигнала от 0 до 3,3 В. Моделирование преобразователя на транзисторном уровне во временной области проводилось в Cadence Virtuoso. Частота дискретизации была установлена равной 1 МГц. Потребляемая мощность составила около 1,8 мВт. Размеры разработанной топологии кристалла преобразователя составили 103 мкм на 109 мкм. При частоте входного сигнала, равной 50 кГц, и амплитуде 1,55 В, моделирование с учетом экстракции паразитных параметров топологии аналого-цифрового преобразователя показывает на выходе значение отношения сигнала к шуму и искажениям 36,48 дБ (эффективное число бит равно 5,77 бит).

Ключевые слова: аналого-цифровой преобразователь, генератор, управляемый напряжением, линейность передаточной характеристики, цифровой синтез

Финансирование: Производство интегральной микросхемы было выполнено за счет средств Министерства науки и высшего образования Российской Федерации в рамках федерального проекта «Подготовка кадров и научный фундамент для электронной промышленности» по государственному заданию на выполнение опытно-конструкторской работы «Разработка методики прототипирования элементной базы электроники на отечественных микроэлектронных производствах на основе сервиса MPW».

Для цитирования: Bystrov V.D., Morozov D.V., Pilipko M.M. An 8-bit wide input swing analog-to-digital converter based on voltage-controlled oscillator // Computing, Telecommunications and Control. 2025. Т. 18, № 3. С. 123–130. DOI: 10.18721/JCSTCS.18311

Introduction

The trend towards an increase in the number of sensors in areas such as the IoT, medical care and mechanical engineering has led to an increased need to use many analog-to-digital converters (ADCs). Since many devices are autonomous, ADCs must provide low power consumption [1–4]. Also, with the scaling of CMOS technology, there is a tendency to transfer the signal processing complexity to the digital domain [5]. The mentioned trends are supported by an ADC based on a volt age-controlled oscillator (VCO). The VCO-based ADC block diagram is shown in Fig. 1 (DU stands for digital unit). The analog signal is sent to the VCO input. A VCO converts the input signal level into the frequency of the output pulses. The digital unit counts the number of pulses per sampling cycle and generates the corresponding binary code at the output. The typical values of the bit depth of the ADC type in question depend on the technology used. For technologies larger than 100 nm, bit depths of up to 10 bits are more common [6, 7]. In cases of implementation using technology less than 100 nm, a bit depth of 15 bits is often found [8–10].

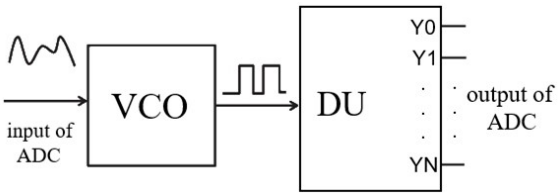


Fig. 1. Block diagram of an ADC based on a VCO

Circuit of the VCO

Fig. 2 shows the developed circuit of the VCO in complementary metal-oxide-semiconductor (CMOS) technology. The VCO core is three inverters connected in a ring. These inverters are formed by transistors M1–M6. Also, an output inverter on transistors M17 and M18 is connected to the output of one of the inverters to reduce the rise and fall time of the output pulse signal of the VCO. Limiting transistors M10–M12 and M14–M16 are used to control the frequency of VCO. These transistors limit the current consumed by the VCO core inverters during operation. The lower the maximum current through the limiting transistors, the more time the inverter needs to recharge the parasitic capacity of the next inverter, therefore, the lower the frequency of pulses. With the help of current mirrors formed by transistors M8, M9, M13–M16, the maximum current through the limiting transistors is set. The control current is formed by a voltage converter into a current formed by the transistor M7 and resistors R_1, R_2, R_h, R_s. The linearity of the VCO transfer characteristic is provided by the linear transfer characteristic of the voltage-to-current converter [11]. Resistors R_1, R_2, R_s provide an extension of the linearity range of the transfer characteristic. The resistor R_h is used to adjust the VCO gain [12]. By independently changing the channel width of the limiting transistors, the channel width of the transistor M7, and the values of the resistors R_h, R_s, it is possible to set the required maximum and minimum pulse generation frequency of the VCO.

Design of the ADC digital unit

The digital unit of the proposed ADC consists of a subtracting counter and a parallel register. The counter counts the number of pulses received at its input during the sampling period. The counter's bit depth is equal to the ADC's bit depth, in our case, it is 8 bits. The register is necessary to fix the counter value according to the sampling cycle. To design the digital unit of the ADC, automatic synthesis of digital circuits from the hardware description was applied. The hardware descriptions of the counter, the parallel register, and the ADC digital unit are shown in Listings 1–3, respectively.

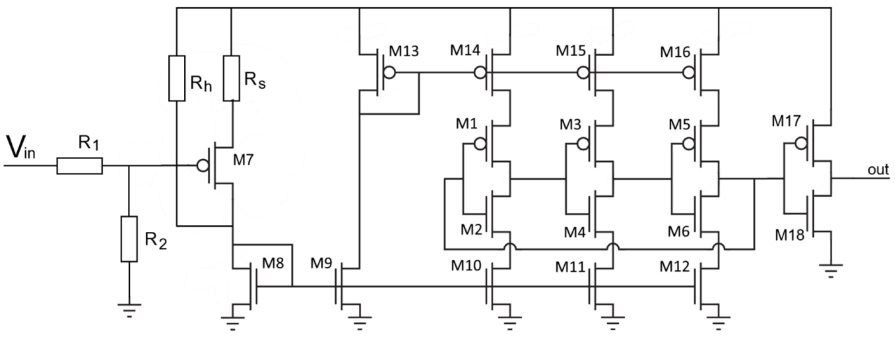


Fig. 2. Circuit of the VCO

Listing 1. Hardware description of the counter

```
module sabstrcnt(r, c, x, y);

	input r, c, x;
	output reg [7:0]y;
	always @( negedge r, posedge x, posedge c)
	begin
		if(!r)
			y <= 'd255;
		else
		begin
			if (c)
				y <= 'd255;
			else
				y <= y-'d1;
		end
	end

endmodule
```

Listing 2. Hardware description of the parallel register

```
module parr(r,c,x,y);

	input r, c;
	input [7:0]x;
	output reg [7:0]y;
	always @( negedge r, posedge c)
		if (!r)
			y <= 0;
		else
			y <= x;

endmodule
```

Listing 3. Hardware description of the ADC digital unit

```
module DU(r,c,x,y);

	input r,c,x;
	wire [7:0]w;
	output [7:0]y;

	sabstrcnt A(r,c,x,w);
	parr B(r,c,w,y);

endmodule
```

The digital unit of the ADC was synthesized using the built-in CAD tools. The circuit was synthesized from standard cells using 180 nm CMOS technology with a supply voltage of 1.8 V from Mikron JSC. The circuit of the digital unit is shown in Fig. 3.

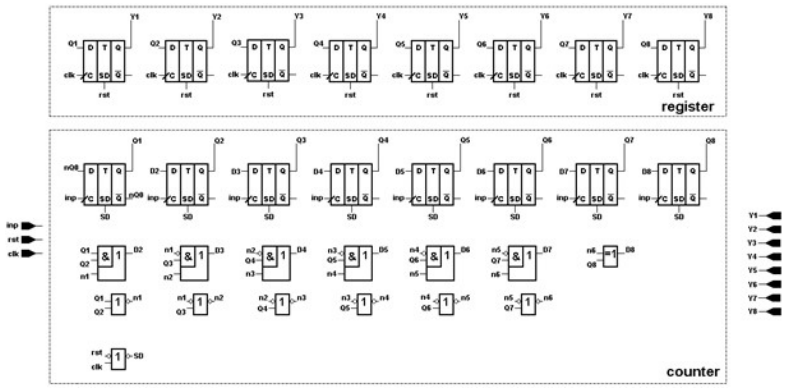


Fig. 3. Circuit of the digital unit of the ADC

Layout of the ADC

Fig. 4 shows the layout of the designed ADC based on the VCO. The VCO is located in the upper left part of the layout. To the right of the VCO, three inverters connected in series are placed. These three inverters convert the amplitude of the VCO output pulses from 3.3 to 1.8 V, because the digital unit of the ADC is designed using cells with a supply voltage of 1.8 V.

The quantization error of the VCO-based ADC depends on the difference between the rise and fall times of the pulse, herewith it is minimal if the difference is zero [13]. The digital unit of the ADC is located at the bottom of the layout in Fig. 4. The D flip-flop cells of the counter are located closest to the VCO, while the combinational logic is located below them. The D flip-flop cells of the output register are located on the right and bottom sides of the layout of the ADC digital unit. The dimensions of the ADC layout are 103 μm by 109 μm.

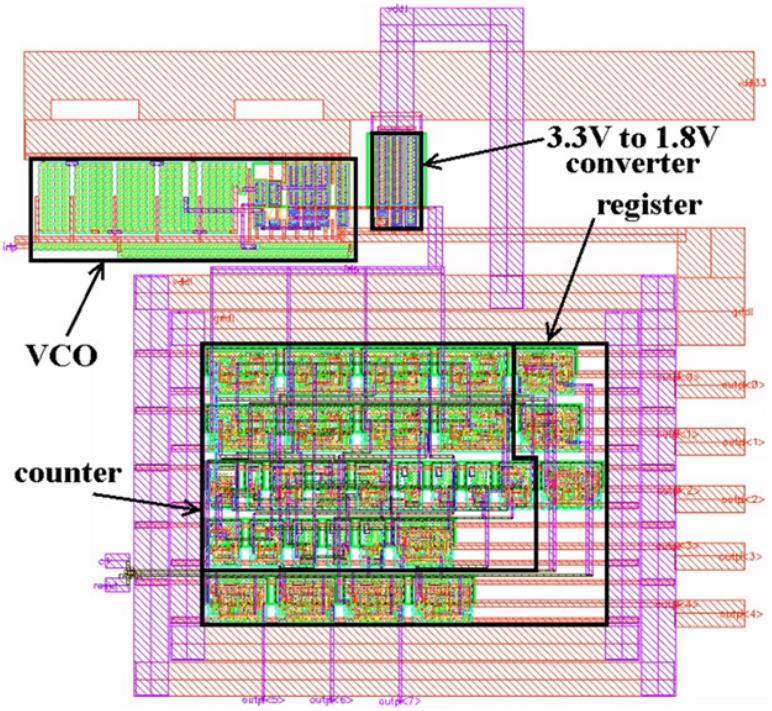
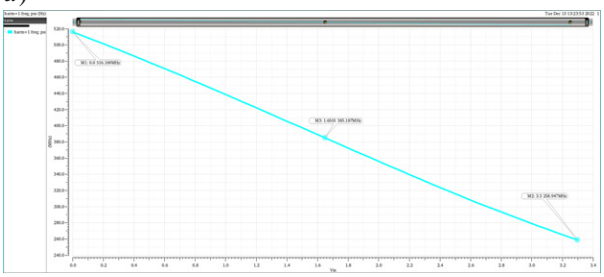


