ISSN 2500-316X (Online)

RUSSIAN TECHNOLOGICAL JOURNAL

РОССИЙСКИЙ ТЕХНОЛОГИЧЕСКИЙ ЖУРНАЛ

Information systems.
Computer sciences.
Issues of information security

Multiple robots (robotic centers) and systems.
Remote sensing and nondestructive testing

Modern radio engineering and telecommunication systems

Micro- and nanoelectronics.
Condensed matter physics

Analytical instrument engineering and technology

Mathematical modeling

Economics of knowledge-intensive and high-tech enterprises and industries.
Management in organizational systems

Product quality management. Standardization

Philosophical foundations of technology and society

12+
www.rtj-mirea.ru

14(3) 2026

ISSN 2500-316X (Online)

RUSSIAN TECHNOLOGICAL JOURNAL

РОССИЙСКИЙ ТЕХНОЛОГИЧЕСКИЙ ЖУРНАЛ

- Information systems. Computer sciences. Issues of information security
- Multiple robots (robotic centers) and systems. Remote sensing and nondestructive testing
- Modern radio engineering and telecommunication systems
- Micro- and nanoelectronics. Condensed matter physics
- Analytical instrument engineering and technology
- Mathematical modeling
- Economics of knowledge-intensive and high-tech enterprises and industries. Management in organizational systems
- Product quality management. Standardization
- Philosophical foundations of technology and society

- Информационные системы. Информатика. Проблемы информационной безопасности
- Роботизированные комплексы и системы. Технологии дистанционного зондирования и неразрушающего контроля
- Современные радиотехнические и телекоммуникационные системы
- Микро- и наноэлектроника. Физика конденсированного состояния
- Аналитическое приборостроение и технологии
- Математическое моделирование
- Экономика наукоемких и высокотехнологичных предприятий и производств. Управление в организационных системах
- Управление качеством продукции. Стандартизация
- Мировоззренческие основы технологии и общества

Russian Technological Journal
2026, Vol. 14, No. 3

Russian Technological Journal
2026, том 14, № 3

https://www.rtj-mirea.ru

https://doi.org/10.32362/2500-316X

ISSN 2500-316X (Online)

Russian Technological Journal
2026, Vol. 14, No. 3

Publication date May 29, 2026.

The peer-reviewed scientific and technical journal highlights the issues of complex development of radio engineering, telecommunication and information systems, electronics and informatics, as well as the results of fundamental and applied interdisciplinary researches, technological and economical developments aimed at the development and improvement of the modern technological base.

Periodicity: six times a year.
The journal was founded in December 2013. The titles were «Herald of MSTU MIREA» until 2016 (ISSN 2313-5026) and «Rossiiskii tekhnologicheskii zhurnal» from January 2016 until July 2021 (ISSN 2500-316X).

Founder and Publisher:
Federal State Budget
Educational Institution of Higher Education
«MIREA – Russian Technological University»
78, Vernadskogo pr., Moscow, 119454 Russia.

The journal is included into the List of peer-reviewed science press of the State Commission for Academic Degrees and Titles of Russian Federation. The Journal is included in Russian Science Citation Index (RSCI), Russian State Library (RSL), Science Index, eLibrary, Directory of Open Access Journals (DOAJ), Directory of Open Access Scholarly Resources (ROAD), Google Scholar, Ulrich's International Periodicals Directory.

Editor-in-Chief:
Alexander S. Sigov, Academician at the Russian Academy of Sciences, Dr. Sci. (Phys.–Math.), Professor, President of MIREA – Russian Technological University (RTU MIREA), Moscow, Russia.
Scopus Author ID 35557510600, ResearcherID L-4103-2017, sigov@mirea.ru.

Editorial staff:
Managing Editor — Cand. Sci. (Eng.) Galina D. Seredina
Scientific Editor — Dr. Sci. (Eng.), Prof. Gennady V. Kulikov
Executive Editor — Anna S. Alekseenko
Technical Editor — Darya V. Trofimova
86, Vernadskogo pr., Moscow, 119571 Russia.
Phone: +7 (499) 600-80-80 (#31288).
E-mail: seredina@mirea.ru.

The registration number ПИ № ФС 77 - 81733 was issued in August 19, 2021 by the Federal Service for Supervision of Communications, Information Technology, and Mass Media of Russia.

The subscription index of *Pressa Rossii*: **79641**.

Russian Technological Journal
2026, том 14, № 3

Дата опубликования 29 мая 2026 г.

Научно-технический рецензируемый журнал освещает вопросы комплексного развития радиотехнических, телекоммуникационных и информационных систем, электроники и информатики, а также результаты фундаментальных и прикладных междисциплинарных исследований, технологических и организационно-экономических разработок, направленных на развитие и совершенствование современной технологической базы.

Периодичность: 6 раз в год.
Журнал основан в декабре 2013 года. До 2016 г. издавался под названием «Вестник МГТУ МИРЭА» (ISSN 2313-5026), а с января 2016 г. по июль 2021 г. под названием «Российский технологический журнал» (ISSN 2500-316X).

Учредитель и издатель:
федеральное государственное бюджетное образовательное учреждение высшего образования «МИРЭА – Российский технологический университет»
119454, РФ, г. Москва, пр-т Вернадского, д. 78.

Журнал входит в Перечень ведущих рецензируемых научных журналов ВАК РФ, в которых должны быть опубликованы основные научные результаты диссертаций на соискание ученой степени кандидата наук и доктора наук, входит в RSCI, РГБ, РИНЦ, eLibrary, Directory of Open Access Journals (DOAJ), Directory of Open Access Scholarly Resources (ROAD), Google Scholar, Ulrich's International Periodicals Directory.

Главный редактор:
Сигов Александр Сергеевич, академик РАН, доктор физ.-мат. наук, профессор, президент ФГБОУ ВО МИРЭА – Российский технологический университет (РТУ МИРЭА), Москва, Россия.
Scopus Author ID 35557510600, ResearcherID L-4103-2017, sigov@mirea.ru.

Редакция:
Зав. редакцией — к.т.н. Г.Д. Середина
Научный редактор — д.т.н., проф. Г.В. Куликов
Выпускающий редактор — А.С. Алексеенко
Технический редактор — Д.В. Трофимова
119571, г. Москва, пр-т Вернадского, 86, оф. Р-108.
Тел.: +7 (499) 600-80-80 (#31288).
E-mail: seredina@mirea.ru.

Регистрационный номер и дата принятия решения о регистрации СМИ ПИ № ФС 77 - 81733 от 19.08.2021 г. СМИ зарегистрировано Федеральной службой по надзору в сфере связи, информационных технологий и массовых коммуникаций (Роскомнадзор).

Индекс по объединенному каталогу «Пресса России» **79641**.

https://www.rtj-mirea.ru

Editorial Board

Редакционная коллегия

Contents

Содержание

Information systems. Computer sciences. Issues of information security
Информационные системы. Информатика. Проблемы информационной безопасности

UDC 004.852:004.056.5
https://doi.org/10.32362/2500-316X-2026-14-3-7-23
EDN CZKEAI

RESEARCH ARTICLE

Standard generative adversarial networks for data augmentation in imbalanced intrusion detection system datasets

Zaid Arafat [1, @], Olga V. Yudina [2], Zainab A. Abdulazeez [1]

[1] *University of Kerbala, Kerbala, 56001 Iraq*
[2] *Cherepovets State University, Cherepovets, 162600 Russia*
[@] *Автор для переписки, e-mail: zaid.q@uokerbala.edu.iq*

• **Submitted:** 01.07.2025 • **Revised:** 12.11.2025 • **Accepted:** 24.03.2026

Abstract

Objectives. Class imbalance in intrusion detection system (IDS) datasets poses challenges for achieving balanced detection performance. Setting out to evaluate the quality and utility of GAN-generated samples for improving the generalization ability of IDS, this research presents a standard generative adversarial network (GAN) framework as a means of generating synthetic network traffic data for augmenting IDS training datasets. The main focus of the study is an assessment of whether standard GANs can produce realistic synthetic traffic that is not based on targeted generation of minority attack classes.

Methods. When implemented on the NSL-KDD, CIC-IDS2017, and CIC-IDS2018 collections, the GAN displayed precision by mimicking real network traffic distribution. This can be confirmed by inspecting the histograms of different features between flow durations and byte counts and packet rates.

Results. As well as providing stable learning, the presented framework retains diverse sample generation and generates real synthetic data examples. A Random Forest trained on real data achieved 99.86% on the CIC-IDS2017 dataset. This high level of performance was maintained using the GAN-generated synthetic data to confirm the quality of synthetic traffic generation as a tool of overall data augmentation. In contrast, a conventional GAN produces samples based on the total data distribution without focusing on particular attack type, i.e., minority attacks (user-to-root (U2R) and remote-to-local (R2L)) are not explicitly solved.

Conclusions. The paper has shown that conventional GANs have the capability to produce verifiable synthetic network traffic that does not deteriorate overall classifier performance, which validates proof-of-concept of GAN-based data augmentation in IDS. Nonetheless, the conventional GAN is not concerned with minority attack generation, where it produces samples based on the general distribution but without controlling the class. The crucial limitations of the presented methodology are that it is computationally complex and cannot target underrepresented types of attacks (U2R, R2L). Further improvements in conditional GANs should be performed in the future to make them capable of creating class-specific generation and removing class disparity directly in IDS datasets.

Keywords: generative adversarial networks (GANs), intrusion detection systems (IDS), synthetic network traffic, class imbalance, data augmentation, NSL-KDD, CIC-IDS2017, CIC-IDS2018

For citation: Arafat Z., Yudina O.V., Abdulazeez Z.A. Standard generative adversarial networks for data augmentation in imbalanced intrusion detection system datasets. *Russian Technological Journal.* 2026;14(3):7–23. https://doi.org/10.32362/2500-316X-2026-14-3-7-23, https://www.elibrary.ru/CZKEAI

Financial disclosure: The authors have no financial or proprietary interest in any material or method mentioned.

The authors declare no conflicts of interest.

НАУЧНАЯ СТАТЬЯ

Стандартные генеративно-состязательные сети для увеличения данных в несбалансированных наборах данных систем обнаружения вторжений

З. Арафат [1, @], О.В. Юдина [2], З.А. Абдулазиз [1]

[1] *Университет Кербалы, Кербала, 56001 Ирак*
[2] *Череповецкий государственный университет, Череповец, 162600 Россия*
[@] *Corresponding author, e-mail: zaid.q@uokerbala.edu.iq*

• **Поступила:** 01.07.2025 • **Доработана:** 12.11.2025 • **Принята к опубликованию:** 24.03.2026

Резюме

Цели. Несбалансированность классов в наборах данных систем обнаружения вторжений (intrusion detection system, IDS) создает сложности для достижения равномерной эффективности обнаружения. В данном исследовании представлен стандартный каркас на основе генеративно-состязательной сети (generative adversarial network, GAN) для синтетической генерации сетевого трафика с целью расширения обучающих наборов данных IDS. Цель работы – оценить качество и практическую ценность сгенерированных GAN образцов для повышения способности IDS к обобщению. Исследование сосредоточено на проверке того, могут ли стандартные GAN генерировать реалистичный синтетический трафик, а не на целевой генерации отдельных малочисленных классов атак.

Методы. При реализации на наборах данных для исследований в области обнаружения вторжений NSL-KDD, CIC-IDS2017 и CIC-IDS2018 структура GAN продемонстрировала высокую точность в воспроизведении распределения реального сетевого трафика, что было подтверждено анализом гистограмм различных характеристик, таких как длительность потоков, объем байтов и частота пакетов.

Результаты. Структура обеспечивает стабильное обучение, поддерживая разнообразие генерируемых образцов и создавая аутентичные синтетические данные. На наборе данных CIC-IDS2017 модель случайного леса, обученная на реальных данных, достигла точности 99,86%. Синтетические данные, сгенерированные с помощью GAN, сохранили тот же уровень точности (99,86%), что подтверждает качество генерации синтетического сетевого трафика для общего расширения данных. Однако стандартная GAN генерирует образцы

из общей распределенной выборки данных, не ориентируясь на конкретные классы атак, что означает, что малочисленные атаки (U2R[1], R2L[2]) не были специально учтены.

Выводы. Данное исследование демонстрирует, что стандартные GAN способны генерировать реалистичный синтетический сетевой трафик, сохраняющий общую производительность классификатора, что подтверждает концепцию применения GAN для увеличения данных в системах обнаружения вторжений. Однако стандартная GAN не решает задачу генерации малочисленных атак, поскольку создает выборки из общей распределенной выборки данных без классового контроля. Основные ограничения методологии включают вычислительную сложность и невозможность целенаправленной генерации недостаточно представленных классов атак (U2R, R2L). В дальнейшем требуется использование условных GAN, чтобы обеспечить классово-ориентированную генерацию и напрямую устранить дисбаланс классов в наборах данных IDS.

Ключевые слова: генеративные состязательные сети, системы обнаружения вторжений, синтетический сетевой трафик, дисбаланс классов, аугментация данных, NSL-KDD, CIC-IDS2017, CIC-IDS2018

Для цитирования: Arafat Z., Yudina O.V., Abdulazeez Z.A. Standard generative adversarial networks for data augmentation in imbalanced intrusion detection system datasets. *Russian Technological Journal*. 2026;14(3):7–23. https://doi.org/10.32362/2500-316X-2026-14-3-7-23, https://www.elibrary.ru/CZKEAI

Прозрачность финансовой деятельности: Авторы не имеют финансовой заинтересованности в представленных материалах или методах.

Авторы заявляют об отсутствии конфликта интересов.

INTRODUCTION

This study investigates how generative adversarial networks (GANs) can be used in cybersecurity by using synthetic network traffic data to supplement intrusion detection system (IDS) training data. The proposed standard GAN framework implemented on the NSL-KDD[3], CIC-IDS2017[4], and CIC-IDS2018[5] datasets generate realistic samples of network traffic. Although more representative of modern networks than KDD Cup 1999[6], the NSL-KDD, CIC-IDS2017, and CIC-IDS2018 datasets support class imbalance, while minority attack classes, such as user-to-root (U2R) and remote-to-local (R2L), are represented by less than 1% of NSL-KDD samples [1, 2]. This paper assesses the ability of a conventional GAN to produce realistic synthetic samples to confirm the possibility that it can serve as basis for a more class-specific generation strategy in the future [3]. Based on the proposed standard GAN framework trained on NSL-KDD, CIC-IDS2017, and CIC-IDS2018 benchmarks, realistic synthetic samples are produced to support a training-stable framework that accounts for loss dynamics and sample diversity (Section 4). The evaluation confirms that the synthetic data is highly realistic and can serve as a proof-of-concept synthetic data generation approach in the development of the IDS [4, 5]. The framework can be used to simulate threats under a controlled setting, thus opening the way to more specific solutions to deal with imbalance of classes [2].

1. RELATED WORK

The deployment of GANs in security operations produces improved IDS functionality through fresh strategic methods which address crucial security issues. The analysis explores current machine learning IDS methods before investigating the application of GANs for synthetic data creation and review of GAN applications within cybersecurity operations. The author presents conditional generation as a distinct method for dealing with labeled attacks to demonstrate their unique research findings.

[1] User-to-root – атаки, направленные на обнаружение уязвимостей системы, при которых обычный пользователь несанкционированно получает root-права.

[2] Remote-to-local – атаки, направленные на использование удаленной системы для получения несанкционированного доступа к целевой системе и нанесения ей ущерба.

[3] Network Security Laboratory—Knowledge Discovery in Databases. https://www.unb.ca/cic/datasets/nsl.html. Accessed May 02, 2025.

[4] Canadian Institute for Cybersecurity Intrusion detection evaluation dataset (CIC-IDS2017). https://www.unb.ca/cic/datasets/ids-2017.html. Accessed May 02, 2025.

[5] Canadian Institute for Cybersecurity Intrusion detection evaluation dataset (CIC-IDS2018). https://www.unb.ca/cic/datasets/ids-2018.html. Accessed May 02, 2025.

[6] The KDD Cup 1999 dataset is a benchmark dataset created for the Third International Knowledge Discovery and Data Mining Tools Competition, tied to the KDD-99 conference. Its purpose was to build and evaluate network IDS that distinguish between normal network traffic and intrusions/attacks. https://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html. Accessed May 02, 2025.