Fig. 4. VCO-based ADC layout

a)



b)

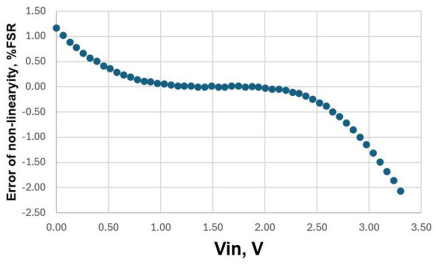


Fig. 5. Simulation of the transfer characteristic of the VCO

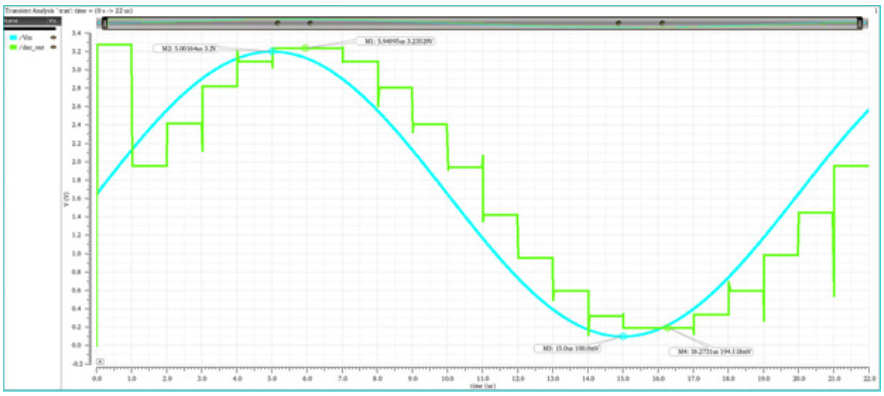


Fig. 6. ADC input and output signals

Simulation of ADC operation

A simulation of the operation of the VCO-based ADC was carried out. Since the characteristics of the ADC depend on the linearity of the transfer characteristic of the VCO, the dependence of the frequency of VCO on the input signal level was built. Fig. 5a shows the transfer characteristic of the VCO obtained using the PSS analysis. Fig. 5b shows a graph of the dependence of the nonlinearity error on the ADC input signal level. The average error of the nonlinearity of the VCO gain is 0.43% of the full scale of measurements (FSR).

Fig. 6 shows the input and output signals of a VCO-based ADC. The amplitude of the input signal was set to 1.55 V, and the frequency was 50 kHz. The outlier in the output signal of the ADC at the initial time occurred due to the initial values in the output register. Using the built-in Cadence Virtuoso tools, the ADC dynamic characteristics were determined based on the output signal spectrum. The values for the first sampling cycle of the ADC output signal were not used when calculating the parameters. A signal-to-noise ratio (SNDR) of 36.48 dB has been achieved which means an effective number of bits (ENOB) of 5.77 bits.

Conclusion

An 8-bit voltage-controlled oscillator-based ADC with a range of input values from 0 to 3.3 V using 180 nm CMOS technology from Mikron JSC with a supply voltage of 3.3 V in the analog part and 1.8 V in the digital part is presented. The simulation of the ADC in the time domain was performed in Cadence Virtuoso. The sampling rate was set to 1 MHz. The power consumption is about 1.8 mW. With an input frequency of 50 kHz and an amplitude of 1.55 V, the ADC provides an output SNDR of 36.48 dB (ENOB is 5.77 bits). This low-power design is applicable in the IoT, medical care and other autonomous devices.

REFERENCES

1. **Gielen G.G.E., Hernandez L., Rombouts P.** Time-encoding analog-to-digital converters: Bridging the analog gap to advanced digital CMOS-part 1: Basic principles. *IEEE Solid-State Circuits Magazine*, 2020, Vol. 12, No. 2, Pp. 47–55. DOI: 10.1109/MSSC.2020.2987536

2. **Li S., Sanyal A., Lee K., Yoon Y., Tang X., Zhong Y., Ragab K, Sun N.** Advances in voltage-controlled-oscillator-based δσ ADCs. *IEICE Transactions on Electronics*, 2019, Vol. 102, No. 7, Pp. 509–519. DOI: 10.1587/transele.2018CTI0001

3. **Anand T., Makinwa K.A.A., Hanumolu P.K.** A VCO based highly digital temperature sensor with 0.034 C/mV supply sensitivity. *IEEE Journal of Solid-State Circuits*, 2016, Vol. 51, No. 11, Pp. 2651–2663. DOI: 10.1109/JSSC.2016.2598765

4. **Tu C.-C., Wang Y.-K., Lin T.-H.** A low-noise area-efficient chopped VCO-based CTDSM for sensor applications in 40-nm CMOS. *IEEE Journal of Solid-State Circuits*, 2017, Vol. 52, No. 10, Pp. 2523–2532. DOI: 10.1109/JSSC.2017.2724025

5. **De la Rosa J.M., Schreier R., Pun K.-P., Pavan S.** Next-Generation delta-sigma converters: Trends and perspectives. *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, 2015, Vol. 5, No. 4, Pp. 484–499. DOI: 10.1109/JETCAS.2015.2502164

6. **Quintero A., Buffa C., Perez C., Cardes F., Straeussnigg D., Wiesbauer A., Hernandez L.** A coarse-fine VCO-ADC for MEMS microphones with sampling synchronization by data scrambling. *IEEE Solid-State Circuits Letters*, 2020, Vol. 3, Pp. 29–32. DOI: 10.1109/LSSC.2020.2964158

7. **Ellaithy D.M.** Voltage-controlled oscillator based analog-to-digital converter in 130-nm CMOS for biomedical applications. *Journal of Electrical Systems and Information Technology*, 2023, Vol. 10, Art. no. 38. DOI: 10.1186/s43067-023-00109-x

8. **Shibata H., Taylor G., Schell B., Kozlov V., Patil S., Paterson D. et al.** 16.6 an 800MHz-BW VCO-based continuous-time pipelined ADC with inherent anti-aliasing and on-chip digital reconstruction filter. *2020 IEEE International Solid-State Circuits Conference* (*ISSCC*), 2020, Pp. 260–262. DOI: 10.1109/ISSCC19947.2020.9062917

9. **Wu T.-F., Chen M.S.-W.** 16.7 A 40MHz-BW 76.2dB/78.0dB SNDR/DR noise-shaping nonuniform sampling ADC with single phase-domain level crossing and embedded nonuniform digital signal processor in 28nm CMOS. *2020 IEEE International Solid-State Circuits Conference* (*ISSCC*), 2020, Pp. 262–264. DOI: 10.1109/ISSCC19947.2020.9063022

10. **Zhong Y., Li S., Tang X., Shen L., Zhao W., Wu S., Sun N.** A second-order purely VCO-based CT ΔΣ ADC using a modified DPLL structure in 40-nm CMOS. *IEEE International Solid-State Circuits*, 2020, Vol. 55, No. 2, Pp. 356–368. DOI: 10.1109/JSSC.2019.2948008

11. **Andryzcik S., Graham D.W.** Linearization of voltage-controlled oscillators using floating-gate transistors. *IEEE Transactions on Circuits and Systems II: Express Briefs*, 2021, Vol. 68, No. 7, Pp. 2337–2341. DOI: 10.1109/TCSII.2021.3059109

12. **Voelker M., Pashmineh S., Hauer J., Ortmanns M.** Current feedback linearization applied to oscillator based ADCs. *IEEE Transactions on Circuits and Systems I: Regular Papers*, 2014, Vol. 61, No. 11, Pp. 3066–3074. DOI: 10.1109/TCSI.2014.2327302

13. **Razavi B.** *Design of CMOS Phase-Locked Loops: From Circuit Level to Architecture Level*. Cambridge (UK): University Printing House, 2020. DOI: 10.1017/9781108626200

INFORMATION ABOUT AUTHORS / СВЕДЕНИЯ ОБ АВТОРАХ

Vitaly D. Bystrov
Быстров Виталий Дмитриевич
E-mail: bystrov.vd@edu.spbstu.ru

Dmitry V. Morozov
Морозов Дмитрий Валерьевич
E-mail: dvmorozov@inbox.ru
ORCID: https://orcid.org/0000-0003-3403-0120

Mikhail M. Pilipko
Пилипко Михаил Михайлович
E-mail: m_m_pilipko@rambler.ru
ORCID: https://orcid.org/0000-0003-3813-6846

Submitted: 20.05.2025; Approved: 12.09.2025; Accepted: 22.09.2025.
Поступила: 20.05.2025; Одобрена: 12.09.2025; Принята: 22.09.2025.

Computing, Telecommunication and Control, 2025, Vol. 18, No. 3, Pp. 131–143.
Информатика, телекоммуникации и управление. 2025. Том 18, № 3. С. 131–143.

Software and Hardware of Computer, Network, Telecommunication, Control, and Measurement Systems

Компьютерные сети, вычислительные, телекоммуникационные, управляющие и измерительные систем

Research article
DOI: https://doi.org/10.18721/JCSTCS.18312
UDC 004.896

AUTOMATION OF PREPARATION AND DEPLOYMENT OF INFORMATION INFRASTRUCTURE OF CLOUD SERVICES USING THE ANSIBLE TOOL

A.A. Kiseleva[1] ✉, *I.V. Nikiforov*[1], *M.V. Shishko*[2], *S.M. Ustinov*[1]

[1] Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russian Federation;

[2] Gazpromneft Information Technology Operator LLC, St. Petersburg, Russian Federation

✉ aleksandrakiseleva2001@gmail.com

Abstract. The preparation and deployment of information technology infrastructure when implementing corporate collaboration platforms in industrial companies takes up to 40% of total time spent on a project. Automation of these processes helps reduce indicated time costs. This article proposes an approach to automate the preparation and deployment of information infrastructure, as well as to automate the installation of cloud services using the Ansible application. The R7-office platform is chosen as the implemented solution, which has basic necessary functionality for working with documents of various formats, enables collaborative document editing and meets state requirements for import substitution of components. The approach is based on an algorithm for preparing information technology infrastructure and installing the target platform using a software tool written in the Python programming language. It enables automatic parsing of formalized requirements, generating access rights requests for users and creating playbooks for configuring the infrastructure using Ansible. According to the experimental results, the implementation of the proposed approach reduces the labor intensity of the infrastructure preparation process by 45%.