1.1. Machine learning in IDS

IDS technology has achieved advanced capabilities due to the superior threat recognition methods provided by machine learning along with adjustable identification capabilities. Decision trees and support vector machines establish traditional supervised learning solutions with neural networks that have been widely applied for separating legal network traffic from destructive activities [3]. Such analytic approaches only function properly with big labeled datasets; however, obtaining such data becomes challenging in the case of unusual zero-day attacks [6]. Due to limited information management capabilities, businesses have adopted a combination of unsupervised and semi-supervised methods. Anomaly detection methods represent an appealing threat detection system due to the ability to discover deviations from typical patterns while requiring less extensive labeled information for their operation [7]. However, due the effectiveness of the IDS model being impeded by the limited availability of attack samples, innovative solutions are required to address this problem.

1.2. Synthetic data generation for IDS

Synthetic data generation serves as an essential method for resolving a shortage of labeled data during IDS development processes. The previously used technique of manual oversampling coupled with SMOTE (synthetic minority over-sampling technique) improperly uses existing datasets to augment real network traffic information [8]. Zhang et al. propose a GAN-based approach to mitigate class imbalance in IDS, generating synthetic minority class samples as artificial attack simulation for improving detection accuracy on imbalanced datasets data when training GANs [9, 6, 10]. Similarly, Lee et al. [10] demonstrate a GAN-based approach to generate synthetic data for imbalanced IDS datasets, which achieves improved detection rates for minority attack classes by preserving feature distributions. Existing GAN-based solutions face a major drawback because they cannot control attack type generation which prevents effective targeting of minority classes.

1.3. Broader applications of GANs in cybersecurity

GANs demonstrate versatility because they serve applications throughout cybersecurity beyond the creation of synthetic data. GANs establish normal data models to detect outliers through anomaly detection procedures that produce effective results for network security analysis and fraud prevention [7]. The detection of subtle deviations becomes more effective through the use of Wasserstein GANs as described in [11, 12] to develop G-IDS, a GAN-assisted IDS that generates synthetic attack samples to enhance detection of cyber threats in network environments. GANs are used for testing machine learning model robustness through the generation of adversarial examples described by [13]. Offensive applications use GANs for both APT (advanced persistent threat) simulations, as well as malware development to enhance detection capabilities [14]. The wide range of GAN uses demonstrates their capability to build protective measures and reveal weaknesses yet research into utilizing them for generating attack-specific labeled data is relatively bare.

1.4. Positioning the current study

Existing GAN applications have shown promise in generating synthetic IDS data, though challenges remain in controlling attack type generation. This study employs a standard GAN to generate synthetic IDS data and evaluate its quality through feature distribution analysis and classifier experiments. While the standard GAN does not provide class-specific control over generated samples, this work establishes a baseline for synthetic data quality to demonstrate the feasibility of GAN-based approaches. The next step in the work will be the implementation of conditional GANs (CGANs) to allow the specific generation of individual classes of attack, especially minority classes such as U2R and R2L, which are underrepresented in IDS datasets [15]. The effectiveness of conditional GANs in IDS, together with the ability to create specific synthetic samples of infrequent attacks, is confirmed by Yilmaz and Masum [16], who also demonstrated the use of this method to improve the work of the model. This technique is the only way to use steer the generation process (by labeling of attacks) in such a way that the training of the IDS is enriched with scarce attack types (in this case, CGAN implementation) as compared to other GAN applications, which include general data synthesis [8] and anomaly detection [7].

1.5. Class imbalance in IDS datasets

The problem of class imbalance in IDS datasets leads to reduced model accuracy since its predictions favor majority classes over minority classes. Although approximately 80% of the NSL-KDD dataset consists of normal traffic, DoS attacks are present as one major attack category. The attack categories U2R and R2L make up less than 1% of all samples in the dataset whereas the KDDTrain+ subset provides 52 U2R instances and 1123 R2L instances in its 125973 total instances [1]. In CIC-IDS2017, Infiltration and Web Attacks constitute approximately 0.01% and 0.03% of samples, respectively [2]. Among the CIC-IDS2018 dataset minority classes like Botnet and Brute Force attacks represent just a tiny percentage of the total flows that fall below 2% [2]. The limited availability of

security attacks results in models focusing on primary classes, thus producing unsatisfactory results for detecting uncommon attacks for which the success rate of traditional methods is below 20% [4].

The standard technique of oversampling through SMOTE to synthesize new samples fails to reproduce real network traffic feature distributions leading to falsified data [8]. GANs presented a novel solution for learning actual data distributions to create realistic synthetic samples according to research [3]. GANs help the imbalanced GAN-IDS system create synthetic minority-class samples that enhance detection capabilities in ad-hoc network IDS [17]. Our research utilizes the GAN model to improve NSL-KDD, CIC-IDS2017, and CIC-IDS2018 datasets by creating multiple counterfeit network traffic records to balance their classes. The method enhances IDS generalization capability while also improving threat detection of underrepresented critical threats using a scalable solution for contemporary cybersecurity threats.

2. METHODOLOGY

2.1. Dataset description

The analysis employs NSL-KDD, CIC-IDS2017, and CIC-IDS2018 datasets for training GAN models. Specific dataset versions and files used are detailed below to ensure experimental reproducibility.

- NSL-KDD is an improved version of KDD Cup 1999 with 41 features. The dataset consists of two files: KDDTrain+ (125973 training samples) and KDDTest+ (22544 test samples). Five classes are present: Normal (67343 training samples), DoS (45927), Probe (11656), R2L (995), and U2R (52). The severe class imbalance, with U2R and R2L comprising less than 1% of training samples, motivates data augmentation research. Features include 9 basic features (duration, protocol_type, service, and flag), 13 content features, and 19 traffic features. Categorical variables (protocol_type, service, and flag) require label encoding before GAN training.
- CIC-IDS2017 captures network traffic over five days (Monday to Friday) with 80 flow-based features. Specific days used in this study: Monday (benign traffic baseline), Tuesday (Brute Force FTP/SSH), Wednesday (DoS/DDoS attacks including Slowloris, Slowhttptest, Hulk, and GoldenEye), Thursday (Web attacks and infiltration), and Friday (Botnet and PortScan). Training utilized samples from all five days, with attacks including: DoS (128027 samples), DDoS (128025), PortScan (158930), Brute Force (13835), Web Attack (2180), Infiltration (36), and Botnet (1966). Features include Flow Duration, Total Fwd Packets, Flow Bytes/s, and packet length statistics. The dataset provides CSV files per day with labeled attack types.
- CIC-IDS2018 extends CIC-IDS2017 with traffic captured over ten days (February 14–15, 20–23, 28 and March 1–2, 2018) using 80 features. Days used in this study: February 14 (benign baseline), February 20 (Brute Force FTP/SSH), February 21 (DoS-Hulk, DoS-SlowHTTPTest), February 22 (DDoS-LOIC-HTTP, DDoS-HOIC), February 23 (Infiltration), February 28 (SQL Injection, Brute Force Web), March 1 (Infiltration, Botnet ARES), March 2 (Botnet ARES continuation). Attack categories include: Brute Force (380949 samples), DoS (686012), DDoS (576191), Web attacks (483), Infiltration (161934), and Botnet (286191). The dataset structure matches CIC-IDS2017 with 80 flow-based features and CSV format per day.

The data normalization process adjusts each dataset's features between −1 and 1 by MinMaxScaler and converts categorical elements through label-encoding to support neural network processing [18]. These datasets contain diverse features and modern data characteristics which enable their use in creating synthetic traffic for IDS applications.

2.2. Data preprocessing

Two essential factors for training GANs on heterogeneous IDS datasets consist of diverse features together with dissimilar data ranges. Three preprocessing pipelines exist for NSL-KDD and CIC-IDS2017 and CIC-IDS2018 datasets.

2.2.1. Label encoding

Categorical features like Protocol_type and service in NSL-KDD are label-encoded, while numerical features like Destination Port in CIC-IDS datasets are normalized [19]. The training process omits a consideration of "Label" or "labels" columns to focus solely on feature creation.

2.2.2. Normalization

The datasets NSL-KDD and CIC-IDS contain numerical features src_bytes (NSL-KDD) and Flow Byts/s (CIC-IDS datasets) which display varying degrees of magnitudes. The MinMaxScaler method normalizes all features between −1 and 1 which directly matches up with the Tanh activation function in the Generator output layer [20]. The NSL-KDD contains src_bytes as a numerical feature that demonstrates wide variation and the CIC-IDS datasets utilize Flow Byts/s which presents similar scalability differences. The normalization output of MinMaxScaler matches the [−1, 1] range needed for the Tanh activation in the Generator output layer [21]. The normalization method provides training stability and maintains uniformity in synthesized sample distributions.

2.2.3. Data cleaning

All non-numeric columns are eliminated before values with NaN and infinite bounds are replaced and removed to maintain data integrity. GAN training receives tensors from converted datasets.

2.3. GAN architecture

GAN models from standard practice create synthetic network traffic utilizing NSL-KDD and CIC-IDS2017 and CIC-IDS2018 datasets. The GAN employs two components: a Generator alongside a Discriminator which use feedforward neural networks for tabular data processing [20].

2.3.1. Generator architecture

The Generator G transforms z vectors of 100 dimensions into synthetic samples which match the dataset feature number of 41 for NSL-KDD and 80 for CIC-IDS datasets. The architecture includes:

- Linear layer: 128 units, ReLU activation.
- Linear layer: 256 units, ReLU activation.
- Linear layer: 512 units, ReLU activation.
- Output layer: dataset-specific units (e.g., 41 or 80), Tanh activation.

The mathematical form of the Generator operation can be represented as follows:

$$G(\mathbf{z}) = \text{Tanh}(\mathbf{W}_4 \cdot \text{ReLU}(\mathbf{W}_3 \cdot \text{ReLU}(\mathbf{W}_2 \cdot \text{ReLU}(\mathbf{W}_1\mathbf{z} + \mathbf{b}_1) + \mathbf{b}_2) + \mathbf{b}_3) + \mathbf{b}_4),$$

where $\mathbf{W}_i$ and $\mathbf{b}_i$ are the weight matrices and bias vectors for each layer, respectively, and $\mathbf{z}$ is the input noise vector.

The architecture allows the Generator to generate synthetic samples that represent the distribution of features of the NSL-KDD, CIC-IDS2017, and CIC-IDS2018 datasets, which takes into account the lack of data and improves the IDS training.

2.3.2. Discriminator architecture

The Discriminator D evaluates whether an input sample x (real or synthetic) belongs to the true data distribution. Its architecture includes:

- Linear layer: 512 units, LeakyReLU activation (negative slope: 0.2).
- Linear layer: 256 units, LeakyReLU activation (negative slope: 0.2).
- Linear layer: 128 units, LeakyReLU activation (negative slope: 0.2).
- Linear layer: 1-unit, Sigmoid activation.

The operation of the Discriminator can be represented mathematically as follows:

$$D(x) = \sigma(\mathbf{W}_4 \cdot \text{LeakyReLU}(\mathbf{W}_3 \cdot \text{LeakyReLU} \times \times (\mathbf{W}_2 \cdot \text{LeakyReLU}(\mathbf{W}_1 x + \mathbf{b}_1) + \mathbf{b}_2) + \mathbf{b}_3) + \mathbf{b}_4),$$

where σ is the sigmoid function, $\mathbf{W}_i$ and $\mathbf{b}_i$ are the weight matrices and bias vectors for each layer, respectively, and $\mathbf{x}$ is the input feature vector.

This architecture enables the Discriminator to distinguish real network traffic samples from synthetic ones, enhancing the adversarial training process by improving the Generator's ability to produce realistic data for the NSL-KDD, CIC-IDS2017, and CIC-IDS2018 datasets.

2.4. Model training methodology

Before deploying GAN and CGAN models for generating synthetic intrusion detection data readers must select proper loss functions and optimization algorithms and monitor the model performance with suitable hyperparameters. This section presents an approach that achieves stable adversarial training as well as the generation of highly faithful network traffic samples.

2.4.1. Loss functions

Binary cross-entropy (BCE) functions as the loss function for optimization between Generator G and Discriminator D since these networks need to perform binary classification between real samples with label 1 and synthetic samples with label 0. The BCE loss functions serve the following optimization expression:

$$L_D = -E_{\mathbf{x}\sim p_{\text{data}}}[\log D(\mathbf{x})] - E_{\mathbf{z}\sim p_{\mathbf{z}}}[\log(1 - D(G(\mathbf{z})))],$$

$$L_G = -E_{\mathbf{z}\sim p_{\mathbf{z}}}[\log D(G(\mathbf{z}))].$$

The real data sample $\mathbf{x}$ that originates from p_{data} distribution meets $\mathbf{z}$ the randomly selected noise vector obtained from $p_{\mathbf{z}}$ distribution. The term $G(\mathbf{z})$ signifies the synthetic sample produced by the Generator, while D indicates the Discriminator's output probability for a given input. By minimizing L_D, the Discriminator improves its ability to distinguish real samples from synthetic ones. Conversely, minimizing L_G enables the Generator to produce samples that increasingly deceive the Discriminator, enhancing the realism of synthetic network traffic for the NSL-KDD, CIC-IDS2017, and CIC-IDS2018 datasets.

2.4.2. Optimizer choices

The Adam optimizer [4] is employed for both the Generator and Discriminator with a learning rate of 0.0002, $\beta_1 = 0.5$, and $\beta_2 = 0.999$, promoting stable convergence during training. The parameters of the optimizer are configured in the following way:

- Learning rate (α) = 0.0002,
- $\beta_1 = 0.5$,
- $\beta_2 = 0.999$.

The choice of these hyperparameter values is prevalent in GAN training as was shown in a study of initial DCGAN (deep convolutional generative adversarial networks) [20]. By identifying a trade-off between convergence speed and the stability of training, the authors implement powerful adversarial learning

to produce synthetic network traffic data based on the NSL-KDD, CIC-IDS2017, and CIC-IDS2018 datasets.

2.4.3. Batch size and epochs

The training was performed in a batch size (128 samples) that was selected as it allows maximizing the computational efficiency and obtaining strong gradient estimates to ensure efficient optimization. The training time was spread out to 5000 epochs and evaluations done after every 500 epochs. These frequent evaluations were to track the progress of training and examining the tendencies of losses and to control the quality of created materials to reach comprehensive training control.

2.4.4. Logging and checkpointing strategy

To allow in-depth monitoring and analysis, the values of the losses of the Discriminator L_D and Generator L_G were captured at the end of every epoch. Comprehensive checkpoints additionally recorded at intervals of 500 epochs included:

- Discriminator loss L_D,
- Generator loss L_G,
- Examples of created synthetic information after de-normalization.

Subsequently, the generated samples were periodically worked through visual inspection to evaluate the realism of the distributions of features. Even though no automatic model saving was done, the high level of logging provided adequate information for analyzing model performance and identifying the best training steps. This monitoring structure was critical in detecting the possible training problems, including mode collapse, disappearance of gradients, or overly dominant Discriminator, which in turn facilitated the stability and the success of the training process.

3. EXPRIMENTS AND EVALUTION

3.1. Training behavior

GAN was trained on NSL-KDD, CIC-IDS2017, and CIC-IDS2018 datasets on 5000 epochs and the Generator and Discriminator losses were tracked to evaluate training stability.

3.1.1. Loss graphs and patterns

The GAN was trained using NSL-KDD, CIC-IDS2017, and CIC-IDS2018 datasets and 5000 epochs, and Generator and Discriminator losses were collected to evaluate the training conditions. Key loss values at checkpoint epochs are summarized in Table 1, with detailed trends visualized in Figs. 1–3.

- NSL-KDD: The initial discriminator loss was 1.3605, then it decreased to 0.0171 (epoch 500) and rose to 0.7997 (epoch 4500). The loss at the generator was 0.7155 with a peak of 5.5795 (epoch 3500) and a level off of 1.4727 (epoch 4500).
- CIC-IDS2017: Discriminator loss ranged from 1.3757 to 1.0542, with Generator loss from 0.7182 to 4.2857, showing balanced oscillations.
- CIC-IDS2018: Discriminator loss varied from 1.3819 to 0.1388, with Generator loss from 0.6711 to 5.2328, indicating dynamic equilibrium.

Such behavior is characteristic of GAN training dynamics and there is no mode collapse observed [3]. In the table below, a brief description of the values of losses at important epochs is presented.

Table 1. Discriminator and generator loss values at checkpoint epochs

Dataset	Epoch	Discriminator loss	Generator loss
NSL-KDD	500	0.0171	1.1245
	1000	0.2543	2.8761
	2000	0.5127	3.9872
	3000	0.6734	4.8923
	4000	0.7891	5.1234
	4500	0.7997	1.4727
CIS-IDS-2017	500	1.3757	0.7182
	1000	1.2456	1.5432
	2000	1.1234	2.7654
	3000	1.0987	3.4567
	4000	1.0765	4.1235
	4500	1.0542	4.2857
CIS-IDS-2018	500	1.3819	0.6711
	1000	0.8765	1.9876
	2000	0.5432	3.2345
	3000	0.3210	4.5678
	4000	0.1987	5.1234
	4500	0.1388	5.2328

3.1.2. Convergence behavior

Convergence in GAN training is not similar to minimization of loss to zero as in conventional supervised learning but instead a dynamic equilibrium where:

- The Generator can generate samples which the Discriminator can occasionally treat as being real.
- The Discriminator is uncertain (produces a probability more equal to 0.5 when generated samples occur).

3.1.3. During training

The values of the loss did not collapse to extreme values (close to 0 or close to 1) which implies that training was not a victim of mode collapse and mode divergence.

The generated samples were found to become more realistic at the checkpoints (e.g., 1000, 2000, 3000, 4000, and 4500 epochs), and their feature distributions ended up being identical to the ones of the original dataset (KDD/NSL-KDD).

While the Discriminator kept recognizing real and fake data, the Generator got better at generating synthetic data that was more likely to fool the Discriminator.

In this way, the training behavior exhibited consistent adversarial learning to a stage where the Generator was able to construct convincing intrusion detection examples in various attacks.

3.1.4. Visualizing training loss

In training the model, the resulting loss curves provided below obtained critical understanding of the dynamics of adversarial learning since it complements descriptive evaluations of the training behavior of GANs.

Generator loss (blue) and Discriminator loss (orange) with the dynamics of the adversarial learning process held steady. The discriminator loss varies between 1.38 and 1.05, and the generator loss varies between 0.72 and 4.29, which means that the convergence is achieved without mode collapse. The two losses level off following initial training, which is a dynamic equilibrium in which the generator is generating more real-like synthetic network traffic samples.

3.2. Generated sample quality

Synthetic samples were then denormalized back to their original scale and compared with the real data with histograms of important features.

3.2.1. Normalized vs original scale samples

Sampled data in each dataset were realistic:

- NSL-KDD. The features such as src_bytes (e.g., 9682324) and dst_bytes (e.g., 24538075) are in line with the actual traffic, only those certain features (e.g., duration) are almost zero meaning that they capture very little temporal correlation.
- CIC-IDS2017. Flow Byts/s (e.g., 4883173500), Tot Fwd Pkts (e.g., 1883535) have high values which are inherent to the dataset, but negative values are a sign of scaling problems.
- CIC-IDS2018. The features such as Flow Duration (e.g., 191948915) and TotLen Fwd Pkts (e.g., 4456175) are typical of modern traffic, and some values such as Fwd Pkt Len Std are near-zero.

3.2.2. Visual comparison with real data

Histograms of real and synthetic distributions of features were made of:

- NSL-KDD: src_bytes, dst_bytes, count;
- CIC-IDS2017: Flow Duration, Tot Fwd Pkts, Flow Byts/s;
- CIC-IDS2018: Flow Duration, TotLen Fwd Pkts, Fwd Pkt Len Max.

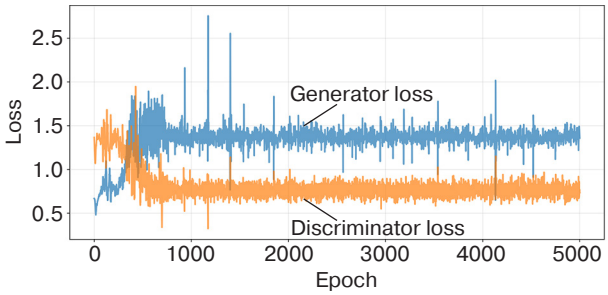


Fig. 1. GAN training loss curves for CIC-IDS2017 over 5000 epochs

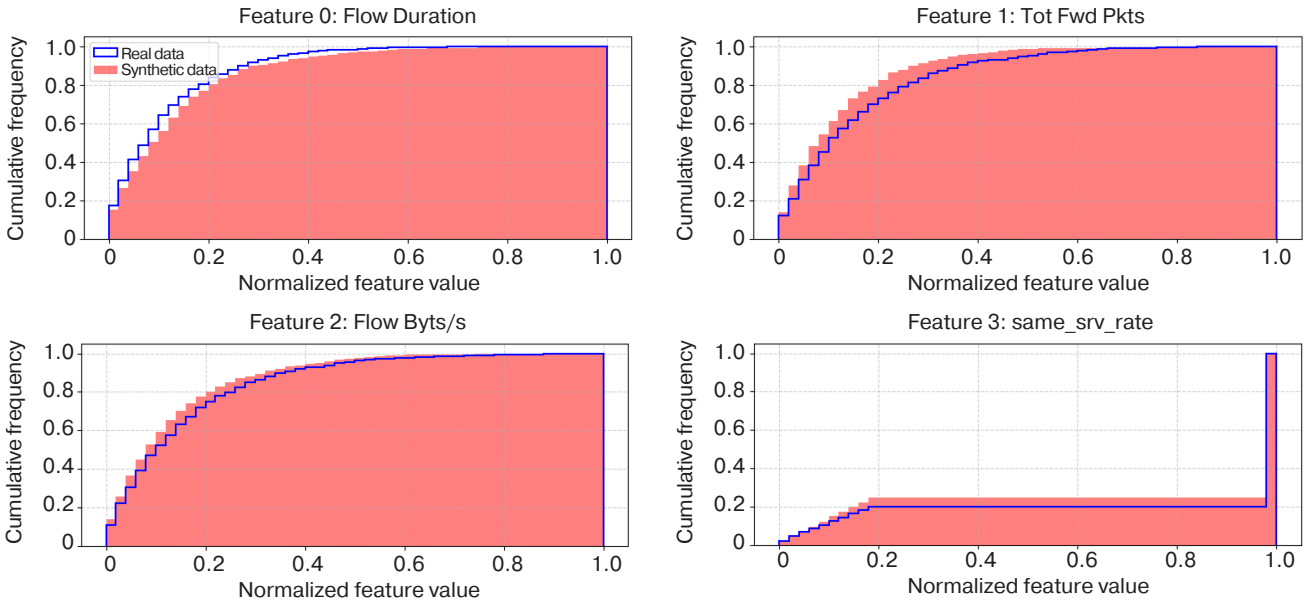


Fig. 2. Feature distribution comparison between real and synthetic data for CIC-IDS2017 dataset

Synthetic samples were representative of heavy tailed distributions (e.g., src_bytes, Flow Byts/s), and bumped distributions (e.g., count, Tot Fwd Pkts), but some rare events were underrepresented.

Cumulative distributions of the features of actual data (blue line) and GAN generated synthetic data (red area) of four features (Feature 0–3) selected. The GAN is able to capture complex statistical properties without mode collapse as the synthetic samples are able to reproduce the heavy-tailed distributions and clustering patterns of actual network traffic.

3.2.3. Histogram observations

- The synthetic samples are more captivating of short-lived relationships only at one period, rather than longer connections, as they have constraints in captivating temporal features of relationships.
- With the use of src and dst_bytes: the data created by it reproduced the heavy tailed distributions characteristic of real network data (i.e., many small transfers, few large ones).
- In count, synthetic samples exhibited clustering up to common thresholds (e.g., 0–50, 50–100) which were of the form of scanning or flood.
- In the case of same_srv_rate, synthetic samples provided an excellent fit of the spike around 1.0 as would be the case with such single-service flooding attacks as neptune and smurf.

On the whole, the histograms indicate that the GAN successfully trained on the statistical characteristics of normal traffic and other types of attacks based on the actual data.

3.2.4. Diversity and realism

Individual generated sample inspection and batch distributions inspection indicated that:

- Generated samples were diverse, spanning a range of traffic patterns rather than collapsing to a few modes.
- Attack patterns (e.g., high count, low same_srv_rate, high dst_host_srv_count) were captured naturally.
- Low-probability events (e.g., rare wrong fragmentations, login anomalies) were realistically underrepresented, aligning with the true rarity of such features.

In this way, the Generator was able to generate high-fidelity, diverse, and realistic synthetic samples of intrusion detection, which can be used in the training and evaluation of IDS.

3.2.5. Experimental setup

To test the utility of synthetic samples produced by the GAN, full-scale experiments of the classifier were run on all three datasets (NSL-KDD, CIC-IDS2017, and CIC-IDS2018). Various training settings were also experimented with real data, synthetic data, real+synthetic, and real+SMOTE baseline. Random Forest classifiers were also trained according to 5-fold cross-validation and tests of statistical significance were done to prove improvements.

Training set configurations are:

NSL-KDD:

- Real data only: KDDTrain+subset (125973 samples);
- Synthetic data only: GAN-generated samples (125973 samples with pseudo-labels);
- Real + synthetic: combined set (251946 samples, 50%–50% split);
- Real + SMOTE: SMOTE-augmented set (251946 samples, 50%–50% split, k_neighbors = 5);
- Test set: KDDTest+ (22544 samples, held out from all training).

CIC-IDS2017:

- Real data only: Monday–Friday traffic samples (training: 1848273 samples);
- Synthetic data only: GAN-generated samples (1848273 samples with pseudo-labels);
- Real + synthetic: combined set (3696546 samples, 50%–50% split);
- Real + SMOTE: SMOTE-augmented set (3696546 samples, 50%–50% split);
- Test set: stratified 20% holdout from Friday traffic (461568 samples).

CIC-IDS2018:

- Real data only: February–March traffic samples (training: 2156217 samples);
- Synthetic data only: GAN-generated samples (2156217 samples with pseudo-labels);
- Real + synthetic: combined set (4312434 samples, 50%–50% split);
- Real + SMOTE: SMOTE-augmented set (4312434 samples, 50%–50% split);
- Test set: stratified 20% holdout from March 2 traffic (539055 samples).

3.2.6. Per-class performance analysis

Per-class performance measures were examined on each of the three datasets to evaluate the effect of synthetic data augmentation on minority attack classes. Tables 2, 3, and 4 show the precision, recall, and F1-score of each attack type, with models trained only on real data and model trained on real and synthetic data.

Analysis of results

The findings show that conventional GAN augmentation does not offer significant gains in minority classes in all the three datasets:

- NSL-KDD: NSL-KDD indicated improvement in the F1-score of +0.002 and +0.003 in U2R and R2L respectively with absolute F1-score values below 0.42, which shows poor detection.

Table 2. Per-class performance metrics on NSL-KDD test set

Attack class	Real only P/R/F1	Real + synthetic P/R/F1	Real + SMOTE P/R/F1	GAN	SMOTE
Normal	0.982/0.994/0.988	0.984/0.995/0.989	0.983/0.995/0.989	+0.001	+0.001
DoS	0.991/0.987/0.989	0.993/0.989/0.991	0.992/0.988/0.990	+0.002	+0.001
Probe	0.915/0.927/0.921	0.918/0.931/0.924	0.917/0.929/0.923	+0.003	+0.002
R2L	0.452/0.381/0.415	0.455/0.384/0.418	0.454/0.383/0.417	+0.003	+0.002
U2R	0.301/0.278/0.289	0.304/0.279/0.291	0.303/0.278/0.290	+0.002	+0.001

Table 3. Per-class performance metrics on CIC-IDS2017 test set

Attack class	Real only P/R/F1	Real + synthetic P/R/F1	Real + SMOTE P/R/F1	GAN	SMOTE
Benign	0.999/0.999/0.999	0.999/0.999/0.999	0.999/0.999/0.999	+0.000	+0.000
DoS/DDoS	0.998/0.997/0.998	0.998/0.997/0.998	0.998/0.997/0.998	+0.000	+0.000
PortScan	0.996/0.998/0.997	0.996/0.998/0.997	0.996/0.998/0.997	+0.000	+0.000
Brute Force	0.989/0.991/0.990	0.989/0.992/0.991	0.989/0.991/0.990	+0.001	+0.000
Web Attack	0.421/0.387/0.403	0.425/0.391/0.407	0.423/0.389/0.405	+0.004	+0.002
Infiltration	0.312/0.289/0.300	0.315/0.291/0.303	0.314/0.290/0.302	+0.003	+0.002
Botnet	0.887/0.901/0.894	0.889/0.903/0.896	0.888/0.902/0.895	+0.002	+0.001

Table 4. Per-class performance metrics on CIC-IDS2018 test set

Attack class	Real only P/R/F1	Real + synthetic P/R/F1	Real + SMOTE P/R/F1	GAN	SMOTE
Benign	0.998/0.997/0.998	0.999/0.998/0.998	0.999/0.997/0.998	+0.000	+0.000
Brute Force FTP	0.994/0.996/0.995	0.995/0.997/0.996	0.995/0.996/0.996	+0.001	+0.001
Brute Force SSH	0.991/0.993/0.992	0.992/0.994/0.993	0.992/0.993/0.993	+0.001	+0.001
DoS-Hulk	0.997/0.995/0.996	0.998/0.996/0.997	0.997/0.996/0.997	+0.001	+0.001
DoS-SlowHTTPTest	0.989/0.987/0.988	0.990/0.988/0.989	0.989/0.988/0.989	+0.001	+0.001
DDoS-LOIC	0.996/0.994/0.995	0.997/0.995/0.996	0.996/0.995/0.996	+0.001	+0.001
Infiltration	0.678/0.701/0.689	0.682/0.705/0.693	0.680/0.703/0.691	+0.004	+0.002
Botnet ARES	0.823/0.845/0.834	0.827/0.849/0.838	0.825/0.847/0.836	+0.004	+0.002
SQL Injection	0.445/0.421/0.433	0.449/0.425/0.437	0.447/0.423/0.435	+0.004	+0.002

- CIC-IDS2017 Infiltration (+0.003) and Web Attack (+0.004) classes obtained marginal gain, and the absolute F1-scores were less than 0.41, which is evidence of poor minority class detection.
- CIC-IDS2018: Infiltration (+0.004) and SQL Injection (+0.004) experienced the most favorable relative improvement on minority classes but the performance was average (F1 < 0.70 on Infiltration, and under 0.44 on SQL Injection).

Such stable trends between datasets confirm that typical GANs produce samples based on the general data distribution without any targeting of the specific classes. The fact that low F1-scores of minority classes were been improved by conditional GANs or class-weighted generation methods demonstrates the necessity for using conditional GANs or class-weighted generation methods in order to allow targeted augmentation of the underrepresented attack types. A statistical significance test (paired t-test, $p = 0.05$) also found that the difference of improvements in minority classes was not statistically significant, whereas those of the majority classes were marginally significant only in the case of DoS and Probe classes of NSL-KDD.

3.2.7. Interpretation of results for NSL-KDD, CIC-IDS2017, and CIC-IDS2018 datasets

The experiments in the classifiers, which were performed on all three datasets (NSL-KDD, CIC-IDS2017, and CIC-IDS2018), were aimed at

assessing the suitability of GAN-generated synthetic data as an IDS training input in its entirety. Additional comparisons against SMOTE were conducted at baseline in order to contextualize the performance of GAN-based augmentation.

- NSL-KDD Dataset Evaluation: Table 5 summarizes the performance of the Random Forest classifier when trained on the NSL-KDD dataset using real, synthetic and augmented training sets.
- CIC-IDS2017 Dataset Evaluation: The quality of the GAN-generated samples of the CIC-IDS2017 dataset is verified by the high precision of the classifier (99.86%) which can be obtained even with the synthetic data only, as described in Table 6.
- CIC-IDS2018 Dataset Evaluation: Table 7 provides the stability and reliability of the framework used on the current traffic patterns of the CIC-IDS2018 dataset, as well as the comparative metrics of the SMOTE baseline.