Keywords: R7-office, collaborative editing, automation, IT infrastructure, Ansible

Citation: Kiseleva A.A., Nikiforov I.V., Shishko M.V., Ustinov S.M. Automation of preparation and deployment of information infrastructure of cloud services using the Ansible tool. Computing, Telecommunications and Control, 2025, Vol. 18, No. 3, Pp. 131–143. DOI: 10.18721/JCSTCS.18312

Научная статья
DOI: https://doi.org/10.18721/JCSTCS.18312
УДК 004.896

АВТОМАТИЗАЦИЯ ПОДГОТОВКИ И РАЗВЕРТЫВАНИЯ ИНФОРМАЦИОННОЙ ИНФРАСТРУКТУРЫ ОБЛАЧНЫХ СЕРВИСОВ С ИСПОЛЬЗОВАНИЕМ ИНСТРУМЕНТА ANSIBLE

А.А. Киселева[1] ✉, *И.В. Никифоров*[1], *М.В. Шишко*[2], *С.М. Устинов*[1]

[1] Санкт-Петербургский политехнический университет Петра Великого, Санкт-Петербург, Российская Федерация;

[2] ООО «Газпромнефть Информационно-Технологический оператор», Санкт-Петербург, Российская Федерация

✉ aleksandrakiseleva2001@gmail.com

Аннотация. Подготовка и развертывание информационно-технологической инфраструктуры при внедрении корпоративных платформ совместной работы в индустриальных компаниях может занимать до 40% общего времени, затрачиваемого на проект. Автоматизация этих процессов помогает снизить указанные временные трудозатраты. В данной работе предложен подход, позволяющий автоматизировать подготовку и развертывание информационной инфраструктуры, а также установку облачных сервисов при помощи приложения Ansible. В качестве внедряемого решения выбрана платформа «Р7-офис», которая обладает основным необходимым функционалом работы с документами различного формата, позволяет совместно редактировать документы, а также удовлетворяет требованиям государства по импортозамещению компонент. В основе подхода лежит алгоритм по подготовке информационно-технологической инфраструктуры и установке целевой платформы с помощью программного средства, написанного на языке программирования Python. Оно позволяет производить автоматический разбор формализованных требований, составление заявок на предоставление прав доступа пользователям и формирование плейбуков для настройки инфраструктуры с помощью Ansible. По результатам проведенных экспериментов внедрение предложенного подхода позволяет сократить трудоемкость процесса подготовки инфраструктуры на 45%.

Ключевые слова: Р7-офис, совместное редактирование, автоматизация, ИТ-инфраструктура, Ansible

Для цитирования: Kiseleva A.A., Nikiforov I.V., Shishko M.V., Ustinov S.M. Automation of preparation and deployment of information infrastructure of cloud services using the Ansible tool // Computing, Telecommunications and Control. 2025. Т. 18, № 3. С. 131–143. DOI: 10.18721/JCSTCS.18312

Introduction

Industrial companies widely use office applications for creating documents, spreadsheet, presentations, as well as for communicating and holding video conferences [1, 2]. For project teams, it is important to be able not only to edit files locally, but also to work with them collaboratively [3], since it saves up to 20% of time due to the absence of the need to transfer documents and synchronize them [4].

There are several products that provide collaborative editing services [5]. Some of them support the deployment on the company's internal servers, while others are available only as cloud services. There are both foreign solutions and domestic analogues [6].

It is also important to consider how much time, money and effort will be spent to implement a particular office suite, as this affects the timing and cost of the final product: the more labor costs are

required to implement a solution, the longer the implementation and the higher its cost [7]. Therefore, the issue of choosing the method for deploying the platform within the infrastructure is also relevant [8].

Thus, an urgent task for companies is the implementation of a Russian corporate collaboration platform for the effective work of employees in projects with the maintenance of common documentation. At the same time, the implementation process itself can take up to 40% of the project time and require additional resources from the project administrator to generate and track various requests and engineers to directly configure the infrastructure. For this reason, it is necessary to develop a way to automate the preparation of the infrastructure and the deployment of software on it [9–13].

Comparative analysis of office packages

There are many solutions for editing a document by a group of people [3]. These solutions have similar functionality and are aimed at solving similar problems. To make a choice, it is necessary to conduct a comparative analysis of office packages [14, 5]. Let us take six of the most popular solutions: Yandex 360 [15, 16], Microsoft 365 [17, 18], Notion [19], Slite[1] (https://slite.com), Notejoy [20], R7-Office [21, 22].

Comparative analysis of office packages (Table 1) was conducted according to the most important criteria from the point of view of implementation [6]. These criteria are listed below.

1. The "Collaboration" criterion indicates the possibility or impossibility of collaborative document editing. This functionality is key in this article.

2. The criterion of "Completeness of functionality" is a cumulative criterion that reflects how extensive the functionality offered by the office suite is. The extent of the functionality affects how many tasks user is able to perform.

3. The "Cost" criterion indicates the cost of the solution's licenses. This criterion is highly important for business, as it affects the total cost of the solution.

4. The criterion of "Import dependence" indicates how possible it is to use the package in the Russian Federation, as well as whether the manufacturer is domestic. This is one of the key criteria for companies within the framework of the import substitution policy.

5. The "Using plugins/integrations" criterion reflects whether additional plugins and integrations are provided by the software. Such functionality is important for the further development of the solution, taking into account the specifics of the company.

6. The criterion of "Scalability" indicates how much the system can expand. This is important when the number of employees in the company is growing.

7. The "Compatibility with Russian operating systems" criterion reflects whether the package can work with various operating systems, including domestic ones. This criterion is also important in terms of the policy of software import substitution in companies.

8. The criterion of "Deployment on your own servers" indicates the possibility of deploying the software on the company's internal corporate network. This criterion is also the key one, as it directly affects the security of using the solution.

9. The "Maximum number of users" criterion reflects how many users an office suite can accommodate, which affects the scalability of the solution and the growth of the number of users during the development of the company.

Each of the listed criteria has a score, as well as a weight. For criteria 1, 4, 5, 6, 7, 8, the score is 0 or 1, depending on the presence or absence of the specified functionality. For criteria 2, 3, 9, there is a range of values where the maximum score is assigned when the criterion is exhaustively matched, and the minimum is assigned in the case of significant restrictions on the criterion in comparison with other office packages. The weight indicates the importance of the criterion within the framework of the studied task of implementing an office suite.

[1] Slite | AI-powered knowledge base, Available: https://slite.com (Accessed 02.09.2025)

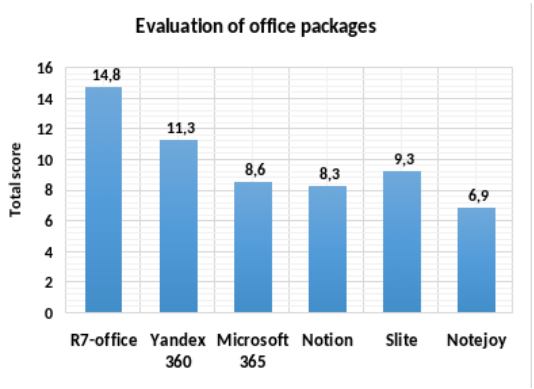


Fig. 1. Evaluation of office packages

As a result of the analysis, each office suite is assigned a total score, which is formed according to the following formula:

$$s = k_1 \times p_1 + k_2 \times p_2 + \ldots + k_n \times p_n,$$

where k_n is the score according to the criterion n, p_n is the weight by criterion n.

Table 1

Comparative analysis of office packages

Criteria	Range	Weight	Yandex 360	Microsoft 365	Notion	Slite	Notejoy	R7-office
Collaboration	0/1	1	1	1	1	1	1	1
Full functionality	1.5	1	5	5	3	3	2	5
Cost	1.5	0.5	3	1	2	4	4	5
Import dependence	0/1	1	1	0	0	0	0	1
Using plugins/integrations	0/1	0.5	0	0	1	1	1	1
Ability to scale	0/1	0.7	1	1	1	1	1	1
Compatibility with Russian operating systems	0/1	1	0	0	0	0	0	1
Deployment of own servers	0/1	1	0	0	0	0	0	1
Maximum number of users	1.3	0.7	3	2	3	3	1	3
Total			11.3	8.6	8.3	9.3	6.9	14.8

To decide on choosing an office suite, a bar chart has been compiled reflecting the results of the comparative analysis (Fig. 1).

Based on the comparative analysis of office packages, it can be concluded that the R7-office shows the best results according to most of the criteria presented.

Comparative analysis of deployment tools

Let us take four of the most popular tools: Puppet [23, 24], Chef [25], Ansible [26–28], SaltStack [29, 30].

Comparative analysis of deployment tools (Table 2) was conducted according to the criteria listed below.

Table 2

Comparative analysis of deployment tools

Criteria	Range	Weight	Puppet	Chef	Ansible	SaltStack
Level of development	1.3	1	1	2	3	3
Ease of use	1.3	0.7	2	2	3	3
Security	1.3	1	3	1	3	2
Main language	1.3	0.5	1	1	3	3
Documentation	1.3	1	3	2	3	1
Flexibility	1.3	0.5	3	3	2	3
Total			10.4	8.4	13.6	11.1

1. The “Level of development” criterion reflects how much effort a specialist of an average competence needs to master the tool. This criterion directly affects the labor input required for its implementation.

2. The criterion of “Ease of use” shows how intuitive the tool is for a specialist. It also affects the level of mastery and comfort of using the tool by a specialist.

3. The “Security” criterion assesses the tool's reliability and safety in deployment, whether it contains vulnerabilities that could allow the leakage of company data or the introduction of malware into the corporate network. This is one of the most important criteria, since the security of user data is a top priority for companies.

4. The criterion of “Main language” indicates the threshold for entry and the ease of mastering the tool, which is determined by the programming language it utilizes.

5. The “Documentation” criterion reflects how fully the system is equipped with the necessary documentation for using the tool. The completeness of the documentation also affects the speed of mastering the tool by a specialist.

6. The criterion of “Flexibility” reflects how flexible the tool is to use for different tasks and on different infrastructures. This criterion determines the breadth of use of the tool.

As with comparison of office packages, each criterion has a score and a weight. Similarly, the point score is a range of values. The weight indicates the importance of the criterion within the framework of the studied task of implementing an office suite.

Each instrument is assigned a total score, which is formed according to the following formula:

$$s = k_1 \times p_1 + k_2 \times p_2 + \ldots + k_n \times p_n,$$

where k_n is the score according to the criterion n, p_n is the weight by criterion n.

To make a choice of an IT infrastructure preparation tool, a bar chart has been compiled reflecting the results of the comparative analysis (Fig. 2).

The tools considered are useful and applicable in various tasks. In this article, Ansible tool is chosen since it shows the best results according to most of the criteria presented.

Proposed solution concept

This paper proposes an approach for automated deployment of information infrastructure in data centers, as well as automated installation of the “R7-office” office package.

The infrastructure preparation and the software deployment using the developed software tool, implementing the proposed approach, takes place according to the stages shown in Fig. 3.