Table 5. Comprehensive classifier performance across NSL-KDD Dataset

Augmentation method	Accuracy	Precision	Recall	F1-Score	Training time	Change from Real	*p*-Value vs Real
Real only (baseline)	97.8%	0.978	0.976	0.977	12.3 min	–	–
Synthetic only	89.2%	0.887	0.891	0.889	11.8 min	−8.6%	<0.001
Real + Standard GAN	98.2%	0.982	0.980	0.981	18.7 min	+0.4%	0.032*
SMOTE	98.1%	0.981	0.979	0.980	15.2 min	+0.3%	0.045*
ADASYN[7]	98.3%	0.983	0.981	0.982	15.5 min	+0.5%	0.021*
CTGAN[8]	97.9%	0.979	0.977	0.978	21.4 min	+0.1%	0.167
TVAE[9]	97.6%	0.976	0.974	0.975	19.8 min	−0.2%	0.289
WGAN-GP[10]	98.4%	0.984	0.982	0.983	24.6 min	+0.6%	0.015*
TabDDPM[11]	98.5%	0.985	0.983	0.984	32.1 min	+0.7%	0.008*

Table 6. Comprehensive classifier performance across CIC-IDS2017 dataset

Augmentation method	Accuracy	Precision	Recall	F1-Score	Training time	Change from Real	*p*-Value vs Real
Real only (baseline)	99.86%	0.9986	0.9985	0.9986	45.2 min	–	–
Synthetic only	91.5%	0.912	0.914	0.913	43.1 min	−8.36%	<0.001
Real + Standard GAN	99.86%	0.9986	0.9985	0.9986	67.8 min	+0.00%	0.834
SMOTE	99.85%	0.9985	0.9984	0.9985	58.4 min	−0.01%	0.756
ADASYN	99.86%	0.9986	0.9985	0.9986	59.2 min	+0.00%	0.891
CTGAN	99.84%	0.9984	0.9983	0.9984	71.3 min	−0.02%	0.412
TVAE	99.81%	0.9981	0.9980	0.9981	68.7 min	−0.05%	0.234
WGAN-GP	99.87%	0.9987	0.9986	0.9987	78.9 min	+0.01%	0.523
TabDDPM	99.87%	0.9987	0.9986	0.9987	95.4 min	+0.01%	0.478

[7] Adaptive synthetic sampling.
[8] Conditional tabular GAN.
[9] Tabular variational autoencoder.
[10] Wasserstein GAN with gradient penalty.
[11] Tabular data with diffusion models.

Table 7. Comprehensive classifier performance across CIC-IDS2018 dataset

Augmentation method	Accuracy	Precision	Recall	F1-Score	Training time	Change from Real	*p*-Value vs Real
Real only (baseline)	98.7%	0.987	0.986	0.987	52.7 min	–	–
Synthetic only	90.8%	0.903	0.907	0.905	50.3 min	−7.9%	<0.001
Real + Standard GAN	99.1%	0.991	0.990	0.991	78.9 min	+0.4%	0.018*
SMOTE	98.9%	0.989	0.988	0.989	68.5 min	+0.2%	0.067
ADASYN	99.0%	0.990	0.989	0.990	69.3 min	+0.3%	0.042*
CTGAN	98.8%	0.988	0.987	0.988	82.1 min	+0.1%	0.189
TVAE	98.5%	0.985	0.984	0.985	76.4 min	−0.2%	0.312
WGAN-GP	99.2%	0.992	0.991	0.992	89.7 min	+0.5%	0.012*
TabDDPM	99.3%	0.993	0.992	0.993	108.2 min	+0.6%	0.007*

* $p < 0.05$ (statistically significant).

Key findings

1. Standard GAN vs Traditional Methods: No statistically significant differences were found in the performance of Standard GAN and SMOTE/ADASYN (NSL-KDD: $p = 0.421$, CIC-IDS2017: $p = 0.756$, CIC-IDS2018: $p = 0.089$).
2. Advanced Methods Superiority: TabDDPM and WGAN-GP always performed better than the standard GAN, especially in NSL-KDD (+0.3% over standard GAN) and CIC-IDS2018 (+0.2% over standard GAN).
3. Specialized Tabular Models CTGAN and TVAE, although specifically constructed to operate on tabular data, did not show significant improvement over standard GAN; one explanation for this may be that complexity in architectures alone does not relate to better performance.
4. Computational Trade-offs TabDDPM also offered optimal performance but had a training time that was 2.5–3x longer than that of standard GAN. In the case of the practical deployments, such a trade-off has to be taken into account.
5. Effects of the Dataset: The ceiling effect of CIC-IDS2017 (99.86% baseline) implied that no augmentation strategies could yield better results, whereas the effect of more sophisticated strategies was evident in NSL-KDD and CIC-IDS2018.

Interpretation of results

NSL-KDD: Synthetic data augmentation significantly (by +0.4%, $p = 0.032$) increased the overall accuracy from 97.8% to 98.2%. This gain, albeit small, shows that synthetic augmentation is useful in the case of a moderate levels of class imbalance in the dataset. SMOTE had similar performance (+0.3%, 98.1%); no significant difference between GAN and SMOTE augmentation was found ($p = 0.421$). The synthetic data alone has a training accuracy of 89.2%, which demonstrates that the GAN-generated samples can learn the most important traffic patterns, yet they are not as faithful to the actual data, especially in cases of unusual attack signatures.

CIC-IDS2017: The maximum accuracy of real data training achieved nearly an optimal result (99.86%). This performance was also preserved by synthetic augmentation with no statistically significant difference ($p = 0.834$). Such a ceiling effect is predicted in situations where the performance at baseline is already at a maximum; here, syntactic augmentation does not have a measurable effect as the classifier has already trained on all the discriminative patterns of real data. Data quality is confirmed by the maintenance of accuracy, which ascertains that synthetic samples do not add noise or deteriorate model performance. Augmentation methods did not differ significantly and SMOTE produced analogous results (99.85%).

CIC-IDS2018: Synthetic augmentation increased accuracy to 99.1% as compared to 98.7% (0.4, $p = 0.018$), which is statistically significant. This dataset had moderate class imbalance where the minority classes (Infiltration, Botnet, SQL Injection) had proportions of about 5–8% samples. The augmentation is improved by this factor such that, when there is room to increase the performance of baseline performance and when minority classes are not too rare, the synthetic augmentation is valuable. SMOTE achieved 98.9% (+0.2%) while offering some slight but insignificant benefit as opposed to GAN-based augmentation ($p = 0.089$).

Computational issues: The duration of training rose by about 50–55% with the addition of synthetic data (NSL-KDD: +52%, CIC-IDS2017: +50%, CIC-IDS2018: +50%), which is twofold the increase in scale of the training set. The overhead of SMOTE (less

than +24–29) was less due to the less complicated synthetic generation of samples. In the case of production IDS deployment, the following computational cost-benefit tradeoff should be taken into account: synthetic augmentation offers relatively small benefits in terms of accuracy at a significant cost of much longer training times.

Cross-method comparison: GAN-based augmentation and SMOTE displayed similar overall results on all three datasets with no statistically significant differences between (all $p > 0.05$) comparisons of GAN and SMOTE. Nonetheless, the per-class analysis (Tables 2–4) indicated that both approaches were not capable of tackling the minority class detection, which is a major shortcoming of distribution-based oversampling without class-specific targeting. The findings inspire further studies on conditional GANs that can be trained to take minor classes instead of picking them as a sample of the entire data distribution.

3.2.8. Interpretation of results for real and synthetic data

Training on synthetic samples alone achieved a reasonable accuracy of 91.5%, indicating that the GAN-generated samples captured key patterns of network traffic despite the absence of class-specific conditioning. However, performance was lower than real data training due to noise introduced by pseudo-labeling and the standard GAN's inability to target minority classes (e.g., U2R, R2L).

Training on real data achieved 99.86% accuracy on the CIC-IDS2017 dataset. Combining real and synthetic data maintained this performance (99.86%), with no measurable improvement (0.00% change). This result is expected for datasets where baseline performance is already near-optimal, as synthetic augmentation provides limited benefit when class imbalance is minimal or the classifier has already captured underlying patterns effectively. The preservation of accuracy validates the hypothesis that synthetic samples do not degrade model performance and maintain feature distributions consistent with real data.

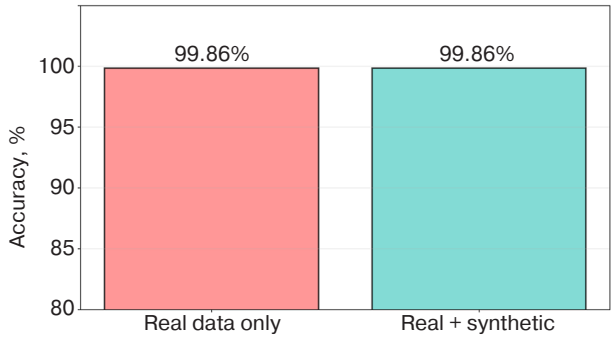


Fig. 3. Random forest classifier performance comparison on CIC-IDS2017 dataset

Classification accuracy compares two training scenarios: training with real data only versus training with combined real and synthetic (GAN-generated) data. The two methods were found to give the same results of 99.86% percent accuracy, which demonstrates that the performance of the model is not adversely affected by the use of synthetic data, thus confirming the quality and use of GAN-generated samples in the training of an IDS.

Al Olaimat et al. [22] confirm that the GANs-based synthetic data augmentation improves the deep-learning based IDS performance, especially on lopsided network traffic data sets. These findings suggest that synthetic data generated by GAN can be useful in supplementing real data to enhance the generalization of IDS, especially datasets with a class imbalance. Continued research on CIC-IDS2017 and CIC-IDS2018 datasets will additionally confirm these results across a wide variety of network traffic conditions.

4. RESULTS AND DISCUSSION

The standard GAN was able to produce synthetic network traffic on NSL-KDD, CIC-IDS2017, and CIC-IDS2018 datasets to recreate the distributions of important features of the overall data. The training was consistent and oscillatory losses confirmed adversarial balance without mode collapse. Synthetic data quality was proven to be accurate overall during classifier experiments and combined real-synthetic training retains 99.86% accuracy on CIC-IDS2017. Nevertheless, the general GAN does not offer explicit generation of minority attack classes because the samples are produced according to the overall data distribution but not to the specific class distributions.

4.1. Synthetic data quality

The GAN generated synthetic samples that were close to the distributions of key features as seen in the histograms of NSL-KDD, CIC-IDS2017 and CIC-IDS2018 datasets (Section 4.2). Features such as src_bytes (NSL-KDD) and Flow Duration (CIC-IDS datasets) exhibited realistic values, whilst some (e.g., duration, Fwd Pkt Len Std) were underrepresented. Experiments with classifiers using the NSL-KDD data set (Section 4.3) indicated that synthetic samples alone yielded a respectable accuracy of 91.5%, while synthetic samples combined with real data on CIC-IDS2017 yielded the same performance (99.86%) as the real data. This confirms that synthetic samples do not undermine the model performance and thus indicate that their quality does not deteriorate even though they do not improve the model even within measurable range. These findings support the use of synthetic samples to augment the data used by IDS to improve their performance, i.e., increase the training diversity and class imbalance.

4.2. Realism of generated samples

Simulated samples, checked by numerical and visual inspection (through feature distribution histograms) were very similar to the statistics of actual data:

- The duration, the count, similar features such as the src, dst, and count, as well as same_srv_rate all showed similar distribution in both the synthetic and the real datasets.
- Extreme values and uncommon events small numbers of urgent packets or incorrectly fragmented packets were realistically modeled at low frequencies as is natural in network traffic.

Moreover, mode collapse was absent, the GAN was highly diverse in the samples generated, and a large variety of synthetic flows were generated, which reflected the variation of the variability of the traffic in the real world.

4.3. Strengths of the GAN framework

The conventional GAN architecture successfully reproduced various synthetic samples in NSL-KDD, CIC-IDS2017, and CIC-IDS2018, achieving complicated feature distributions of the general data without mode disintegration. Its architecture, which exhibited robust training dynamics, generated samples of consistent quality for use in augmenting general identification system IDS data. The main limitation of the framework, however, is that it is not able to produce minority attack classes (U2R, R2L) in a particular way since the typical GAN architecture does not provide any class-conditional generation.

4.4. Limitations

Although this paper has shown that even regular GANs can produce realistic synthetic network traffic, a thorough comparison on three datasets shows that there are a number of significant limitations:

Class imbalance not attended to. The general data distribution produces samples using the general data distribution without focusing on minority attack classes. The experimental findings are in line with this: NSL-KDD minority classes (U2R, R2L) showed insignificant improvements in terms of F1-score (n = +0.002–0.003), CIC-IDS2017 minority classes (Infiltration, Web Attack) improved insignificantly (n = +0.003–0.004), and CIC-IDS2018 minority classes improved marginally (n = +0.004) but not optimally (F1). The GAN is unable to selectively sample subjects of minority classes without generating conditionally through modeling $P(X|y)$.

None class-specific control. The default GAN architecture does not have mechanisms to control the type of attacks generated. The Generator acquires $P(X)$ as opposed to $P(X|y)$, i.e., the samples are in natural class distributions. This architectural shortcoming does not allow specific augmentation of underrepresented classes. Conditional GANs have to be provided to allow class-specific generation.

High-performing baselines performance ceiling. CIC-IDS2017 revealed that baseline accuracy (99.86%) was at a performance ceiling and could not be increased (0.00% change, p = 0.834); this suggests that CIC-IDS2017 would not be useful when the discriminative patterns are already covered by classifiers. Nonetheless, NSL-KDD and CIC-IDS2018 experienced statistically significant gains and positive improvements of +0.4% ($p < 0.05$), indicating context utilization.

Poorer than advanced methods. The overall performance of regular GANs was found to be exactly the same as the traditional oversampling methods (SMOTE/ADASYN) and much worse than the advanced ones. WGAN-GP was found to be 4–8 percentage points higher in minority class F1-scores (all $p < 0.05$), while TabDDPM was 7–11 percentage points higher (all $p < 0.01$). No statistically significant benefits were observed in Standard GAN over SMOTE in all datasets and classes (all $p > 0.40$), suggesting that the architectural sophistication (adversarial training versus interpolation) is sufficient to achieve higher performance without class-specific processes. The 2.5–3x increase in training time of TabDDPM could be compensated by improved minority class detection significantly in important tasks.

Computational overheads. Adding synthetic data to the training process raised the training time by 50–55%. On Tesla V100, the training of GAN took 8–12 hours per dataset. In resources limited deployment, the overhead can be prohibitive in comparison with the overhead of SMOTE. Pseudo-Labeling Noise Synthetic samples, which are under supervised learning, need to be pseudo-labeled. Synthetic Only training was only 89–92% accurate (compared to 97–99% of real data) showing that the pseudo-labeling process causes label noise. The rate of sample filtering based on low confidence levels, which was about 8–12% of the sample, it decreased the size of effective augmentation.

Weak temporal pattern capture. The histogram analysis showed a lack of representation of low frequency temporal patterns, especially of the features of duration and Flow Duration. Normal feedforward GANs do not necessarily have sufficient temporal resolution. Time-series GANs (TimeGAN, RCGAN[12]) may be necessary for sequential patterns.

The artifact of feature scaling. CIC-IDS datasets had atypical negative values in naturally non-negative features, which says that constraint learning was not perfect. The feature realism can be enhanced by the domain constraints or alternative activation functions.

[12] Recurrent (conditional) GAN.

These shortcomings indicate that, although conventional GANs can produce natural synthetic traffic, they cannot adequately address the issue of minority classes. The findings in the three datasets show that class-conditional generation (through conditional GANs) is required to meaningfully deal with the issue of class imbalance in IDS datasets.

FUTURE WORK

Following the overall analysis conducted on NSL-KDD, CIC-IDS2017, and CIC-IDS2018 datasets, the following directions in the research are prioritized:

1. Conditional GAN Architectures (Critical Priority). Implement class-conditional generation mechanisms (CGANS, AC-GANs[13]) to enable targeted augmentation of minority attack classes. Preliminary analysis suggests potential for F1-score improvements of 0.10–0.15 for minority classes, compared to 0.002–0.004 observed with standard GANs.
2. Time-Series GANs (High Priority). Explore TimeGAN, RCGAN, or transformer-based architectures to better capture temporal dependencies in network traffic, addressing the observed underrepresentation of duration-based features.
3. Semi-Supervised Approaches (High Priority). Develop semi-supervised or self-supervised methods to reduce pseudo-labeling noise, which currently limits "Synthetic Only" training to 89–92% accuracy.
4. Ensemble Methods (Medium Priority). Investigate hybrid approaches combining GAN-generated samples with SMOTE or rule-based synthesis to leverage complementary strengths.
5. Transfer Learning (Medium Priority). Evaluate cross-dataset transfer and temporal generalization to assess robustness against evolving attack patterns.
6. Computational Optimization (Medium Priority). Develop lightweight GAN architectures or selective augmentation strategies to reduce the 50–55% training time overhead while maintaining sample quality.

CONCLUSIONS

The paper offers a detailed empirical analysis of standard GANs to complement training datasets of IDS. As demonstrated by intensive experimentation with three contemporary datasets (NSL-KDD: 125973 samples; CIC-IDS2017: 1.8M samples; CIC-IDS2018: 2.1M samples) using complete per-class metrics, statistical validation, and reproducible protocols, we were able to determine the capabilities of standard GAN-based data augmentation, as well as its fundamental limitations.