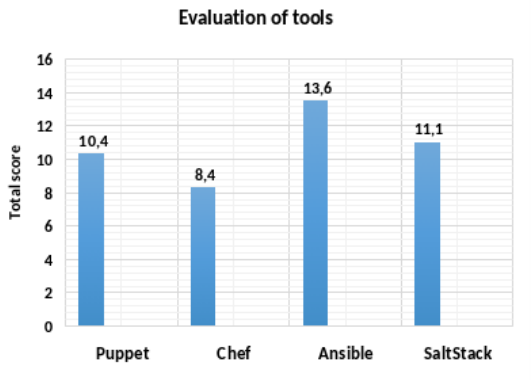


Fig. 2. Evaluation of tools

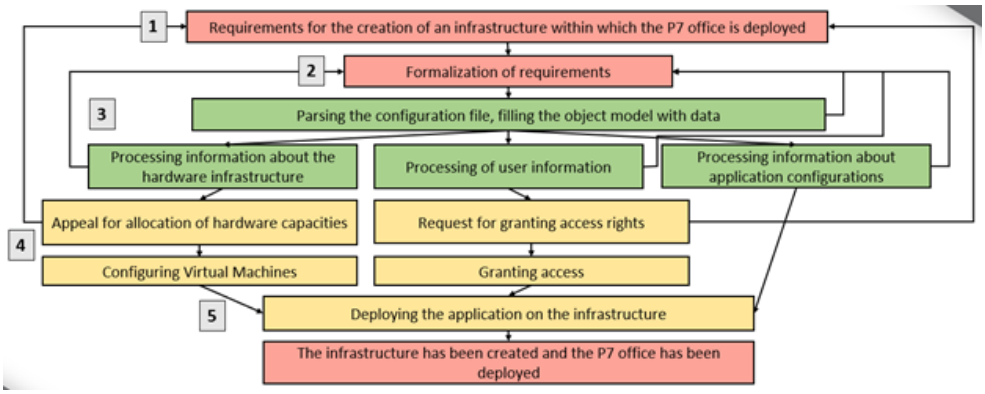


Fig. 3. Scheme of IT infrastructure preparation and software deployment

The stages are as follows:

1. Collecting requirements for the software, the infrastructure for its deployment, as well as a list of users with their roles for access granting.

2. Formalizing requirements into the established format using the instructions. The output is a generated CSV-file containing structured, formalized infrastructure requirements and a list of system users with their roles for access granting. The requirements are formalized in a specific way to ensure the correct operation of the software.

3. Using Python scripts to parse the requirements file and generate configuration files for setup:

– YAML files for hardware configuration.

– A Bash-script for deployment application.

– Data for creating requests for the allocation of capacities and granting system access to users.

The described scripts are implemented within a software tool developed in Python, which aims to automatically parse requirements and generate the necessary output files for further stages.

4. Preparing the cluster and using Ansible [26–28] to configure the servers in accordance with the collected requirements, granting access to users. This includes configuring the OTRS system [31] to process requests for granting access to the office suite for employees and allocating capacity.

5. The final deployment of the application, which results in the infrastructure for the information system.

Implementation

1. Collecting and defining requirements for creating an infrastructure for R7-office deployment

At the initial stage, the system deployment requirements are identified. The minimum required set of requirements is divided into the following categories:

– Hardware infrastructure requirements:
– required infrastructure;
– technical requirements in a formalized form;
– a list of necessary access rights for employees.

2. Formalization of requirements

Using the instructions in CSV format (Fig. 4), the requirements are recorded in a single CSV file in a formalized form. To fill in the configuration file correctly, an instruction has been developed, also in CSV format. The file contains general information about the project with which the system is being implemented.

The structure of the CSV-file is as follows:

– General information about the project is specified at the top of the file: a code for quick search of the project in various information systems of the company, the project name, its purpose and objectives.

– This is followed by information about the system users for whom access needs to be requested, under the keyword "Users". The information is filled in without specific appearance and formatting requirements, except for the column-based organization of information.

– Next, the hardware infrastructure requirements are listed under the keyword "Infrastructure". The information is also filled in without specific appearance and formatting requirements, except for the column-based organization of information.

– After that, the technical requirements for the infrastructure are described under the keyword "Technical requirements". The description is filled in freely, with formalization following specific rules to enable automatic conversion of requirements into configuration files.

To enable correct formalization of infrastructure requirements and their entry into a CSV-file, an instruction has been developed in CSV format. It describes two blocks.

The first block contains:

– a description of the types of YAML files [32] that the application supports, a description of the resulting functionality and a list of supported keywords;

Код проекта	H347620032							
Наименование проек	Внедрение корпоративной платформы совместной работы							
Цель	Проектирование и создание автоматизированной информационной инфраструктуры во внутренней корпоративной сети и развертывания на ней Р7-офиса.							
Задачи	1) Создание архитектуры информационной инфраструктуры для развертывания решения 2) Разработка подхода, позволяющего автоматизировать подготовку и развертывание информационной инфраструктуры и автоматизировать установку сервисов 3) Развертывание Р7-офиса.							
Пользователи								
ФИО	Почта	Роль						
Сидоров Василий Ог	sv@mail.ru	Пользователь						
Янукова Алиса Викто	ya@mail.ru	Администратор						
Кропина Лидия Павл	kl@mail.ru	Пользователь						
Николаев Егор Борис	ne@mail.ru	Администратор ИБ						
Градов Павел Алексе	gp@mail.ru	Пользователь						
Инфраструктура								
Имя сервера	Размещение	Тип	Роль	ОС	IP-адрес	Core	RAM	Дисковая память
1	Внутренняя СПД	IaaS	Сервер совместной работы	Linux	10.43.234	40	56	2048
2	Внутренняя СПД	IaaS	Сервер совместной работы (ре	Linux	10.43.234	24	56	2048
3	Внутренняя СПД	IaaS	Сервер документов	Linux	10.43.234	24	56	2048
4	Внутренняя СПД	IaaS	Сервер документов	Linux	10.43.234	24	56	2048
5	Внутренняя СПД	IaaS	Сервер документов (резервный	Linux	10.43.234	8	56	2048
6	Внутренняя СПД	IaaS	Сервер хранения данных	Linux	10.43.234	8	16	50000
7	Внутренняя СПД	IaaS	Сервер хранения данных (резе	Linux	10.43.234	8	16	50000
8	Внутренняя СПД	IaaS	Сервер мониторинга	Linux	10.43.234	16	32	2048
9	Внутренняя СПД	IaaS	Почтовый сервер	Linux	10.43.234	16	56	2048
Технические требования								
ID	Описание	Формализация						
100	Развертывание Р7-офиса с помощью bash-скрипта на сервере	title: Script; id: 1,2; name: Install python; package: python3; name: Copy file; src: /home/manager/Desktop/he						
101	Развертывание базы данных postgresql на сервере документов	title: PostgreSQL; id: 6,7; name: Install PostgreSQL; package: postgresql; name: Start PostgreSQL service; pac						
102	Развертывание системы мониторинга prometheus на сервере	title: Prometheus; id: 8; name: Install Prometheus; package: prometheus; name: Start Prometheus service; pa						
103	Развертывание почтового сервиса sendmail на почтовом серв	title: Sendmail; id: 9; name: Install Sendmail; package: sendmail; name: Start Sendmail service; package: send						
104	Пример другого сервиса	title: Other; id: 9; name: Install Other; package: other; name: Start Other service; package: other						

Fig. 4. File containing the requirements

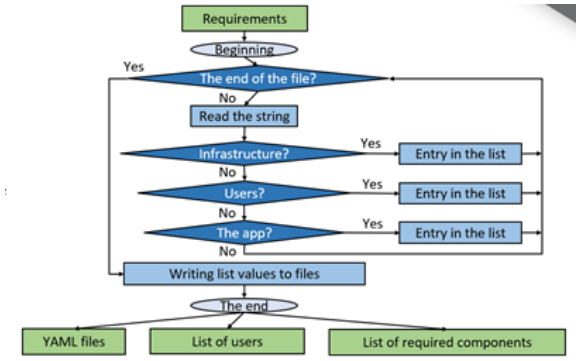


Fig. 5. Requirement analysis algorithm

– an algorithm of actions for cases when the file types implemented in the software are insufficient. In this case, the specialist can integrate support for a new type of service into the existing code.

The second block contains:

– requirements for the formalization of the functionality description;

– requirements for the formation of the source data file.

The instruction file is for informational purposes only for the employee who defines the requirements and is not used by the software. However, incorrect formalization of requirements makes proper use of the software solution impossible.

3. Parsing the configuration file, filling the object model with data

Automatically, using a Python software tool, the file is parsed into three components according to the requirement categories. Next, each part of the configuration file is processed.

The algorithm is shown in Fig. 5.

Description:

– CSV-file text is read line by line until the end of the file is reached;

– requirement category is identified using keywords in the configuration file;

– upon completion of the file analysis, three lists are formed with data corresponding to each category;

– data from each list is processed according to its own algorithm to generate output files;

– received output files are used to create OTRS requests for resource allocation and user access provisioning, as well as for configuring the infrastructure.

The processing of the CSV-file with requirements begins with a general analysis that separates it into three components. This is accomplished using a single library – “csv”.

The file is read line by line. When one of the keywords is encountered: “Users”, “Infrastructure” or “Technical requirements”, – the relevant information is added to a dedicated list: user_list[], infra_list[] or tech_list[] using standard list manipulation methods.

Next, the data related to the list of users and the infrastructure are saved to separate “Users.csv” and “Infra.csv” files for further processing. Data related to technical requirements is passed to the next block of code for further processing and generation of Ansible playbooks.

4. Processing information about hardware infrastructure and users

Algorithm:

– data from the list is converted into text for an e-mail;

– e-mails are automatically sent to the OTRS system;

– the contractor processes and fulfill the requests;

– the request initiator tracks the status of registered requests, using a Python script, within the developed software.

The libraries “smtplib”, “csv”, “email” and “os.path” are used for implementation. The “Users.csv” file is parsed by the software, and an e-mail is generated requesting system access for users. The “Infra.csv” file is not converted in any way, but is sent to the OTRS system as an attachment. Thus, two e-mails with the necessary requests are sent to the system’s e-mail address for request registration.

In response to the created request, an e-mail with the number of the registered requests is sent to the sender. The developed software tool then enables request status tracking. To implement this feature, the libraries “datetime” and “pyotrs” are used. The user enters the tracking numbers into a text file and, after testing the program, receives a file with the status of the specified requests.

An OTRS system is configured to support this software solution. It is an open system for processing requests, which can be customized according to specific requirements. The system is prepared as follows:

– OTRS installation on a virtual machine with initial configuration;
– e-mail integration for sending and receiving system messages, including protocol configuration;
– creation of system agents and customers, as well as a request queue;
– configuration of request generation from e-mails;
– setup of automatic user notifications about the creation of a request.

5. Processing information about application configurations

Algorithm:

– data from the corresponding list is parsed by keywords and converted into several YAML files, serving as playbooks for Ansible.
– using the generated files and Ansible, virtual machines are configured and the office suite is deployed.

From the previously obtained tech_list[], the playbooks for virtual machines are generated.

The initial function decomposes the list into its components. The requirement ID is passed to a function that generates shared files for configuring Ansible: hosts and *_servers, where the * symbol indicates the name of the deployed service. In it, according to the received ID and using the infra_list[], the IP address of the server is allocated and specified in the generated files. The requirements are also separated by the header keywords corresponding to the services being installed and transferred to functions that process each service separately. If a header contains a word unrelated to the keywords, a general YAML file is created for later specialist customization.