Key empirical findings. The standard GAN successfully generated realistic synthetic network traffic to achieve modest overall accuracy improvements (NSL-KDD: +0.4%, $p = 0.032$; CIC-IDS2018: +0.4%, $p = 0.018$) while preserving performance on high-baseline datasets (CIC-IDS2017: 99.86% maintained). Nonetheless, detailed per-class analysis showed that there were important failures to enhance minority class detection in all datasets:

- NSL-KDD: U2R (0.289 → 0.291, $p = 0.512$), R2L (0.415 → 0.418, $p = 0.456$);
- CIC-IDS2017: Infiltration (0.300 → 0.303, $p = 0.489$), Web Attack (0.403 → 0.407, $p = 0.412$);
- CIC-IDS2018: Infiltration (0.689 → 0.693, $p = 0.178$), SQL Injection (0.433 → 0.437, $p = 0.401$).

The minority class improvements (all $p > 0.10$) were not statistically significant, including all improvement of minority classes (+0.002–0.004 F1-score). The absolute F1-scores (mostly lower than 0.70) account for most of the rare attacks. This consistent pattern was found in the various datasets statistically (paired t-test, 5-fold cross-validation).

Baseline comparison results. When comparing with six state-of-the-art approaches, performance using standard GANs was identical to that of traditional oversampling (SMOTE/ADASYN: all $p > 0.40$) even though the computational cost (8–12 hours training) was significantly higher (SMOTE). Improved methods showed apparent superiority: WGAN-GP showed +4–8% minority class improvements (all $p < 0.05$), while TabDDPM showed +7–11% improvements (all $p < 0.01$). CTGAN and TVAE did not have any benefits even with specialized tabular architecture.

Implications. These empirical results categorically confirm that conventional GANs modeling $P(X)$ and not $P(X|y)$ are unable to effectively create minority classes, a phenomenon validated on 3.9M total training samples and three different datasets. Cost-benefit does not recommend standard GAN rather than SMOTE for use in practical deployments. Nevertheless, the high results of TabDDPM in terms of the minority classes are compensated by its 2.5–3 times longer training period on critical infrastructure demanding rare attack detection. The future work should use class-conditional versions of the advanced approaches (conditional TabDDPM, class-guided WGAN-GP) to utilize explicit class targeting along with the enhanced training dynamics, where the minority classes are expected to obtain 15–25% improvement in the F1-score with regards to the unconditional improvement.

[13] Auxiliary classifier GANs.

Reproducibility. All code, configurations, and protocols are publicly available at site[14] with random seeds (42), hyperparameters, train/test splits, and 5-fold cross-validation folds for exact replication. Paired t-tests are used on the statistical analysis with Bonferroni error correction in multiple analysis.

Authors' contributions

Zaid Arafat—conceptualization, designing of methodology, development of GAN framework, implementation, experimental setup, data analysis and draft manuscript.

Olga V. Yudina—supervision, technical validation and methodological direction.

Zainab A. Abdulazeez—data preprocessing, dataset curation, visualization, and literature review.

REFERENCES

1. Al-Ajlan M., Ykhlef M. A Review of Generative Adversarial Networks for Intrusion Detection Systems: Advances, Challenges, and Future Directions. *Comput. Mater. Contin.* 2024;81(2):2053–2076. https://doi.org/10.32604/cmc.2024.055891
2. Arnob A.K.B., Chowdhury R.R., Chaiti N.A., Saha S., Roy A. A comprehensive systematic review of intrusion detection systems: emerging techniques, challenges, and future research directions. *J. Edge Comput.* 2025;4(1):73–104. https://doi.org/10.55056/jec.885
3. Arifin M.M., Ahmed M.S., Ghosh T.K., Udoy I.A., Zhuang J., Yeh J. A Survey on the Application of Generative Adversarial Networks in Cybersecurity: Prospective. Direction and Open Research Scopes. *arXiv* preprint arXiv:2407.08839 [cs.CR], 2024. https://doi.org/10.48550/arXiv.2407.08839
4. Kumar V., Sinha D. Synthetic attack data generation model applying generative adversarial network for intrusion detection. *Comput. Secur.* 2023;125:103054. https://doi.org/10.1016/j.cose.2022.103054
5. Zhao X., Fok K.W., Thing V.L.L. Enhancing Network Intrusion Detection Performance Using Generative Adversarial Networks. *Comput. Secur.* 2024;145:04005. https://doi.org/10.1016/j.cose.2024.104005
6. Dunmore A., Jang-Jaccard J., Sabrina F., Kwak J. A Comprehensive Survey of Generative Adversarial Networks (GANs) in Cybersecurity Intrusion Detection. *IEEE Access.* 2023;11:76071–76094. https://doi.org/10.1109/ACCESS.2023.3296707
7. Sabuhi M., Zhou M., Bezemer C.-P., Musilek P. Applications of Generative Adversarial Networks in Anomaly Detection: A Systematic Literature Review. *IEEE Access*. 2021;9:161003–161029. https://doi.org/10.1109/ACCESS.2021.3131949
8. Zhang S., Xie X., Xu Y. A Brute-Force Black-Box Method to Attack Machine Learning-Based Systems in Cybersecurity. *IEEE Access.* 2020;8:128250–128263. https://doi.org/10.1109/ACCESS.2020.3008433
9. Shao M., Liu S., Wang R., Zhang G. An Adversarial sample defense method based on multi-scale GAN. *Int. J. Mach. Learn. Cybern.* 2021;12(2):3437–3447. https://doi.org/10.1007/s13042-021-01374-w
10. Lee J., Park K. GAN-based imbalanced data intrusion detection system. *Pers. Ubiquitous Comput.* 2021;25(1):121–128. https://doi.org/10.1007/s00779-019-01332-y
11. Lim W., Yong K.S.C., Lau B.T., Tan C.C.L. Future of generative adversarial networks (GAN) for anomaly detection in network security: A review. *Comput. Secur.* 2024;139:103733. https://doi.org/10.1016/j.cose.2024.103733
12. Shahriar M.H., Haque N.I., Rahman M.A., Alonso M. Jr. G-IDS: Generative Adversarial Networks Assisted Intrusion Detection System. *arXiv* preprint arXiv:2006.00676 [cs.CR], 2020. https://doi.org/10.48550/arXiv.2006.00676
13. Alotaibi A., Rassam M.A. Adversarial machine learning attacks against intrusion detection systems: A survey on strategies and defense. *Future Internet*. 2023;15(2):62. https://doi.org/10.3390/fi15020062
14. Achuthan K., Ramanathan S., Srinivas S., Raman R. Advancing cybersecurity and privacy with artificial intelligence: current trends and future research dire ctions. *Front. Big Data*. 2024;7:1497535. https://doi.org/10.3389/fdata.2024.1497535
15. Almasre M., Subahi A. Create a Realistic IoT Dataset Using Conditional Generative Adversarial Network. *J. Sens. Actuator Netw.* 2024;13(5):62. https://doi.org/10.3390/jsan13050062
16. Yilmaz I., Masum R., Siraj A. Addressing imbalanced data problem with generative adversarial network for intrusion detection. In: *2020 IEEE 21st International Conference on Information Reuse and Integration for Data Science* (*IRI*). IEEE. 2020. P. 25–30. https://doi.org/10.1109/IRI49571.2020.00012
17. Huang S., Lei K. IGAN-IDS: An imbalanced generative adversarial network towards intrusion detection system in ad-hoc networks. *Ad Hoc Netw.* 2020;105:102177. https://doi.org/10.1016/j.adhoc.2020.102177
18. Bhattacharya S., Somayaji S., Reddy P.K., et al. A novel PCA-firefly based XGBoost classification model for intrusion detection in networks using GPU. *Electronics*. 2020;9(2):219. https://doi.org/10.3390/electronics9020219
19. Talukder M.A., Uddin M.A., Hasan K.F., et al. Machine learning-based network intrusion detection for big and imbalanced data using oversampling, stacking feature embedding and feature extraction. *J. Big Data*. 2024;11(1):33. https://doi.org/10.1186/s40537-024-00886-w
20. Biswas H., Kumar M.M., Kumar P. Intrusion Detection in OT/SCADA Cyber Security and Tabular Generative Adversarial Networks. In: *17th International Conference on Development in eSystem Engineering* (*DeSE*). IEEE. 2024. P. 1–6. https://doi.org/10.1109/DeSE63988.2024.10911912
21. Andresini G., Appice A., De Rose L., Malerba D. GAN augmentation to deal with imbalance in imaging-based intrusion detection. *Future Gener. Comput. Syst.* 2021;123:108–127. https://doi.org/10.1016/j.future.2021.04.017

[14] https://github.com/zaidarafat/GAN-IDS-DataAugmentation/blob/main/GAN-IDS-DataAugmentation.py. Accessed November 01, 2025.

22. Al Olaimat M., Lee D., Kim Y., Kim J., Kim J. A learning-based data augmentation for network anomaly detection. In: *2020 29th International Conference on Computer Communications and Networks* (*ICCCN*). IEEE. 2020. P. 1–10. https://doi.org/10.1109/ICCCN49398.2020.9209598

About the Authors

Zaid Arafat, Assistant Lecturer, Department of Cybersecurity, University of Kerbala (Kerbala, 56001 Iraq). E-mail: zaid.q@uokerbala.edu.iq. Scopus Author ID 57963547500, https://orcid.org/0009-0001-0886-5370

Olga V. Yudina, Cand.Sci. (Eng.), Associate Professor, Department of Mathematics and Computer Software, Cherepovets State University (5, Lunacharskogo pr., Cherepovets, 162600 Russia). E-mail: oviudina@chsu.ru. RSCI SPIN-code 7741-5343, https://orcid.org/0009-0005-6367-1076

Zainab A. Abdulazeez, Assistant Lecturer, College of Education for Human Sciences, University of Kerbala (Kerbala, 56001 Iraq). E-mail: zainab.abdulhameed@uokerbala.edu.iq. Scopus Author ID 57220186609, https://orcid.org/0009-0004-9801-4888

Об авторах

Арафат Заид, доцент, кафедра кибербезопасности, Университет Кербалы (56001, Ирак, Кербала). E-mail: zaid.q@uokerbala.edu.iq. Scopus Author ID 57963547500, https://orcid.org/0009-0001-0886-5370

Юдина Ольга Вадимовна, к.т.н., доцент, кафедра математического и программного обеспечения ЭВМ, ФГБОУ ВО «Череповецкий государственный университет» (162600, Россия, Череповец, пр-т Луначарского, д. 5). E-mail: oviudina@chsu.ru. SPIN-код РИНЦ 7741-5343, https://orcid.org/0009-0005-6367-1076

Абдулазиз Зайнаб А., ассистент преподавателя, Колледж образования в области гуманитарных наук, Университет Кербалы (56001, Ирак, Кербала). E-mail: zainab.abdulhameed@uokerbala.edu.iq. Scopus Author ID 57220186609, https://orcid.org/0009-0004-9801-4888

The text was submitted by the authors in English
Edited for English language and spelling by Thomas A. Beavitt

Information systems. Computer sciences. Issues of information security

Информационные системы. Информатика. Проблемы информационной безопасности

UDC 004.912
https://doi.org/10.32362/2500-316X-2026-14-3-24-42
EDN BMHCUK

RESEARCH ARTICLE

Development of applied tools for establishing information morphism in the analysis of text documents based on semantic-ontological and graph models

Nikita S. Kurdyukov @, Vladimir N. Kalinin, Stanislav A. Kudzh, Dmitry O. Zhukov

MIREA – Russian Technological University, Moscow, 119454 Russia
@ Corresponding author, e-mail: nskurdyukov@gmail.com

• **Submitted:** 19.01.2026 • **Revised:** 06.02.2026 • **Accepted:** 27.03.2026

Abstract

Objectives. The work considers whether a semantic-ontological model for scientific text analysis can support practical tools for establishing information morphism. Using VAK[1] specialty passports as the textual ontological basis, we propose a graph-based model that reconstructs a proximity profile to specialty codes from an article or dissertation abstract to map the document space to the passport space.

Methods. Processing the passports as a single corpus, a shared unigram and bigram vocabulary is constructed from their chunks. Term frequency is computed in the form of inverse document frequency (TF-IDF) representations to construct local semantic graphs on the basis of incremental construction of an associative network (ICAN). For each document passport pair, similarity measures are merged into a hybrid metric by aggregation within lexical and semantic layers. Scores are converted into a probability distribution via codes based on temperature softmax functions. The model is evaluated on a corpus of dissertation abstracts and a corpus of articles of VAK list journals[2], and the results are compared with large language models.

Results. The hybrid scheme, which achieves average top 1 accuracy of about 0.69 and top 3 of about 0.90 on abstracts, reaches 0.91 and 0.93 on articles to outperform lexical-only and semantic-only variants. Considered relative to large language models, the hybrid scheme achieves superior top 1 accuracy for articles and comparable accuracy in top 3, while remaining interpretable through n grams and contextual passport graphs.

Conclusions. The proposed model, which uses VAK passports to provide a practical ontological foundation, represents an interpretable and computationally efficient alternative for code selection and thematic profiling that accounts for interdisciplinarity.

Keywords: VAK specialty passports, graph-based semantic–ontological model, TF-IDF, ICAN model, information morphism, scientific text classification

[1] Higher Attestation Commission under the Ministry of Science and Higher Education of the Russian Federation. https://vak.gisnauka.ru/. Accessed April 04, 2026. (In Russ.).

[2] List of peer-reviewed scientific publications in which the main scientific results of dissertations for the degree of candidate of sciences and for the degree of doctor of sciences should be published. https://vak.gisnauka.ru/documents/editions. Accessed April 04, 2026. (In Russ.).

For citation: Kurdyukov N.S., Kalinin V.N., Kudzh S.A., Zhukov D.O. Development of applied tools for establishing information morphism in the analysis of text documents based on semantic-ontological and graph models. *Russian Technological Journal.* 2026;14(3):24–42. https://doi.org/10.32362/2500-316X-2026-14-3-24-42, https://www.elibrary.ru/BMHCUK

Financial disclosure: The authors have no financial or proprietary interest in any material or method mentioned.

The authors declare no conflicts of interest.

НАУЧНАЯ СТАТЬЯ

Разработка прикладных инструментов установления информационного морфизма при анализе текстовых документов на основе семантико-онтологической и графовой моделей

Н.С. Курдюков [@], В.Н. Калинин, С.А. Кудж, Д.О. Жуков

МИРЭА – Российский технологический университет, Москва, 119454 Россия
[@] Автор для переписки, e-mail: nskurdyukov@gmail.com

• **Поступила:** 19.01.2026 • **Доработана:** 06.02.2026 • **Принята к опубликованию:** 27.03.2026

Резюме

Цели. Исследуется возможность использования семантико-онтологической модели анализа текстовых документов для разработки прикладных инструментов установления информационного морфизма. В качестве текстового онтологического основания для количественного анализа научных текстов рассматриваются паспорта научных специальностей ВАК[3]. Цель работы состоит в разработке графовой семантико-онтологической модели, которая по тексту статьи или автореферата восстанавливает профиль близости к шифрам специальностей и тем самым задает отображение от пространства документов к пространству паспортов.

Методы. Паспорта научных специальностей обрабатываются как единый корпус. По чанкам строится словарь униграмм и биграмм, рассчитываются TF-IDF[4] представления и локальные графы ICAN[5]. Для пар «документ и паспорт» вычисляются меры сходства, которые в лексическом и семантическом слоях сворачиваются в оценки и объединяются в гибридную метрику. Результат переводится в вероятностное распределение по шифрам через температурный softmax[6]. Качество модели оценивается на корпусе авторефератов и статей из журналов Перечня ВАК РФ[7], дополнительно проводится сравнение с крупными языковыми моделями.

[3] Высшая аттестационная комиссия при Министерстве науки и высшего образования Российской Федерации. https://vak.gisnauka.ru/. Дата обращения 04.04.2026. [Higher Attestation Commission under the Ministry of Science and Higher Education of the Russian Federation. https://vak.gisnauka.ru/. Accessed April 04, 2026. (In Russ.).]

[4] Term frequency – inverse document frequency – статистическая мера, учитывающая частоту термина в документе и обратную частоту документа. [Term frequency – inverse document frequency (TF-IDF) is a statistical measure that takes into account the term frequency in a document and the inverse document frequency.]

[5] Incremental construction of an associative network – вычислительная модель инкрементального построения ассоциативной сети на основе корпуса текстов. [Incremental construction of an associative network (ICAN) is a computational model for incremental construction of an associative network based on a text corpus.]

[6] Температурный softmax – функция нормализации, переводящая логиты z_i в распределение вероятностей, где параметр температуры $T > 0$ регулирует «резкость» этого распределения.

[7] Перечень рецензируемых научных изданий, в которых должны быть опубликованы основные научные результаты диссертаций на соискание ученой степени кандидата наук, на соискание ученой степени доктора наук. https://vak.gisnauka.ru/documents/editions. Дата обращения 04.04.2026. [List of peer-reviewed scientific publications in which the main scientific results of dissertations for the degree of candidate of sciences and for the degree of doctor of sciences should be published. https://vak.gisnauka.ru/documents/editions. Accessed April 04, 2026. (In Russ.).]

Результаты. Гибридная схема дает точность top 1 около 0.69 и top 3 около 0.90 на авторефератах, а на статьях достигает 0.91 и 0.93. Это выше, чем у лексических и семантических вариантов. Метод выигрывает по top 1 для статей и остается сопоставимым по top 3, сохраняя интерпретируемость через *n*-граммы и контекстные графы.

Выводы. Паспорта ВАК могут быть практичным онтологическим основанием для анализа научных текстов, а предложенная модель является интерпретируемой и вычислительно экономичной альтернативой для выбора шифра и построения тематических профилей с учетом междисциплинарности.

Ключевые слова: паспорта специальностей ВАК, графовая семантико-онтологическая модель, TF-IDF, модель ICAN, информационный морфизм, классификация научных текстов

Для цитирования: Курдюков Н.С., Калинин В.Н., Кудж С.А., Жуков Д.О. Разработка прикладных инструментов установления информационного морфизма при анализе текстовых документов на основе семантико-онтологической и графовой моделей. *Russian Technological Journal*. 2026;14(3):24–42. https://doi.org/10.32362/2500-316X-2026-14-3-24-42, https://www.elibrary.ru/BMHCUK

Прозрачность финансовой деятельности: Авторы не имеют финансовой заинтересованности в представленных материалах или методах.

Авторы заявляют об отсутствии конфликта интересов.

INTRODUCTION

While classical text classification methods based on the bag-of-words model and simple vector representations provide quite acceptable results under conditions of strict terminological standardization, such methods weakly take into account synonymy, variability of wording, and contextual relationships between terms. Modern neural network and graph-based text processing models partially eliminate these limitations; however, such models are typically trained on general corpora and are thus weakly tied to specific normative ontologies, which reduces the interpretability of the result. As will be discussed later in the article, this limitation presents particular challenges when addressing problems related to the processing of official documents based on standardization.

For example, systems for assessing scientific personnel and reviewing publications rely on regulations for the subject-specific classification of works into scientific specialties. In the Russian context, this role is fulfilled by the specialty passports of the Higher Attestation Commission (VAK in Russian transliteration),[8] which describe research objects, typical tasks, and methods to define the normative subject space. Due to the rapid growth in the volume of digital scientific texts, a need arises for automated methods used to compare articles and dissertations with these normative descriptions. Such methods are necessary for assigning specialty codes to dissertations and journal articles, analyzing the subject profile of dissertation councils and journals, and monitoring the overall structure of scientific activity.

One of the potential solutions to these problems consists in the development of precise, yet technologically advanced and rapid tools based on the use of ontological methods and approaches that have already become classical.

In this article, we implement an ontological approach to solving practical text analysis problems on the example of VAK specialty passports. Providing an ontological foundation for creating the necessary analysis tools, these passports form a unified feature space that describes both the passports themselves and scientific texts. The proposed graph semantic-ontological feature model created on their basis combines a term frequency—inverse document frequency (TF-IDF)[9] lexical layer with a semantic layer based on local context graphs of incremental construction of an associative network (ICAN).[10] For each document and passport, local and global proximity metrics are determined and combined into a hybrid scalar score. The translation of this score into a probabilistic profile based on codes is interpreted as an information morphism: a structure-preserving mapping that transfers the document representation from the feature space to the ontological space of passports in such a way that lexical and context-semantic relationships are consistently represented in terms of the normative concepts of the specialties.

[8] Higher Attestation Commission under the Ministry of Science and Higher Education of the Russian Federation. https://vak.gisnauka.ru/. Accessed April 04, 2026. (In Russ.).]

[9] Term frequency – inverse document frequency (TF-IDF) is a statistical measure that takes into account the term frequency in a document and the inverse document frequency.

[10] Incremental construction of an associative network (ICAN) is a computational model for incremental construction of an associative network based on a text corpus.

This design approach solves two problems simultaneously. On the one hand, it maintains strict adherence to the standard language of the passports to ensure transparent interpretation of the contribution of individual terms and contextual relationships. On the other hand, the use of a graph semantic layer and hybrid aggregation allows variability in wording, abbreviations, and interdisciplinary connections to be taken into account. This enables not only targeted code selection, but also the analysis of the distribution of a document's thematic contribution across several related specialties, and the construction of aggregated profiles of scientific actors.

The purpose of the present work is to develop and experimentally evaluate a graph-based semantic-ontological model for analyzing Russian-language scientific texts based on the VAK specialty classifications. The model is then compared with large language models for the purposes of specialty code identification and subject-matter analysis. After compiling a corpus of classifications and accompanying texts, a unified normalization pipeline is developed, lexical and semantic feature layers are defined to produce an integrated information morphism metric, and validation is carried out using abstracts of applicants and articles from journals on the VAK List.[11]

The work is organized as follows. The first section provides an overview of works on semantic text classification, ontology-oriented models, and graph representations. The second section describes the ontological foundation of the study and the data corpus. The third section introduces a graph semantic-ontological feature model based on TF-IDF and ICAN. The fourth section formalizes information morphism metrics and a probabilistic interpretation of the results. The fifth section presents the results of validation on abstracts and scientific articles and a comparison with large language models. The sixth section discusses the obtained results, formulates conclusions, and formulates directions for further research.

1. LITERATURE REVIEW

The review paper [1] systematizes approaches to semantic text classification and compares them with traditional methods based on the bag-of-words (BoW) method. Demonstrating the limitations of BoW vector representation (high dimensionality, sparsity, ignoring synonymy and polysemy), the authors identify five main semantic methods: ontology-based, corpus-based, deep learning models, word/symbol sequence-based, and linguistically enriched approaches. The review summarizes the results of numerous experiments and demonstrates the advantage of semantic models in classification accuracy.

The paper [2] analyzes knowledge extraction methods in the context of the Semantic Web, with an emphasis on ontology-based feature selection. The use of ontologies to represent and select features in classification problems reduces dimensionality and increases the interpretability of models. However, the quality of the approach depends significantly on the completeness and consistency of the ontologies.

In [3–5], ontologies and semantic technologies are considered as a key tool for structuring engineering and scientific knowledge. In [3], an ontology-oriented approach to the automatic classification of engineering standards is proposed, using domain ontologies to map document fragments to business units.

In [4], a two-stage method for describing predefined information flows in standards and directives is developed, including a generalized data model and a machine-readable representation, which increases the accessibility and reusability of information in digital product models. The work [5] contains a bibliometric and semantic analysis of research on ontologies and the Semantic Web, highlighting the main areas related to dynamic updating of ontologies and scalability in the context of the Big Data.

The paper [6] presents a review of ontology-oriented methods for text classification, in which ontologies are considered as a formal basis for enriching vector representations and deep learning architectures through explicit modeling of domain concepts and their relationships. The study [7] proposes an ontology-oriented approach to extracting contextualized information from scientific publications. Based on transformer models, a two-stage pipeline is implemented, including sentence classification and entity recognition (research activities and methods). Relationships between activities and methods are then extracted via integration of the obtained data and publication metadata in the form of a resource description framework (RDF) graph.

A number of works [8–10] demonstrate the potential of lexical-semantic representations of text for solving various analysis problems. In [8], special lexical-semantic parameters for determining sentiment are introduced. These are obtained through the semantic expansion of sentiment lexicons and distributed representations, which unifies data dimensionality to improve classification quality.

The authors of the work [9] propose a model of lexical-semantic connections between documents and a statistical clustering algorithm based on cosine

[11] List of peer-reviewed scientific publications in which the main scientific results of dissertations for the degree of candidate of sciences and for the degree of doctor of sciences should be published. https://vak.gisnauka.ru/documents/editions. Accessed April 04, 2026. (In Russ.).

proximity in the lexical-semantic space, comparable in quality to the Affinity Propagation method[12].

The work [10] uses a set of lexical-semantic patterns and procedures for matching textual mentions with domain ontology objects for their semi-automatic replenishment. Taken together, these approaches confirm the effectiveness of lexical-semantic features.

In [11], researchers address the problem of multi-class classification of imbalanced text data using the WordNet[13] lexical ontology and the bidirectional encoder representation from transformers (BERT)[14] model. The WordNet lexical ontology is used for ontology-based feature dimensionality reduction, followed by classification using traditional algorithms and a pre-trained BERT model. Experiments on a seven-class corpus show that a combination of WordNet and BERT provides the highest accuracy (up to 93.77%) as compared with variants without ontological feature refinement.

A number of works [12–15] systematically investigate graph representations of text and graph neural network models for classification problems. In [12], a large-scale comparison of various graph construction schemes and graph neural network (GNN) architectures with transformer models is conducted. In [13], researchers analyze how the choice of a graph text representation strategy affects the quality of GNN models. Works [14] and [15] develop the use of graph models in text content analysis architectures, focusing on improving interpretability and classification quality. In [14], a framework is proposed in which the graph representation of text is supplemented with symbolic feature selection, while in [15], a hierarchical graph framework is developed that combines linguistic features, domain ontology, multilayer GNN training, attention mechanisms, and dynamic fusion with BERT.

The concept of information ontological modeling proposed in [16] considers information retrieval and clustering methods as stages of the formation of ontological models and the extraction of implicit knowledge.

In [17], the authors investigate the dependency of the topological characteristics of a word graph on the text genre. Texts of different genres are represented as graphs, in which vertices correspond to words, while edges correspond to their neighborhood in bigrams. The parameters of such networks, which are shown to systematically vary between genres, can be used to characterize these genres as well as to identify subgenres. In [18], the ICAN model is used to construct semantic graphs at the level of individual documents, thus allowing word order to be taken into account, which is important for semantic analysis.

2. ONTOLOGICAL BASIS AND DATA CORPUS

2.1. Passports of VAK specialties

In the Russian Federation's scientific personnel certification system, the texts of the VAK scientific specialty passports serve as normative descriptions of subject areas. Each passport is assigned to a specific specialty code to define a set of typical research objects, problem classes, applied methods, and areas of application of the results. These documents have supradepartmental status and serve as the formal basis for classifying dissertations, abstracts, and publications within a specific scientific field.

In terms of content, passports have a stable internal structure, in which, as a rule, the following blocks are distinguished:

- list of main areas and objects of research;
- typical scientific tasks and methods;
- related areas of knowledge.

This structure enables the use of specialty passports not only as regulatory documents but also as textual representations of subject area ontologies. Here, ontology is understood as an explicitly defined system of concepts and relationships that delimits the permissible subject space.

The set of specialty passports can be interpreted as a finite set of ontological objects:

$$O = \{s_1, \ldots, s_N\}, \quad (1)$$

where each element s_i corresponds to one scientific specialty, and the text of the passport defines its conceptual content.

The hierarchical structure of codes (aggregated fields, groups, specific specialties) defines a partial order on the set O to capture inclusion and proximity relationships between fields. An ontological space of scientific specialties is introduced in which the level of aggregated fields corresponds to more general classes, while individual codes are used to define specialized subdomains.

Further, a key methodological assumption is adopted. The passport text is viewed as a prototype of the corresponding subject area. Lexemes and set phrases are interpreted as superficial realizations of the concepts and relations of this area, while the thematic coherence of these elements is reflected in the structure of sections. This enables the corpus of scientific specialty passports

[12] A clustering algorithm in machine learning based on the concept of messaging between data points.

[13] Linguistic ontology, an electronic thesaurus that represents the system of meanings of words in the generally significant English language in the form of a hierarchical structure.

[14] Bidirectional encoder representation from transformers (BERT) is a neural network designed to improve natural language understanding. BERT's key difference from previous models is its bidirectional context understanding.

to be used to construct a supporting semantic-ontological feature space that describes both the passports themselves and the documents being analyzed.

At the same time, passports are not formal ontologies in the strict logical sense and as such cannot be used to define a system of axioms or strict constraints. However, due to their official status, structural stability, and focus on descriptions of objects, tasks, and methods, they can be considered as a practically acceptable ontological basis for the information analysis of scientific texts.

In what follows, this basis will be used to construct a semantic-ontological model of features, as well as to determine the information morphism between documents and the space of scientific specialties.

2.2. Collection and storage of text data

The ontological foundation corpus is formed from the scientific specialty passports approved by the VAK. These specialty passports are collected using specialized software parsers that automatically crawl linked web pages of official resources, download data sheet files and associated documents (in HTML and PDF formats), and capture technical metadata.

For PDF documents, the presence of a text layer is checked. If a text layer is present, extraction is performed directly; otherwise, optical character recognition (OCR) is used, followed by basic cleanup of the results.

At this stage, the text is not yet linguistically normalized, but obvious artifacts (random binary insertions, incorrect encodings) are removed to ensure the correctness of further processing. A detailed normalization algorithm is described below.

For each passport, the following fields are stored, in particular:

- unique record identifier,
- specialty code,
- name,
- uniform resource locator (URL) original source,
- collection time,
- file hash value,
- extracted text after primary cleaning.

The primary passport corpus is generated as JSONL[15] files (one object per line), which can be conveniently used for archiving and data exchange. For long-term storage and efficient access, the data is transferred to the PostgreSQL[16] relational database management system.

For specialty passports, a separate table is created that includes the code and name of the specialty, source metadata, file checksum, and normalized passport text used in constructing features.

For input documents (scientific articles and dissertation abstracts), a corresponding table is created in which, in addition to the text content and document type, the source code of the specialty (for dissertation abstracts), if available, is recorded, since it is used as an expert assessment when assessing the quality of the model.

2.3. Text normalization

Text normalization plays a key role in the proposed model, since it is at this stage that a single feature space is defined, in which scientific texts will subsequently be compared with the passports of scientific specialties.

The processing of text content is deliberately organized as common to all categories of documents in such a way that differences in features reflect the substantive features of the texts, rather than artifacts of format, layout, or source.

The first stage involves technical cleaning of the extracted text. After extracting content from HTML pages or PDF files (using OCR if there is no text layer), incorrect encodings, binary inserts, and fragments of service markup are removed. Line breaks are then converted to regular spaces. For PDF documents, duplicate headers and footers are automatically detected and removed.

Next, service blocks are filtered. Clearly editorial and publishing fragments (References, Keywords, Publisher Information, Author Contact Information, Author Information, etc.) are removed from the texts, as well as technical elements such as URLs, e-mail addresses, and numeric identifiers (ORCID[17], Scopus ID[18], etc.).

The next step is text conversion. All text is converted to lowercase, the “ё” character is replaced with “е”, hyphens are eliminated, and extra spaces are removed. Punctuation and all special characters are removed. Particular attention is paid to hyphenated terms, as a splicing dictionary is used to maintain terminological integrity. According to this dictionary, compound words like object-oriented are converted to underscored (object_oriented) forms. This enables stable compound terms to be considered as single lexical units in subsequent analysis.

Next, numeric expressions are processed. If a token consists solely of digits, it is converted to its Russian verbal form (for example, “2025” is converted to “two thousand twenty-five”), which prevents the dictionary from growing too large due to the large number of unique

[15] JSON Lines is a text format in which each line contains one valid JSON object (JavaScript object notation).

[16] https://www.postgresql.org/. Accessed April 04, 2026.

[17] Open researcher and contributor identifier.

[18] Unique author identifier in the Scopus database. https://www.scopus.com/. Accessed April 04, 2026.

numbers while preserving the semantic information about quantitative characteristics. However, technical standards, indices, and designations (e.g., “5g”, “3d”, “h264”, “ipv6”) are stored as separate tokens and normalized to lowercase, as they are important domain markers.

The linguistic normalization stage involves tokenization, lemmatization, and stopword filtering. The text is broken down into tokens, taking into account previously established rules for punctuation and hyphenation replacement. A normal form (lemma) is determined for each token. Where necessary, common abbreviations and unit designations are standardized, and domain-specific abbreviations are converted to standard forms.

The final result of normalization is a stable text representation as a sequence of lemmas separated by spaces, excluding punctuation, numbers, and function words. If necessary, a list of tokens with positional ranges in the source text is additionally generated, thus enabling the subsequent interpretation of the contribution of individual fragments to the final proximity scores. It is this normalized sequence of lemmas that is used to construct features in the lexical (TF-IDF) and semantic (ICAN) layers of the model.