After a general analysis of the requirements, the YAML file is prepared for a specific service. Each supported service has a dedicated function. The structure of each such function is similar but differs in the specifics of configuring a particular service. The requirements are processed and converted into the YAML format of the playbook. For supported services like the PostgreSQL database, a .sql file is also generated to automatically add user information to the database.

6. Configuring virtual machines using Ansible

The configuration of the infrastructure using Ansible is as follows:

– initial configuration of virtual machines is performed: node manager and worker nodes, preparing them for Ansible operation;
– playbooks obtained by parsing the configuration file are transferred to a virtual machine that plays the role of node manager, and are launched sequentially, configuring the worker nodes and deploying the R7-office system.

To set up a virtual machine and deploy a R7-office, the following playbooks are created and used:

– a playbook for configuring the PostgreSQL database, which stores information about users and their roles for access control;
– a playbook for setting up the Prometheus monitoring server;
– a playbook for configuring the Sendmail mail server, enabling office-mail system integration;
– a playbook for launching a bash script, that handles the R7-office deployment program.

Future developments of the software solution may include additional playbooks for configuring supplementary services.

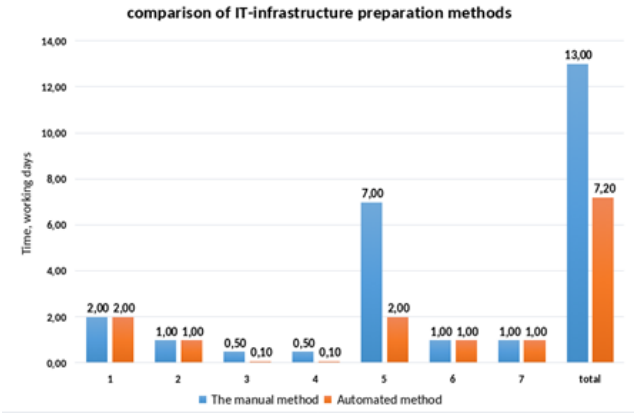


Fig. 6. Comparison of IT infrastructure preparation approaches

Experimental studies

Based on operational experience, an approximate time estimate has been made for the manual approach to parsing requirements for infrastructure preparation. This approach requires specialists to manually create playbooks for Ansible and submit requests for resource allocation and user access rights provisioning [4]. On average, for this method, with the number of servers up to 5, the number of users up to 100 and the number of deployed services up to 5, an average specialist needs up to two weeks to complete these tasks.

The calculation of the time taken by the proposed automated approach is carried out under constrained laboratory conditions using two nodes. According to the results, the full cycle of automated preparation of the IT infrastructure takes approximately 7 days. The automation of requirements analysis and configuration file generation reduced specialist workload by a factor of 5. The calculation of the values is shown in Table 3. Yellow indicates the phases that are automated using the proposed approach.

Table 3

Calculation of the time required to prepare the IT infrastructure in working days

Phase	A manual approach, (days)	An automated approach, (days)
Collecting requirements	2	2
Writing to a file	1	1
File analysis, formation of documents with parts of the requirements	0.5	0.1
Sending requests for access rights	0.5	0.1
Creating playbooks for setting up the infrastructure	7	2
Configuring Ansible on virtual machines	1	1
Launching playbooks	1	1
Total	13	7.2

A visual comparison of the time spent is shown on the graph (Fig. 6). It shows the number of points for each stage and in total.

The resulting temporary advantage is calculated using the following formula:

$$t = \left(t_1 - t_2\right) \div t_1 \times 100,$$

where t_1 is the initial time for the implementation of the manual approach, t_2 is the time obtained due to automation.

Thus, the average software deployment time during the implementation of the proposed approach is reduced by 45%.

Conclusion

The research conducted in the work and the comparative analysis presented showed that the most suitable package for solving the problem of reducing the company's resources by implementing a Russian corporate collaboration platform is the R7-office package, and the most successful tool for automating deployment according to a set of criteria is the Ansible tool.

The approach proposed in this work allows for automation of the preparation and configuration of the IT infrastructure, as well as the installation of the "R7-office" office package. The approach has the following distinctive features: automatic generation of configuration files for infrastructure configuration, files for requesting capacities and for providing access to users, as well as remote configuration of worker nodes and software deployment. The approach is implemented in a software tool using Ansible and Python technologies.

Experimental studies conducted in a local environment consisting of two nodes allow us to conclude that the time spent on implementation using the proposed solution can be reduced by 45%.

REFERENCES

1. **Gorkavchenko A.A., Zalukaeva E.A., Popenya I.V., Terekhov Ya.A.** Vozmozhnosti onlajn ofisov [Online Office Capabilities]. *Formirovanie i razvitie novoj paradigmy nauki v usloviyah postindustrial'nogo obshchestva: Sbornik statej po itogam Mezhdunarodnoj nauchno-prakticheskoj konferencii* [*Formation and development of a new paradigm of science in the conditions of post-industrial society: Collection of articles based on the results of the International scientific and practical conference*], 2023, Pp. 91–94.

2. **Mori G., Buzzi M.C., Buzzi M., Leporini B., Penichet V.M.R.** Collaborative editing for all: the google docs example. *Universal Access in Human-Computer Interaction. Applications and Services* (*UAHCI 2011*), 2011, Vol. 6768, Pp. 165–174. DOI: 10.1007/978-3-642-21657-2_18

3. **Kirvas V.A.** Sovremennye oblachnye informacionnye tekhnologii sovmestnoj raboty [Modern cloud information technologies for collaboration]. *Ekspertnye ocenki elementov uchebnogo processa: programma i materialy XIX mezhvuzovskoj nauchno-prakticheskoj konferencii* [*Expert assessments of elements of the educational process: program and materials of the XIX interuniversity scientific and practical conference*], 2017, Vol. 25, Pp. 35–43.

4. **Gorodnyaya L.V.** An approach to the measure of the labour required for programming. *Scientific Services & Internet*. 2020. Vol. 22, Pp. 192–209. DOI: 10.20948/abrau-2020-3

5. **Nikonenko A.N.** Methods and criteria for selection of suppliers. *World science: problems and innovations*, 2016, Vol. 2, Pp. 60–62.

6. **Sakoyan A.R., Puchkov V.I.** Challenges of the IT industry in the context of import substitution in 2024. *Lomonosovskie chteniya. Aktual'nye voprosy fundamental'nyh i prikladnyh issledovanij* [*Expert assessments of elements of the educational process: program and materials of the XIX interuniversity scientific and practical conference*], 2024, Pp. 16–20.

7. **Brooks F.P., Jr.** *The Mythical Man-Month* (*Essays on Software Engineering*). Reading: Addison-Wesley Publishing, 1975.

8. **Basynya E.A., Lukina M.S.** Avtomatizirovannaya ustanovka i konfigurirovanie servernyh reshenij [Automated installation and configuration of server solutions]. *Sovremennye materialy, tekhnika i tekhnologii* [*Modern materials, equipment and technologies*], 2016, Vol. 2, No. 5, Pp. 21–26.

9. **Ivlev V.A., Nikiforov I.V.** Aktual'nost' avtomatizirovannoj nastrojki infrastruktury IT-proekta [The relevance of automated configuration of IT project infrastructure]. *Sovremennye tekhnologii v teorii i praktike programmirovaniya* [*Modern technologies in the theory and practice of programming*], 2022, Pp. 79–80.

10. **Ivlev V.A., Mironenkov G.V., Nikiforov I.V., Kovalev A.D.** Generaciya informacionno-tekhnologicheskoj infrastruktury proekta na osnove neformalizovannyh trebovanij [Generation of the project's information technology infrastructure based on informal requirements]. *Sovremennye tekhnologii v teorii i praktike programmirovaniya* [*Modern technologies in the theory and practice of programming*], 2023, Pp. 242–244.

11. **Ivlev V.A., Nikiforov I.V., Yusupova O.A.** Automation method for configuring IT infrastructure for IT projects. *Proceedings of International Conference on Digital Transformation: Informatics, Economics, and Education* (*DTIEE2023*), 2023, Vol. 12637, Art. no. 126370D. DOI: 10.1117/12.2680779

12. **Phan T., Li W.-S.** Automated Configuration of System Infrastructure for SOA-Based Enterprise Computing. *2008 IEEE International Conference on e-Business Engineering*, 2008, Pp. 205–212. DOI: 10.1109/ICEBE.2008.97

13. **Putra W.R.A., Nurwa A.R.A., Priambodo D.F., Hasbi M.** Infrastructure as code for security automation and network infrastructure monitoring. *MATRIK: Jurnal Manajemen, Teknik Informatika, dan Rekayasa Komputer*, 2022, Vol. 22, No. 1, Pp. 201–214. DOI: 10.30812/matrik.v22i1.2471

14. **Panasenko S.V., Panasenko V.A.** Criteria for vendor-enterprises in the market of corporate search. *Newsletter of North-Caucasus Federal University*, 2015, Vol. 47, No. 2, Pp. 144–150.

15. **Kuznecova A.P.** Rossijskie oblachnye servisy dlya biznesa ot kompanij VK i "Yandeks" [Russian cloud services for business from VK and Yandex.]. *Social'no-ekonomicheskaya transformaciya v sovremennom mire: regional'nye proyavleniya* [*Socio-economic transformation in the modern world: regional manifestations*], 2022, Pp. 72–77.

16. **Kudryashov A., Nikitina E.** Comparative analysis of Yandex Office Packages.Documents, OpenOffice, WPS Office. *Nauchnoe soobshchestvo studentov XXI stoletiya, tekhnicheskie nauki* [*Scientific community of students of the 21st century, technical sciences.*], 2023, Pp. 52–59.

17. **Wilson K.** Microsoft office 365. In: *Using Office 365*, 2014, Pp. 1–14. DOI: 10.1007/978-1-4302-6686-0_1

18. **Habraken J.** *Microsoft Office Inside Out* (*Office 2021 and Microsoft 365*). Redmond (Washington): Microsoft Press, 2021.

19. **Mutovina N.V., Sahanov R.R.** Upravlenie IT proektami v Notion [IT Project Management in Notion]. *Proektnyj menedzhment v Kazahstane: sostoyanie, problemy i perspektivy* [*Project management in Kazakhstan: status, problems and prospects*]. 2021.

20. **Burgonov O.V., Konstantinova O.I., Platonov K.A.** Tools of digital transformation of research and educational organizations. *Journal of Legal and Economic Studies*, 2022, Vol. 4, Pp. 263–270. DOI: 10.26163/GIEF.2022.68.50.041

21. **Sharatova N.V.** Perekhod na R7-ofis v ramkah importozameshcheniya. Sravnitel'naya harakteristika s Microsoft Office [Transition to P7-office as part of import substitution. Comparative characteristics with Microsoft Office]. *Morskie tekhnologii: problemy i resheniya – 2023* [*Marine Technologies: Problems and Solutions – 2023*], 2023, Pp. 458–462.