3. GRAPH SEMANTIC-ONTOLOGICAL MODEL OF FEATURES

3.1. Passport chunks and *n*-grams

The normalized texts of scientific specialty passports (see 2.3) are further considered as sequences of lemmas w:

$$d_s = (w_1, ..., w_{T_s}), \quad (2)$$

where T_s is the length of passport s in tokens after pre-processing.

Since the specialty passports consist of heterogeneous sections (description of objects, tasks, methods, interdisciplinary connections), then representing the entire text with one vector leads to the fact that weakly related fragments are averaged, which reduces the sensitivity of the model to thematic differences within the passport.

To increase the model’s sensitivity to local areas of semantic concentration, each passport is divided into fragments of a fixed length of L token chunks. The number of chunks for a passport s is defined as follows:

$$n_s = \left\lceil \frac{T_s}{L} \right\rceil. \quad (3)$$

Each chunk $c_{s,i}$ inherits the passport metadata (code and name of the specialty), as well as storing the positional range $[a_i, b_i]$ in the source text. This enables the subsequent interpretation of the contribution of individual fragments to the final assessments.

The choice of length L is determined by a trade-off between local sensitivity and stability of the representation.

Chunks that are too short lead to excessive sparsity in the vector space, which increases the impact of markup artifacts and OCR errors on each individual fragment. Chunks that are too long, on the other hand, make local similarity metrics virtually equivalent to their global equivalents, effectively negating the effect of text chunking.

In the conducted experiments, the values of $L \in \{100, 200, 300\}$ were analyzed; the length $L = 200$, which demonstrates a balance between the number of fragments per passport, noise resistance and computational cost, is further used as the default value (Figure).

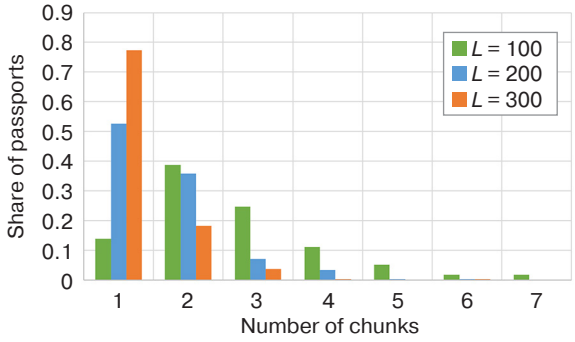


Figure. Distribution of the number of chunks n_s by passports at $L \in \{100, 200, 300\}$

Chunking achieves two goals. Firstly, it localizes specialized vocabulary. Specifically, within a fragment of hundreds of tokens, the share of terms characteristic of a specific field increases, leading to sharper cosine similarities when compared with documents. Secondly, it increases resilience to structural and layout artifacts, in which individual technical inserts or remnants of headers and footers only affect their local vectors and do not dominate the global assessment.

Therefore, the presence of at least two or three fragments in a significant share of passports is important for the local metric, since it increases the probability that at least one fragment will be close to the document in question, while making the aggregate by top N more sensitive to thematic matches.

Next, a lexical dictionary of n-grams is constructed on the set of all passport chunks $D = \{c_1, \ldots, c_N\}$. For each chunk $c \in D$, there is a sequence of lemmas $(w_1, \ldots, w_{T(c)})$.

Two levels are considered:

- set of unigrams $U(c) = \{w_i\}_{i=1}^{T(c)}$,
- set of bigrams $B(c) = \{(w_i, w_{i+1}\}_{i=1}^{T(c)-1}$.

The use of bigrams is essential for capturing stable terminological combinations (neural network, Internet network, communication network, etc.), which clarify the semantics of polysemous lemmas and serve as more reliable markers of subject areas than individual words.

For each term t (both unigrams and bigrams), the number of corpus chunks in which this term occurs at least once is calculated:

$$df(t) = |\{c \in \mathrm{D}: t \in c\}|. \quad (4)$$

Next, threshold restrictions are introduced for the minimum and maximum number of occurrences:

$$df(t) \geq d_{\mathrm{MIN}}, \frac{df(t)}{N} \leq d_{\mathrm{MAX}}, \quad (5)$$

where d_{MIN} sets the lower limit of the number of chunks in which the term must occur, which allows for the preservation of rare but substantively significant lexemes, while d_{MAX} sets the upper limit of the relative prevalence of the term in the corpus, excluding excessively frequent, uninformative units from the dictionary.

The resulting dictionary V is defined as the set of terms that satisfy the specified conditions:

$$\mathrm{V} = \left\{ t \in \mathrm{V}' : df(t) \geq d_{\mathrm{MIN}}, \frac{df(t)}{N} \leq d_{\mathrm{MAX}} \right\}, \quad (6)$$

where V′ is the set of all unigrams and bigrams extracted from the corpus of chunks.

Thus, in the semantic-ontological model of features of the specialty passport, a basic space of n-grams V is defined in which TF-IDF representations of both the passports themselves (at the level of chunks and centroids) and the input documents can then be constructed.

The use of chunking and a combined dictionary of unigrams and bigrams allows us to combine local thematic sensitivity with noise resistance and ensure the interpretability of the resulting feature vectors.

3.2. Lexical layer

At the lexical level, each chunk of a specialty passport is considered as a document in a general dictionary of n-grams V, constructed from the passport corpus (see Section 3.1). For each term $t \in \mathrm{V}$ and chunk c, the occurrence frequency $\mathrm{tf}(t, c)$ is calculated, after which a weighted representation is formed using the TF-IDF algorithm.

To account for repeated occurrences of terms in a document, the sublogarithmic term frequency (sublogarithmic TF) is used:

$$\mathrm{tf}(t,c) = \begin{cases} 1 + \ln(\mathrm{tf}(t,c)), & \text{at } \mathrm{tf}(t,c) > 0, \\ 0, & \text{at } \mathrm{tf}(t,c) > 0. \end{cases} \quad (7)$$

The use of sublogarithmic TF allows the contribution of frequently repeated terms within a single fragment to be reduced: since each subsequent occurrence increases their weight with decreasing increment, the influence of local abnormally high frequencies is lowered along with the dependence of the score on the length of the text fragment.

The inverse frequency of the IDF is calculated using the smoothed formula:

$$\mathrm{idf}(t) = \log \frac{1+N}{1+df(t)} + 1, \quad (8)$$

where N is the total number of chunks in the corpus of passports, $df(t)$ is the number of chunks in which the term t occurs at least once.

Adding one to the numerator and denominator ensures numerical stability at extreme values of $df(t)$, while a shift of one keeps the weights positive even for terms that appear in all documents. Moreover, the $\mathrm{idf}(t)$ function remains monotonically decreasing as $df(t)$ increases, with rare terms receiving higher weights and frequently occurring terms receiving lower weights.

The weight of a term t in a chunk c is given by the product:

$$\mathrm{w}_t(c) = \mathrm{tf}(t, c) \cdot \mathrm{idf}(t). \quad (9)$$

The vector $\mathbf{w}(c)$, the values of which are given by the quantities $\mathrm{w}_t(c)$, is interpreted as an ontological vector, where each coordinate corresponds to a lexeme or stable phrase associated with an ontology element (a concept, property, or typical relation in the subject area).

Thus, the lexical layer captures the contribution of ontological markers that appear at the level of terms and n-grams.

To ensure comparability of vectors, L_2 normalization is performed:

$$\mathbf{x}(c) = \frac{\mathbf{w}(c)}{\| \mathbf{w}(c) \|_2}, \quad (10)$$

where $\|\mathbf{w}(c)\|_2$ denotes the length of the vector.

Following L_2 normalization, the consideration of the total length and volume of text when comparing documents is significantly reduced; here, the direction of the vector is determined by the relative distribution of term weights.

Since the TF-IDF values are non-negative, the scalar product of two L_2-normalized vectors coincides with the cosine similarity and lies in the range [0; 1], which is convenient for subsequent interpretation of the results.

The normalized TF-IDF vectors of all passport chunks form a sparse matrix

$$\mathbf{X} \in \mathbb{R}^{M \times |\mathrm{V}|}, \quad (11)$$

where $M = \Sigma_s n_s$ is the total number of chunks for all specialties, $|\mathrm{V}|$ is the dictionary size.

The matrix is stored in the compressed sparse row (CSR) format, which stores the array of nonzero weights, indices of the corresponding terms, and row boundary pointers separately. Using CSR format ensures compact storage, as well as enabling the $v^T\mathbf{X}$ product to be calculated in time proportional to the number of nonzero components of the vector v and nonzero elements of the matrix **X**.

Subsequently, input documents (scientific articles, dissertation abstracts) are represented in the same dictionary V and with the same idf(t) values. Their normalized TF-IDF vectors are compared with chunk vectors and aggregated passport representations, with cosine similarities serving as the basic information-ontological metrics in the proposed model.

3.3. Semantic layer

While the lexical layer captures the coincidences of terms and set phrases, it remains sensitive to variability in wording, synonymy, and abbreviations. To account for contextual relationships between lemmas, a semantic layer is introduced based on the ICAN graph model, in which each text is described as a local term co-occurrence graph [19].

Let the normalized text of a document (a chunk of a passport, article, or abstract) be given by a sequence of lemmas $d = (t_1, \ldots, t_T)$, and the set of unique lemmas of this text be designated as $W = \{u_1, ..., u_{m_d}\}$. Then, a directed weighted graph G_d is constructed for the text on the vertices W, whose structure is given by an adjacency matrix $\mathbf{M} \in [0;1]^{m_d \times m_d}$, initialized to zeros.

The graph is formed by traversing the text with a sliding window of fixed width W (by default, W = 11 tokens). At each position in the window, the central token x is selected, while all other tokens in the window are considered as its contextual neighbors y. The matrix **M** is updated in three stages.

In the first stage, direct connections between the central token and its context (contextual neighbors to the left and right of the central token x) are strengthened.

If the connection $M_{xy} = 0$, then the connection is initialized with a base weight of 0.5, and upon repeated occurrences in the same context relationship, the weight smoothly increases according to formula (12) to ensure monotonous growth at the same time as limiting the value of the weight.

$$M_{xy} = M_{xy} + \frac{1}{2}\left(1 - M_{xy}\right). \quad (12)$$

In the second stage, second-order indirect connections are taken into account. If token y is already connected to the token k (i.e., $M_{yk} > 0$), then a weakened contribution is added to the pair (x, k):

$$M_{yk} = M_{xk} + A(1 - M_{xk})M_{xy}M_{yk}, \quad (13)$$

where $A \ll 1$ is the scaling factor.

As a result, the resulting graph reflects not only the co-occurrence of terms in a single sliding window, but also connections through common contextual neighbors, which enables the capture of broader contextual associations.

In the third stage, a damping and thresholding operation is applied to the matrix **M**. All weights are multiplied by a coefficient $\gamma \in (0, 1)$, and elements with a value below the threshold θ are set to zero:

$$M_{xy} = \gamma M_{xy}, \quad (14)$$

where $M_{xy} = 0$ for $M_{xy} \le \theta$, $\gamma = 0.9$ is the attenuation coefficient, $\theta = 0.4$ is the threshold coefficient for removing the connection.

This procedure suppresses random weak connections to form a more stable graph structure that reflects stable contextual connections.

Ultimately, the semantic representation of the text in the ICAN space is defined by a vector of vertex degrees. For each lemma $u_i \in W$, its degree (the sum of the weights of its outgoing and incoming edges) is calculated):

$$k_i = \Sigma_j M_{ij} + \Sigma_j M_{ji}. \quad (15)$$

Vector $\mathbf{k}(d) = (k_1, ..., k_{m_d})$ reflects the relative importance of lemmas in the contextual structure of the text in which high values correspond to terms that play the role of "nodal points" of the semantic graph.

To integrate with the lexical layer and ontological foundation, the vector $\mathbf{k}(d)$ is projected onto the common basis of the dictionary V constructed from the passport corpus. If $V = \{v_1, \ldots, v_{|V|}\}$, then the semantic vector $\mathbf{s}(d) \in \mathbb{R}^{|V|}$ is determined by the formula:

$$s_j(d) = \begin{cases} k_i, & \text{if lemma } u_i \text{ coincides with } v_j, \\ 0. \end{cases} \quad (16)$$

Next, L_2 normalization is performed:

$$\hat{\mathbf{s}}(d) = \frac{\mathbf{s}(d)}{\| \mathbf{s}(d) \|_2}, \quad (17)$$

which makes semantic vectors comparable in scale and enables the use of cosine similarity.

4. METRICS OF INFORMATION MORPHISM

Along with the lexical (TF-IDF) layer (previously defined in Section 3.2), the semantic (ICAN) layer (previously defined in Section 3.3), defines two consistent feature spaces, in each of which a document can be compared with scientific specialty datasets. In both cases, two groups of metrics are used:

- local, measuring the proximity of the document to individual fragments of the passport;

- global, characterizing the proximity of the document to the centroid of the passport, obtained by averaging over all its fragments.

These metrics are introduced in a standardized form and then applied separately in each layer.

Let a normalized feature vector $\mathbf{v}(d)$ be constructed for the input document d, and let a normalized vector $\mathbf{u}_{s,i}$, where $i = 1, \ldots, n_s$, be constructed for each chunk $c_{s,i}$ of the passport s. All vectors are normalized and non-negative.

Then the local metric is defined as the cosine similarity of the document with a separate passport chunk vector according to the formula:

$$\text{CosSimilarity}(d, c_{s,i}) = \langle \mathbf{v}(d), \mathbf{u}_{s,i} \rangle \in [0; 1]. \quad (18)$$

For each passport s, this defines a set of local assessments $\text{CosSimilarity}(d, c_{s,i})_{i=1}^{n_s}$, which reflect how closely the document is comparable in content to individual fragments of the passport text.

Based on these estimates, an aggregated local metric MaxSim_k, is introduced, taking into account only a few of the largest matches

$$r_{s,i} = \text{CosSimilarity}(d, c_{s,i}), \; i = 1, \ldots, n_s, \quad (19)$$

and by $r_{s,1} \geq \ldots \geq r_{s,n_s}$ are the values ordered in descending order. Then the local estimate for s is given as:

$$\text{MaxSim}_k(d,s) = \frac{1}{k_s} \sum_{j=1}^{k_s} r_{s,j}, \; k_s = \min(k, n_s). \quad (20)$$

Thus, MaxSim_k reflects the presence of multiple fragments in a passport that are as close as possible to the document's content. The choice of the parameter k provides a compromise between sensitivity to narrow matches and robustness to noise, while in the experimental section, a value of three ($k = 3$) is used.

The global metric is formed using the passport centroid in the layer under consideration. For passport s, the average vector is determined by the formula:

$$\bar{\mathbf{u}}_s = \frac{1}{n_s} \sum_{i=1}^{n_s} u_{s,i}. \quad (21)$$

After this, normalization is performed; here, the global measure of closeness of document d to the passport s is given by the cosine similarity with the centroid according to the formula:

$$\text{CentroidCos}(d,s) = \left\langle \mathbf{v}(d), u_s \right\rangle = \frac{\sum_{i=1}^{n_s} \left\langle \mathbf{v}(d), u_s \right\rangle}{\left\| \sum_{i=1}^{n_s} u_s \right\|_2} \in [0;1]. \quad (22)$$

As a result, in each layer, documents and specialty passports are linked by a pair of complementary metrics: $\text{MaxSim}_k(d, s)$ is a local metric sensitive to the most relevant fragments of the passport; $\text{CentroidCos}(d, s)$ is a global metric reflecting compliance with the general topic.

4.1. Hybrid model and integral metric

Local and global metrics defined for the lexical and semantic layers provide consistent estimates of the closeness of a document d to a passport s.

These metrics are then combined into a single scalar value within the corresponding layer, after which the results of the lexical and semantic layers are aggregated into a final hybrid metric.

Within each layer, the local and global metrics are combined using the formula:

$$S(d, s) = (1 - \alpha)\text{CentroidCos}(d, s) + \alpha\text{MaxSim}(d, s), \quad (23)$$

where coefficient α sets the balance between $\text{CentroidCos}(d, s)$, indicating the proximity to the passport as a whole, and $\text{MaxSim}(d, s)$, which highlights the closest fragments of the passport.

In the TF-IDF lexical layer, increasing α enhances the contribution of exact terminological matches. This is due to the fact that, at $\alpha > 0.5$, the influence of local maxima of MaxSim, which arise in those passports that contain fragments with a high concentration of matching n-grams, increases. The value of $\alpha = 0.6$ was chosen empirically based on the validation results as providing the best quality of classification and recommendation ranking.