22. **Zaslavskij A.A.** R7 ofis kak otechestvennaya al'ternativa paketu Microsoft Office [P7 office as a domestic alternative to the Microsoft Office package]. *Interaktivnoe obrazovanie* [*Interactive education*], 2022, Vol. 2, Pp. 25–27.

23. **Loope J.** *Managing Infrastructure with Puppet: Configuration Management at Scale*. Sebastopol (California): O'Reilly Media, 2011.

24. **Rhett J.** *Learning Puppet 4: A Guide to Configuration Management and Automation*. Sebastopol (California): O'Reilly Media, 2016.

25. **Yaremchuk S.** Upravlenie konfiguraciej s Chef [Configuration management with Chef]. *Sistemnyj administrator* [*System administrato*r], 2010, Vol. 7–8, Pp. 28–33.

26. **Medvedev O.S.** Nastrojka Linux-serverov s pomoshch'yu centralizovannogo upravleniya Ansible [Setting up Linux servers with centralized Ansible management]. *54-ya nauchnaya konferenciya aspirantov,*

magistrantov i studentov BGUIR [*54th scientific conference of postgraduate, master's and undergraduate students of BSUIR*], 2018, P. 335.

27. **Hochstein L.** *Ansible: Up and Running: Automating Configuration Management and Deployment the Easy Way*. Sebastopol (California): O'Reilly Media, 2017.

28. **Vorontsov Y.A., Mikhailova E.K.** Combined architecture of the configuration management system for computing infrastructure. *International Journal of Open Information Technologies*, 2021, Vol. 9, No. 11, Pp. 17–21.

29. **Anthony S.** *SaltStack for Configuration Management*. Bethlehem (Pennsylvania): Lehigh University, 2018.

30. **Hosmer B.** Getting started with Salt Stack – the other configuration management system built with Python. *Linux journal*, 2013. Available: https://www.linuxjournal.com/content/getting-started-salt-stack-other-configuration-management-system-built-python (Accessed 01.09.2025)

31. **Shabaev T.A.** *Avtomatizaciya processa obsluzhivaniya klientov s pomoshch'yu sistemy OTRS Free* [*Automate your customer service process with OTRS Free*], Bachelor's Thesis, Tomsk Polytechnic University, 2016.

32. **Eriksson M., Hallberg V.** *Comparison between JSON and YAML for data serialization*, Bachelor's Thesis, Royal Institute of Technology, 2011.

INFORMATION ABOUT AUTHORS / СВЕДЕНИЯ ОБ АВТОРАХ

Aleksandra A. Kiseleva
Киселева Александра Александровна
E-mail: aleksandrakiseleva2001@gmail.com
ORCID: https://orcid.org/0009-0004-8263-3289

Igor V. Nikiforov
Никифоров Игорь Валерьевич
E-mail: igor.nikiforovv@gmail.com
ORCID: https://orcid.org/0000-0003-0198-1886

Maxim V. Shishko
Шишко Максим Вадимович
E-mail: shishko.mv.post@gmail.com
ORCID: https://orcid.org/0009-0005-3406-7005

Sergey M. Ustinov
Устинов Сергей Михайлович
E-mail: usm50@yandex.ru
ORCID: https://orcid.org/0000-0003-4088-4798

Submitted: 03.12.2024; Approved: 26.06.2025; Accepted: 17.07.2025.
Поступила: 03.12.2024; Одобрена: 26.06.2025; Принята: 17.07.2025.

Computing, Telecommunication and Control, 2025, Vol. 18, No. 3, Pp. 144–153.
Информатика, телекоммуникации и управление. 2025. Том 18, № 3. С. 144–153.

System Analysis and Control
Системный анализ и управление

Research article
DOI: https://doi.org/10.18721/JCSTCS.18313
UDC 004.021

DEVELOPMENT OF A DUAL-LOOP METHOD OF INTELLIGENT TRAFFIC LIGHT CONTROL BASED ON REINFORCEMENT LEARNING AND HOURLY DISTILLATION OF PHASE STRATEGIES

A.M. Sazanov ✉ , *V.P. Shkodyrev*, *S.M. Ustinov*

Peter the Great St. Petersburg Polytechnic University,
St. Petersburg, Russian Federation

✉ arseny.sazanov@gmail.com

Abstract. With increasingly complex urban dynamics, as well as increasing demands for the sustainability of urban mobility and introduction of cognitive technologies into transport infrastructure, the paper proposes a dual-loop method for intelligent traffic light control based on reinforcement learning and phase strategy distillation procedures. The first level implements real-time control through an RL-agent, while the second one generates backup hourly plans based on statistics of its behavior. The method is based on a system-discrete model taking into account stochastic traffic parameters and permissible control constraints. The simulation conducted in SUMO for a real intersection demonstrates a significant reduction in average transport delay compared to classical control, confirming the efficiency, sustainability and scalability of the approach. The obtained results substantiate the possibility of practical implementation of the model within the framework of intelligent transport systems of large cities and for laying the engineering foundation for hybrid urban mobility management architectures.

Keywords: reinforcement learning, intelligent traffic light control, dual-loop control architecture, traffic light controller, traffic management and control

Citation: Sazanov A.M., Shkodyrev V.P., Ustinov S.M. Development of a dual-loop method of intelligent traffic light control based on reinforcement learning and hourly distillation of phase strategies. Computing, Telecommunications and Control, 2025, Vol. 18, No. 3, Pp. 144–153. DOI: 10.18721/JCSTCS.18313

Научная статья
DOI: https://doi.org/10.18721/JCSTCS.18313
УДК 004.021

РАЗРАБОТКА ДВУХКОНТУРНОГО МЕТОДА ИНТЕЛЛЕКТУАЛЬНОГО СВЕТОФОРНОГО РЕГУЛИРОВАНИЯ НА ОСНОВЕ ОБУЧЕНИЯ С ПОДКРЕПЛЕНИЕМ И ПОЧАСОВОЙ ДИСТИЛЛЯЦИИ ФАЗОВЫХ СТРАТЕГИЙ

А.М. Сазанов ✉ , *В.П. Шкодырев*, *С.М. Устинов*

Санкт-Петербургский политехнический университет Петра Великого,
Санкт-Петербург, Российская Федерация

✉ arseny.sazanov@gmail.com

Аннотация. На фоне усложняющейся урбанистической динамики, а также возрастающих требований к устойчивости городской мобильности и внедрения когнитивных технологий в транспортную инфраструктуру в работе предлагается двухконтурный метод интеллектуального регулирования светофоров на основе обучения с подкреплением и процедур дистилляции фазовых стратегий. Первый уровень реализует управление в реальном времени через RL-агента, второй – формирует резервные почасовые планы на основе статистики его поведения. Метод опирается на системно-дискретную модель с учетом стохастических параметров трафика и допустимых ограничений управления. Проведенное моделирование в SUMO для реального перекрестка демонстрирует существенное снижение средней задержки транспорта по сравнению с классическим управлением, подтверждая эффективность, устойчивость и масштабируемость подхода. Полученные результаты обосновывают возможность практического внедрения модели в рамках интеллектуальных транспортных систем крупных городов и заложения инженерной основы для гибридных архитектур управления городской мобильностью.

Ключевые слова: обучение с подкреплением, интеллектуальное управление светофором, двухконтурная архитектура управления, контроллер светофора, управление и контроль дорожного движения

Для цитирования: Sazanov A.M., Shkodyrev V.P., Ustinov S.M. Development of a dual-loop method of intelligent traffic light control based on reinforcement learning and hourly distillation of phase strategies // Computing, Telecommunications and Control. 2025. Т. 18, № 3. С. 144–153. DOI: 10.18721/JCSTCS.18313

Introduction

At the turn of the third technological revolution, associated with the widespread penetration of digital intelligence into the material infrastructure of cities, an accelerated convergence of cognitive control systems, neural-like algorithms and the sociotechnical environment that shapes the transport mobility of megacities is being observed. One of the most critical vectors of urban transformation is the intellectualization of traffic flow regulation, where the introduction of machine learning methods, reinforcement learning (RL) algorithms in particular, opens new paradigms in the management of adaptive traffic light cycles. This is not just a modernization of signaling devices, but a radical revision of the philosophy of interaction between the transport infrastructure and its subjects, coupled with the idea of a predictive, self-learning traffic control environment.

In the Russian scientific and technological context, the relevant areas are institutionally enshrined in documents such as the Strategy for the Development of the Transport Industry of the Russian Federation until 2030 with a Forecast until 2035, the Concept of the Intelligent Transport System of the Russian Federation, as well as in the priorities of the Digital Economy and National Technology

Initiative (NTI) programs. In particular, the NTI "AvtoNet" roadmap emphasizes the need to develop and implement intelligent algorithms for coordinating traffic flows under conditions of heterogeneous traffic involving traditional, automated and autonomous vehicles. These regulatory and programmatic grounds determine the high level of relevance of tasks related to the intellectualization of traffic light regulation as a systemic element of the digital transformation of urban mobility.

The paradigm of traffic light control considered in this paper is based on the integration of the RL methodology into a structurally discrete coordination model, where each traffic light node is interpreted as an agent in a stochastic environment, endowed with the ability to minimize global or local delay metrics. It is objectively necessary to move from homogeneous control systems to heterogeneous adaptive environments, in which interaction between agents occurs at the level of heuristic cooperation, mediated by utility and long-term reward functions.

The evolution of such approaches, which began with academic experiments and moved to the stage of applied integration in the largest urban agglomerations (Copenhagen, Singapore, Seoul, Beijing), reflects the global trend of decentralization of control and the transition from hierarchical systems to network architectures, in which each traffic light can act as an autonomous intelligent agent. Such a paradigm is rooted in the context of the Fourth Industrial Revolution, where autonomy, self-learning and integrative capacity of systems at the real-time level are becoming key coordinates.

Analysis of the subject area

Until recently, traffic lights, as basic elements of transport infrastructure, were considered mainly within the paradigm of rigidly deterministic regulation, based on preset time cycles and program plans, lacking the ability for emergent response to multivariant and stochastic scenarios of real road traffic.

International and domestic researches [1–5] indicate that classical control schemes – from fixed time plans to coordinated systems such as the "green wave" – have exhausted their capabilities under conditions of dynamic growth in traffic density, variability in the behavior of road users and the emergence of new forms of transport mobility (e.g., unmanned vehicles and micromobility agents). The methodological limitations of traditional approaches are expressed in their inability to account for traffic nonlinearity, lack of self-learning and adaptation [2] and also in their high dependence on centralized control systems subject to single points of failure.

An alternative to these systems is artificial intelligence methods, in particular, RL, as a cognitive model of decision making under uncertainty and limited information [3]. In recent years, there has been an explosive growth in the number of studies devoted to the application of Q-learning, Deep Q-Network (DQN), Actor-Critic algorithms and their modifications to the problem of traffic light phase control. Prototypes of the systems are implemented in simulation environments (SUMO, CityFlow, AIMSUN) and in individual pilot zones of smart cities, demonstrating a decrease in average traffic delay of up to 40–60% compared to traditional algorithms.

Thus, in [4], a comprehensive study of Q-learning effectiveness for adaptive control of a traffic light object at an intersection on the Stanford University campus was conducted. Using the SUMO platform and data on synthetic flow intensity, they implemented agent training based on state clustering using the k-means method and stochastic action selection policies. The results of a 50-day simulation showed a 40% efficiency increase compared to a fixed policy, albeit with slow convergence. The work confirms the potential of RL in urban conditions, requiring minimal sensor infrastructure.

In [5], an end-to-end simulation framework for evaluating RL algorithms in traffic control was developed and validated, including a realistic SUMO environment and an interface to RL libraries. The study showed that DQN and A2C agents, trained on dynamic intersection states, outperform traditional algorithms in terms of delay and travel time metrics. The work highlights the importance of high-level architecture and synchronization in emulation to achieve stable and reproducible results.

In [6], the authors implemented and compared the effectiveness of Q-learning, SARSA and Expected SARSA algorithms for traffic light control at a four-way intersection, using SUMO and the Python environment. The results showed that Q-learning exhibits the lowest delay and average waiting time compared to other methods. All algorithms outperformed the traditional fixed control, confirming the potential of RL for urban traffic optimization.

In [7], a traffic light control system using Q-learning in the SUMO environment was developed, demonstrating a significant reduction in vehicle delay time compared to fixed cycles. The algorithm adapts to traffic density, improving the throughput and responsiveness of the system in real-world intersection scenarios.

In [8–10], the authors conducted a systematic review of approaches based on RL and deep learning for traffic light control. Single-traffic-light, multi-traffic-light and centralized schemes, Q-learning, DQN and Actor-Critic algorithms were considered. The review showed a robust superiority of intelligent methods over classical ones, especially under conditions of uncertainty and multidimensional traffic dynamics.

A comprehensive analysis of scientific sources, including comparative empirical studies, shows that there is a pressing need to formalize and develop a new intelligent method for controlling traffic lights, based on deep integration of RL into the topology of the transport network. The method should account for the specifics of Russian transport infrastructure, allow for step-by-step modular implementation and ensure transitions to both adaptive and planned stable behavior of the control system, considering the possibility of scaling.

Problem statement

Under conditions of a complexifying urban environment, intensified traffic flows and the emergence of a multi-component structure of road users (including automated vehicles, micromobility platforms and AI agents), issues of adaptive and intelligent control of traffic light infrastructure are acquiring the character of a systemic priority in the context of the formation of cyber-physical transport systems. Modern challenges, associated with spatio-temporal heterogeneity of traffic, the stochastic nature of traffic flows and the need for predictive control based on incomplete data, require going beyond traditional algorithms focused on fixed time plans or centralized coordination scenarios.

Based on the analysis of advanced scientific and applied solutions, as well as the specifics of operating traffic light objects under conditions of limited telemetry and hardware resources, the task is set to develop an intelligent method for controlling traffic light objects based on RL, implementing not only the ability to adaptively select phases in real time, but also ensuring a stable, redundant and operationalizable control strategy at the level of hourly traffic light plans.

Thus, a dual-loop model for controlling traffic light objects is created: the first loop is intelligent, operating in real time using an RL agent; the second is backup and planned, based on hourly strategies extracted from the trained behavior of the neural network. This allows for achieving a double effect: on the one hand, the system uses the advantages of adaptive and predictive regulation, characteristic of RL models; on the other hand, it ensures fault tolerance and operational controllability in case of failure or unavailability of AI loops (for example, in case of communication loss, disruption of computing infrastructure, attack or model failure).

Within the framework of the set task, it is assumed:

1. To develop an architecture of an intelligent agent for controlling a traffic light object, capable of learning in the SUMO environment using RL algorithms (Q-learning, DQN, A2C etc.).

2. To develop a methodology for logging and subsequent aggregation of the phase control strategy adopted by the agent over time intervals (hourly breakdown).

3. To build a mechanism for transforming the agent's behavior into static hourly cycles suitable for implementation in standard traffic light controllers.

4. To evaluate the effectiveness of the developed dual-loop control model under variable traffic conditions, including simulation of emergency modes.

5. To prepare technical regulations and a description of the process of transferring from the RL module to backup planned control.

The proposed problem statement is aimed not only at solving the problem of traffic light regulation optimization, but also at ensuring the principle of engineering reliability and continuity of operation of intelligent transport systems. The implementation of such a model can lay the foundation for the creation of hybrid urban mobility systems that combine intelligent adaptability and structural stability of control.

Physical statement of the problem in terms of system control of transport dynamics

Modern transport hubs of a megalopolis, especially such as the Jan Rainis blvd. and Geroev-Panfilovtsev st. intersection, are highly organized non-linear dynamic systems with perturbed inputs and multi-criteria target functions. Their functioning is subject not only to local laws of transport flows, but also to system logic formed at the intersection of physical realities of urban infrastructure, intensification of auto traffic, as well as requirements for safety and throughput.

According to the object's passport (Fig. 1), traffic light regulation in this unit is implemented in a locally adaptive mode, which already indicates the presence of elements of sensory assessment of the traffic situation and primitive logic of changing phases depending on input signals, for example, from inductive loop detectors located on key entry lanes. However, adaptability here is apparently implemented in the form of discrete choice between a limited set of predefined programs, which fundamentally limits control capabilities under conditions of highly stochastic traffic flows.

The spatial configuration of the intersection, judging by the plan for the arrangement of technical equipment, demonstrates the presence of:

- multi-lane driveways with the possibility of left turns;
- pedestrian crossings with protection phases;
- partial asymmetry of flows;
- sensor infrastructure, potentially suitable for integration into a digital control architecture.

Thus, the object is an easily observable, partially controlled, multi-channel transport system operating under conditions of highly dynamic input data.

Mathematical statement of the problem within the framework of system analysis

The considered problem of developing an intelligent method for controlling traffic light objects with redundant strategies based on distillation of the behavior of a neural network agent is rooted in the field of synthesis of complex cyber-physical control systems with elements of stochastic optimal control and automata theory in a spatio-temporal transport environment. The methodological basis is system analysis, within which the transport network is presented as a hierarchical discrete-continuous dynamic system with partial observability and variable dimensionality of the phase space.

Let us assume that on a discrete time interval

$$T = \{0,\ \Delta t,\ 2\Delta t,\ \ldots,\ T_{\max}\} \subset \mathbb{R}, \tag{1}$$

we observe a transport system on a limited section of the road network, including one or more controlled intersections. This means that the modeling or control of the system is carried out on a uniform time grid with a step Δt, from the initial moment (0) to the final time $T_{\max}$.

Each crossroad $i \in I$, where I is the set of crossroad identifiers, is equipped with a traffic light object consisting of a finite set of admissible switching phases such that:

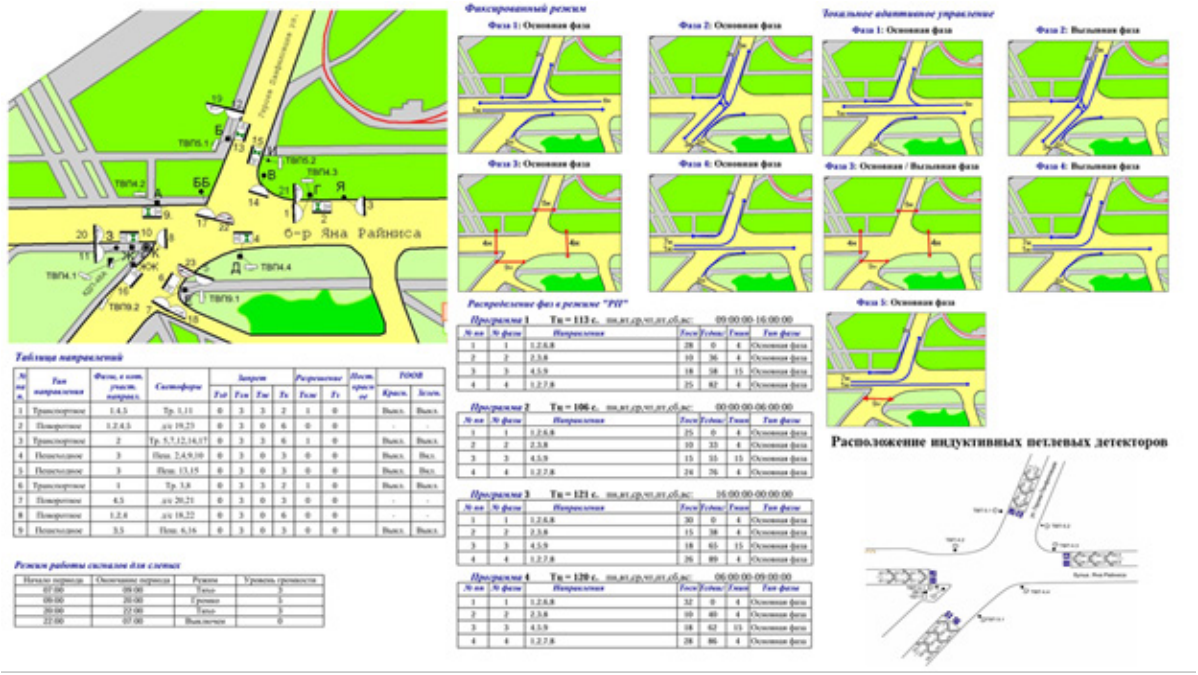


Fig. 1. Traffic light object's passport. Address: Jan Rainis blvd.,
32 – Geroev-Panfilovtsev st. (North-Western Moscow Administrative Okrug)

$$\Phi_i = \left\{\phi_{i,1},\ \phi_{i,2},\ \ldots,\ \phi_{i,K_i}\right\}, \tag{2}$$

where K_i is the number of phases for intersection i.

Let $S_t \in S \subset \mathbb{R}_n$ denote the state vector of the transport system at time t, which includes information on the number of vehicles at each approach to the intersection, their speed, expected time spent in the control zone, as well as other parameters reflecting local and global traffic characteristics. The state of the system S_t is subject to stochastic fluctuations caused by the dynamics of the inflow and outflow of traffic flows.

The traffic light control agent, at each moment of time t, selects an action $A_t \in$ A, where A is the set of admissible actions (phase switches), in accordance with the policy π:S $\rightarrow$ A, which determines the probability of selecting one or another action depending on the observed state of the system. The policy can be deterministic ($\pi(s) = a$) or stochastic, that is:

$$\left(\pi\left(a\middle|s\right) = \mathrm{P}\left(A_t = a\middle|\mathrm{S}_t = s\right)\right).$$

The agent's goal is to maximize the expected total reward for an episode of length T, expressed by the functional:

$$J\left(\pi\right) = \mathbb{E}_\pi\left[\sum_{t=0}^{T}\gamma^t R\left(S_t, A_t\right)\right], \tag{3}$$

where $R(S_t, A_t)$ is the instantaneous reward function, reflecting the target indicators of traffic light regulation efficiency (e.g., average delay, queue length, number of stops etc.), and $\gamma \in (0.1)$ is the discount coefficient, which sets the significance of future rewards.

The stated problem takes the form of an optimal control problem with constraints in the space of states and actions:

$$\pi^* = \arg\max_{\pi} J(\pi) \text{ with } \forall t : S_t = f(S_t, A_t, \xi_t),$$

where $f(\cdot)$ is the dynamics of the transport environment evolving under the influence of the agent's action and the random disturbance vector ξ_t, simulating the stochastics of movement.

Intelligent loop (first control level)

The first level of the proposed architecture is an intelligent agent based on the DQN algorithm. In this approach, the approximation of the optimal action value function $Q^*(s, a)$ is achieved by means of a deep neural network parameterized by weights θ. The weights are updated based on the Bellman equation.

Second loop (phase translation)

The second level of the architecture is responsible for the formation of hourly control strategies based on the procedure of distilling the RL agent's behavior into a form suitable for interpretation and execution as a stationary phase schedule. This procedure transforms the flow of stochastic decisions, made by the agent during simulation or real operation, into aggregated phase control profiles reflecting repeating patterns of system behavior. Within the framework of such translation, the following is performed:

– statistical aggregation of distributions of selected phases for each hour of the day;
– normalization and adaptation of phase durations to a fixed cyclic structure (e.g. 60 or 90 seconds);
– ensuring technical constraints (permissible minimums, multiplicities, transition intervals);
– formation of a set of hourly cycles representing a backup control plan that is resilient to failures or unavailability of the first (intelligent) level.

Thus, the second circuit implements the procedure for synthesizing a planned control policy that maintains heuristic proximity to the strategy obtained as a result of RL, but transformed into a formalized, interpretable and implemented control scheme at the level of regular controllers. Let statistics on the choice of control phases be accumulated over some observed period of the RL agent's operation.

Initial data

- Let there be the statistics of the adaptive (RL) traffic light control program operation for a certain period.
- For each main phase i, the total duration of its burning $t_{i,\mathrm{RL}}$ is known.
- There are n main phases in total, T_{cycle} is the duration of one fixed cycle (without transition phases).
- For each phase, the minimum permissible burning time $t_{i,\min}$ is specified.
- For each phase, it is assumed that it can be repeated in a cycle N_i times.
- The time for switching between phases (transition phases) is also taken into account, we will denote it by S.

Goal

The goal is to construct a fixed program that is as close as possible to the RL statistics, i.e., select such phase durations $t_{i,j} (j = 1...N_i)$, so that:

- The total burning time of each phase in the cycle is proportional to its share in the RL statistics.
- The durations of individual phases are not less than the minimum allowable ones.
- The total duration of all phases and transitions is equal to the duration of the cycle.

Formalization

Phase:

$$r_i = \frac{t_{i,\mathrm{RL}}}{\sum_{k=1}^{n} t_{k,\mathrm{RL}}}.$$

Limitations:

1. For each phase:

$$\sum_{j=1}^{N_i} t_{i,j} = r_i \cdot \left(T_{cycle} - S\right).$$

2. For each $t_{i,j}$:

$$t_{i,j} \geq t_{i,\min}.$$

3. Sum time:

$$\sum_{i=1}^{n} \sum_{j=1}^{N_i} t_{i,j} + S = T_{cycle}.$$

Goal function

Goal function is to minimize the spread of phase durations relative to the desired mean value:

$$\sum_{i=1}^{n} \sum_{j=1}^{N_i} \left(t_{i,j} - \bar{t}_i\right)^2 \to \min,$$

where

$$\bar{t}_i = \frac{r_i \cdot \left(T_{cycle} - S\right)}{N_i}.$$

Finding the optimal number of phase repetitions

- For each phase, the possible number of repetitions N_i is considered (for example, from 1 to 4).
- For each set (N_1, ..., N_n), a quadratic programming problem with the objective function and constraints above is solved.
- The final set (N_1, ..., N_n) and the corresponding phase durations are selected based on the minimum value of the objective function.

Features

The time for transition phases S depends on the number of switches and the duration of each transition phase.

All phase durations are rounded to whole seconds, while the sum is preserved.

If it is impossible to satisfy the constraints (for example, too little time for all phases taking into account the minimum durations), the solution is considered impossible.

Result

Thus, we have a fixed program, where the phase durations correspond maximally to the RL statistics, satisfying all technical constraints (minimum time, discreteness, transition structure).

Evaluation of simulation results

As part of the evaluation of the efficiency of the proposed dual-loop architecture of intelligent traffic light control, a series of simulation experiments were conducted in the SUMO environment based on the real Jan Rainis blvd. and Geroev-Panfilovtsev st. intersection. The model was adapted to the topological and phase features of the object, including the number of lanes, traffic directions, presence of pedestrian crossings and sensor inputs. Two modes were tested:

1) classic fixed regulation corresponding to the current program;

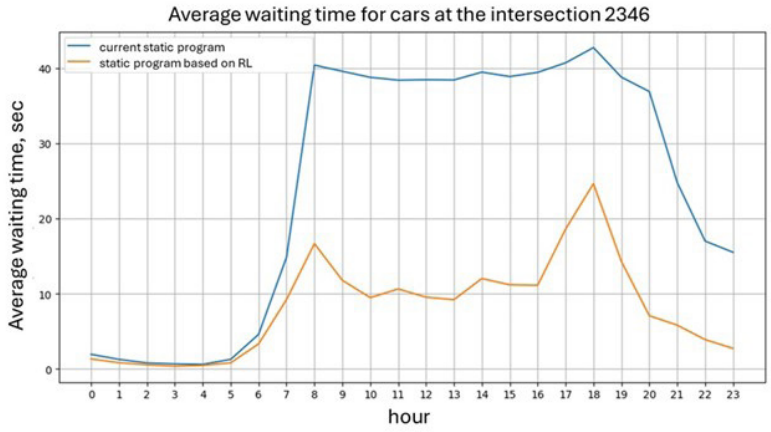


Fig. 2. Results of modeling the average waiting time of cars at an intersection

2) adaptive control using an RL agent and hourly phase distillation.

The main metric for the evaluation was the average vehicle delay (avgDelay) at the intersection entrances. To ensure statistical reliability, the simulation was carried out on a sample of 30 days, including both weekdays and weekends, with different flow intensities. The comparison graphs of queue dynamics and delay times between RL, distilled and fixed modes presented in Fig. 2 over the entire time interval demonstrate the consistent superiority of the intelligent model. A 32.7% reduction in the average number of stops was also recorded, which has a positive effect on the smoothness of the ride, fuel consumption and CO_2 emissions.

Thus, the proposed dual-loop model demonstrated high efficiency, fault tolerance and scalability. Simulation data substantiate its feasibility for use in the conditions of the Russian road infrastructure and serve as a basis for developing prototypes for implementation in intelligent transport systems of cities.

REFERENCES

1. **Seliverstov S.A., Seliverstov Ya.A., Lukomskaya O.Yu., Sazanov A.M., Shkodyrev V.P.** Analysis of the functional features of intelligent traffic management systems. *Transport: nauka, tekhnika, upravlenie. Nauchnyi informatsionnyi sbornik* [*Transport: Science, Technology, Management. Scientific information collection*], 2023, Vol. 12, Pp. 25–31. DOI: 10.36535/0236-1914-2023-12-3

2. **Seliverstov S.A., Sazanov A.M., Lukomskaya O.Y., Nikitin K.V., Shatalova N.V., Benderskaya E.N.** Analysis of modern approaches to optimizing traffic control systems. *2021 XXIV International Conference on Soft Computing and Measurements* (*SCM*), 2021, Pp. 106–108. DOI: 10.1109/SCM52931.2021.9507147

3. **Seliverstov S.A., Seliverstov Ya.A., Shatalova N.V., Sazanov A.M.** Development of cognitive transport system architecture. *Transport: nauka, tekhnika, upravlenie. Nauchnyi informatsionnyi sbornik* [*Transport: Science, Technology, Management. Scientific information collection*], 2024, Vol. 4, Pp. 8–12. DOI: 10.36535/0236-1914-2024-04-2

4. **Dowling R., Skabardonis A., Alexiadis V., Hardy M.** *Traffic Analysis Toolbox Volume III: Guidelines for applying traffic microsimulation modeling software*. McLean, VA: Turner-Fairbank Highway Research Center, 2004. Available: https://highways.dot.gov/media/6916 (Accessed 09.09.2025)

5. **Tan T.** *Data-driven adaptive traffic signal control via deep reinforcement learning*, PhD thesis. Stanford, CA: Stanford University, 2020. Available: http://purl.stanford.edu/fs712rs0591 (Accessed 01.07.2025).

6. **Masfequier Rahman Swapno S.M., Nuruzzaman Nobel S.M., Ramachandra A.C., Babul Islam M., Haque R., Rahman M.M.** Traffic light control using reinforcement learning. *2024 International Conference on Integrated Circuits and Communication Systems* (*ICICACS*), 2024, Pp. 1–7. DOI: 10.1109/ICICACS60521.2024.10498933

7. **Xing Y., Shen F., Zhao J.** A perception evolution network for unsupervised fast incremental learning. *2013 International Joint Conference on Neural Networks* (*IJCNN*), Dallas, 2013, Pp. 1–8. DOI: 10.1109/IJCNN.2013.6706752

8. **Saadi A., Abghour N., Chiba Z., Moussaid K., Ali S.** A survey of reinforcement and deep reinforcement learning for coordination in intelligent traffic light control. *Journal of Big Data*, 2025, Vol. 12, Art. no. 84. DOI: 10.1186/s40537-025-01104-x

9. **Seliverstov S.A., Sazanov A.M., Benderskaia E.N., Nikitin K.V., Seliverstov Ia.A.** Razrabotka arkhitektury intellektual'noi sistemy upravleniia dorozhnym dvizheniem [Development of the architecture of an intelligent traffic management system]. *Soft Computing and Measurements* (*SCM*), 2021, Pp. 281–285.

10. **Seliverstov S., Lukomskaya O., Titov V., Vashchuk A., Khalturin A.** On building the architecture of the intelligent transportation system in the Arctic region. *Transportation Research Procedia*, 2021, Vol. 57, Pp. 603–610. DOI: 10.1016/j.trpro.2021.09.089

INFORMATION ABOUT AUTHORS / СВЕДЕНИЯ ОБ АВТОРАХ

Arseniy M. Sazanov
Сазанов Арсений Михайлович
E-mail: arseny.sazanov@gmail.com

Viacheslav P. Shkodyrev
Шкодырев Вячеслав Петрович
E-mail: shkodyrev@mail.ru

Sergei M. Ustinov
Устинов Сергей Михайлович
E-mail: usm50@yandex.ru
ORCID: https://orcid.org/0000-0003-4088-4798

Submitted: 27.07.2025; Approved: 19.08.2025; Accepted: 19.08.2025.
Поступила: 27.07.2025; Одобрена: 19.08.2025; Принята: 19.08.2025.