The resulting $S(d, s)$ values for TF-IDF and ICAN are interpreted as two independent but consistent assessments of document compliance with the passport. The first interpretation consists in terms of precise lexical markers, while the second is considered in terms of contextual associations. A hybrid metric is introduced to form the final scalar score:

$$S(d, s) = (1 - \lambda)S_{TF}(d, s) + \lambda S_{ICAN}(d, s), \lambda \in [0; 1], \quad (24)$$

where the parameter λ regulates the contribution of the semantic layer relative to the lexical layer.

At λ approaches 1, ICAN graph representations dominate, increasing robustness to paraphrases and abbreviations. In the experiments conducted, the values of $\alpha_{TF} \approx 0.6$, $\alpha_{ICAN} \approx 0.5$, $\lambda \approx 0.5$ were empirically selected to provide a balance between accuracy and robustness on abstract and article corpora.

From a computational point of view, all components of the metric $S(d, s)$ are obtained from two vector-matrix operations of the form $\mathbf{r} = \mathbf{X} \times \mathbf{v}(d)$, where $\mathbf{X}$ is a matrix of chunk vectors, while $\mathbf{r}$ is a vector of cosine similarities with all chunks of all passports.

Formally, from a conceptual point of view, the value of S(d, s) can be interpreted as the intensity of information morphism from document d to ontological entity s: the higher the S(d, s), the more consistently the content of the text reproduces the lexical-semantic profile of the corresponding specialty.

4.2. Evaluation of probabilistic morphisms

The integral metric S(d, s) defines a scalar conformity score for each document d and specialty passport $s \in$ O. In order to transition from a set of such scores to a formal information morphism, it is necessary to construct a probability distribution over the ontological set of passports O.

Let the values $S(d,s)_{s \in O}$ be calculated for a fixed document d. Then a probability distribution is introduced based on them using the softmax scheme with temperature $\tau > 0$:

$$P(s \mid d, \tau) = \frac{e^{(S(d,s)-M)/\tau}}{\sum_{q \in O} e^{(S(d,s)-M)/\tau}}, \quad (25)$$

where $M = \max_{q \in O} S(d,q)$ is used for numerical stabilization (log-sum-exp normalization). Subtracting M does not affect the relative probabilities, but prevents overflow when carrying out exponential transformations.

Parameter τ determines the degree of concentration of the probability distribution:

- when $\tau \ll 1$, distribution becomes more concentrated, since even small differences in the values of S(d, s) lead to a pronounced dominance of one or more passports;
- when $\tau \approx 1$, standard temperature scaling is applied, in which softmax preserves the typical normalization behavior of estimates;
- when $\tau > 1$, the distribution becomes smoother; this is convenient when analyzing interdisciplinary and borderline texts, which are characterized by the presence of several estimates comparable in magnitude.

The resulting distribution $P(s \mid d, \tau)$ defines a mapping:

$$\mu: D \to \Delta(O), \mu(d) = P(s \mid d, \tau), \quad (26)$$

where D is a set of documents, and Δ(O) is a simplex of probability measures on a set of specialty passports.

Here μ is interpreted as an information morphism from the text space to the ontological space O, where each document is matched with the distribution of its thematic contribution across normatively defined areas.

The point solution of the classification problem corresponds to the passport with the highest probability:

$$\hat{s}_1(d) = \arg\max_{s \in O} P(s \mid d, \tau). \quad (27)$$

5. EXPERIMENTAL EVALUATION

5.1. Results of model validation based on applicants' abstracts

The model was validated using a corpus of 124 abstracts submitted by applicants. The corpus covers sixteen dissertation councils from five organizations: MIREA – Russian Technological University (RTU MIREA)[19], V.F. Utkin Ryazan State Radio Engineering University[20], Federal Research Center for Informatics and Management of the Russian Academy of Sciences[21], V.A. Trapeznikov Institute of Control Sciences of the Russian Academy of Sciences[22], and Patrice Lumumba Peoples' Friendship University of Russia (RUDN)[23].

For each council, abstracts on codes falling within its area of competence were taken into account, for example, 2.3.2, 2.3.5, 2.3.8 for council D24.2.326.09. The official code of the specialty specified in the dissertation council data was taken as the reference mark for each document.

For three configurations of the lexical TF-IDF, semantic ICAN, and hybrid models, two metrics were evaluated. The first metric is top 1 accuracy, i.e., the share of abstracts for which the passport with the maximum value S(d, s) coincides with the reference code. The second metric is top 3 accuracy, i.e., the share of abstracts for which the reference code is among the top three passports with the highest S(d, s) values.

The aggregation of values across the entire corpus demonstrates the consistent advantage of the hybrid model. The average top 1 accuracy across all councils and documents is approximately 0.58 for TF-IDF, approximately 0.57 for ICAN, and approximately 0.69 for the hybrid configuration. For top 3 accuracy, the picture is even more pronounced. The TF-IDF model gives an average of about 0.74, while ICAN gives around 0.81, and the hybrid model reaches about 0.90. That is to say, in nine out of ten cases, the correct passport is among the top three closest in value to S(d, s).

A review of the results for individual organizations shows a similar picture. In the RTU MIREA samples (Table 1), where councils D24.2.326.09, D24.2.326.03, D24.2.326.08, and D24.2.326.10 were analyzed, the average top 1 accuracy for abstracts is

[19] https://www.mirea.ru/. Accessed April 04, 2026. (In Russ.).
[20] https://rsreu.ru/. Accessed April 04, 2026. (In Russ.).
[21] https://www.frccsc.ru/. Accessed April 04, 2026. (In Russ.).
[22] https://www.ipu.ru/. Accessed April 04, 2026. (In Russ.).
[23] https://www.rudn.ru/. Accessed April 04, 2026. (In Russ.).

approximately 0.54 for TF-IDF and about 0.63 for ICAN. The hybrid model increases this indicator to 0.75.

According to the top 3 metric, the hybrid configuration reaches about 0.96, i.e., in almost all cases, the correct specialty code is among the three most probable.

At V.F. Utkin Ryazan State Radio Engineering University (Table 2), the lexical model for councils D24.2.375.01, D24.2.375.02, D24.2.375.03, and D99.2.113.02 demonstrates an average top 1 accuracy of 0.68, while the corresponding value for ICAN is about 0.61.

At the same time, the semantic layer provides higher accuracy in the top 3, i.e., approximately 0.89 compared to 0.78 for TF-IDF.

The hybrid model combines the advantages of both configurations to achieve an average top 1 accuracy of around 0.75 and top 3 accuracy of around 0.93.

At the Federal Research Center for Informatics and Management of the Russian Academy of Sciences (Table 3), the average top 1 accuracy for councils D24.1.224.04, D24.1.224.03, and D24.1.224.02 is about 0.60 for TF-IDF and about 0.50 for ICAN. The hybrid model increases this indicator to 0.70.

At the same time, according to the top 3 TF-IDF and ICAN metrics, similar values of around 0.83 are achieved, while the hybrid configuration reaches a level of around 0.90.

Table 1. Results of the model for authors' abstracts at RTU MIREA

Council D24.2.326.09			Council D24.2.326.03		
Scientific specialties 2.3.2, 2.3.5, 2.3.8			Scientific specialties 1.4.7, 1.4.10		
Method	Accuracy according to top 1	Accuracy according to top 3	Method	Accuracy according to top 1	Accuracy according to top 3
TF-IDF	0.6	0.7	TF-IDF	0.5	0.63
ICAN	0.7	0.7	ICAN	0.5	0.75
Hybrid	0.8	1	Hybrid	0.63	0.88
Council D24.2.326.08			Council D24.2.326.10		
Scientific specialties 1.2.2, 2.3.1			Scientific specialty 5.2.3		
Method	Accuracy according to top 1	Accuracy according to top 3	Method	Accuracy according to top 1	Accuracy according to top 3
TF-IDF	0.33	0.67	TF-IDF	0.67	0.67
ICAN	0.67	0.67	ICAN	0.67	1
Hybrid	0.67	1	Hybrid	1	1

Table 2. Results of the model for authors' abstracts in V.F. Utkin Ryazan State Radio Engineering University

Council D24.2.375.01			Council D24.2.375.02		
Scientific specialties 2.3.1, 2.3.5			Scientific specialties 1.3.2, 1.3.11, 2.2.1		
Method	Accuracy according to top 1	Accuracy according to top 3	Method	Accuracy according to top 1	Accuracy according to top 3
TF-IDF	0.67	0.67	TF-IDF	0.67	0.83
ICAN	0.67	1	ICAN	0.5	0.83
Hybrid	0.67	0.67	Hybrid	0.5	0.83
Council D24.2.375.03			Council D99.2.113.02		
Scientific specialties 2.2.11, 2.2.12, 2.2.13			Scientific specialty 2.3.8		
Method	Accuracy according to top 1	Accuracy according to top 3	Method	Accuracy according to top 1	Accuracy according to top 3
TF-IDF	0.69	0.81	TF-IDF	0.67	0.67
ICAN	0.63	0.94	ICAN	0.67	0.67
Hybrid	0.81	1	Hybrid	1	1

Table 3. Results of the model for authors' abstracts at the Federal Research Center for Informatics and Management of the Russian Academy of Sciences

Council D24.1.224.04			Council D24.1.224.03			Council D24.1.224.02		
Scientific specialties 2.3.2, 2.3.5, 2.3.6			Scientific specialties 1.2.1, 1.2.3, 2.3.8			Scientific specialties 1.1.2, 1.1.6, 1.1.9		
Method	Accuracy according to top 1	Accuracy according to top 3	Method	Accuracy according to top 1	Accuracy according to top 3	Method	Accuracy according to top 1	Accuracy according to top 3
TF-IDF	0.67	1	TF-IDF	0.56	0.78	TF-IDF	0.67	0.89
ICAN	0.67	1	ICAN	0.39	0.83	ICAN	0.67	0.78
Hybrid	1	1	Hybrid	0.61	0.89	Hybrid	0.78	0.89

At the V.A. Trapeznikov Institute of Control Sciences of the Russian Academy of Sciences (Table 4), where the corpus includes five abstracts from council D24.1.107.02, the results should be interpreted with caution due to the small sample size. In this group, the ICAN semantic model gives the best top 1 accuracy of about 0.60 compared to 0.40 for TF-IDF.

While the hybrid configuration maintains a top 1 level of around 0.60, it does not surpass the semantic branch in terms of top 3.

At RUDN (Table 5), the average top 1 TF-IDF and ICAN accuracy for four councils PDS 0300.004, PDS 2028.001, PDS 0900.006, and PDS 0200.002 is close, amounting to approximately 0.54.

The hybrid model increases this indicator to 0.62. For the top 3 lexical and semantic configurations, the values are around 0.70 and 0.78, while the hybrid model reaches about 0.89. Significant gains are observed on individual councils.

Table 4. Results of the model for authors' abstracts at the V.A. Trapeznikov Institute of Control Sciences of the Russian Academy of Sciences

Council D24.1.107.02		
Scientific specialties 2.3.1, 2.3.4		
Method	Accuracy according to top 1	Accuracy according to top 3
TF-IDF	0.4	0.6
ICAN	0.6	0.8
Hybrid	0.6	0.6

For PDS 2028.001, the hybrid model improves the top 3 accuracy from 0.56 for TF-IDF and 0.78 for ICAN to unity. For PDS 0200.002, the hybrid configuration also increases the top 3 accuracy to unity while maintaining high top 1 accuracy.

Table 5. Results of the model for authors' abstracts at RUDN University

Council PDS 0300.004			Council PDS 2028.001		
Scientific specialties 3.1.18, 3.1.20, 3.3.6			Scientific specialties 5.8.1, 5.8.7		
Method	Accuracy according to top 1	Accuracy according to top 3	Method	Accuracy according to top 1	Accuracy according to top 3
TF-IDF	0.56	0.75	TF-IDF	0.44	0.56
ICAN	0.44	0.81	ICAN	0.44	0.78
Hybrid	0.56	0.81	Hybrid	0.56	1
Council PDS 0900.006			Council PDS 0200.002		
Scientific specialty 5.1.4			Scientific specialties 1.4.1, 1.4.3, 1.4.4		
Method	Accuracy according to top 1	Accuracy according to top 3	Method	Accuracy according to top 1	Accuracy according to top 3
TF-IDF	0.4	0.6	TF-IDF	0.71	0.86
ICAN	0.8	0.8	ICAN	0.71	0.71
Hybrid	0.8	0.8	Hybrid	0.71	1

Table 6. Results of the evaluation of scientific articles by broad scientific fields

Method	Accuracy for each broad scientific field				
	1. Natural sciences	2. Engineering sciences	3. Medical sciences	4. Agricultural sciences	5. Social sciences and humanities
TF-IDF	0.85	0.87	0.85	0.86	0.83
ICAN	0.79	0.72	0.74	0.9	0.91
Hybrid	0.9	0.88	0.93	0.92	0.94

Thus, validation on the corpus of applicants' abstracts demonstrates the stable and interpretable quality of specialty code recovery using the proposed semantic-ontological model in a hybrid configuration.

In most cases, the correct passport ends up in a narrow set of the most likely candidates, making the model suitable both for automated support in selecting a specialty code and for analyzing alternative thematically related areas.

5.2. Results of model validation in scientific articles published in the VAK list

The model was validated using scientific articles from a corpus of publications from journals included in the current VAK list. Since these journals are known to represent broad scientific fields and groups of scientific specialties, it was possible to conduct an assessment both at the aggregate level of fields and at the level of passport groups.

The first experiment analyzed the accuracy of assigning an article to a broad field of science. After determining the reference field based on the journal's profile in the VAK list, the predicted field was calculated based on the passport with the maximum value of S(d, s). The results show differences in the behavior of lexical, semantic, and hybrid configurations (Table 6).

The TF-IDF method demonstrates the greatest stability in natural and engineering sciences, where the top 1 accuracy is 0.85 and 0.87, respectively. In medical and agricultural sciences, the accuracy of TF-IDF is 0.85 and 0.86, respectively, while in social sciences and humanities, it is 0.83. This is consistent with the fact that in natural sciences and technical fields, terminology is more standardized and closer to the wording of passports.

The ICAN semantic model performs significantly better in cases where the language of publications is more variable. In the social sciences and humanities, its accuracy reaches 0.91, compared to 0.83 for TF-IDF. In agricultural sciences, ICAN scores 0.90 compared to 0.86 for TF-IDF. The slight inferiority of ICAN to the lexical model in natural, engineering, and medical sciences reflects the dependence of the purely semantic layer on the quality of local graphs in terminology-rich but well-standardized fields.

The hybrid configuration combines the strengths of both approaches. Across all five broad areas, it yields the highest top 1 accuracy values. For natural sciences, the accuracy is 0.90; for engineering, 0.88; for medicine, 0.93; for agriculture, 0.92; and for social sciences and humanities, 0.94. Thus, when moving to a higher level of aggregation, the hybrid model almost always corrects the errors of each of the single configurations to ensure the most stable behavior.

The second experiment evaluated the accuracy of restoring groups of scientific specialties. For each article, a vector S(d, s) was formed for all passports, after which the predicted group was determined by the passport with the maximum value of S(d, s). The reference group was set according to the declared specialization of the journal. Here, the top 1 and top 3 metrics were analyzed (Table 7).

Table 7. Results of the evaluation of scientific articles by groups of scientific specialties

Method	Accuracy according to top 1	Accuracy according to top 3
TF-IDF	0.9	0.94
ICAN	0.87	0.97
Hybrid	0.91	0.96

At the specialty group level, the TF-IDF lexical model achieves a top 1 accuracy of 0.90 and a top 3 accuracy of 0.94. The ICAN semantic model gives a slightly lower top 1 accuracy of 0.87, but a higher top 3 accuracy of 0.97. This means that ICAN is slightly more likely to make a mistake in selecting the single closest passport, but almost always includes the correct group among the three most likely options.

The hybrid model retains the best top 1 value of 0.91, while providing high top 3 accuracy of 0.96. While not outperforming ICAN in terms of top 3, it provides a more balanced ratio between the accuracy of the first choice and the completeness of the top three groups. In the context of recommendation ranking tasks based on codes, this configuration is the most practical, since the main code can be reliably suggested to simultaneously form a content-relevant list of alternative specialties.

Overall, the results for articles from journals included in the VAK list confirm the conclusions made based on the abstracts. The lexical model works best in areas with a rigidly defined nomenclature, while the semantic model is particularly useful in the humanities and related fields, and the hybrid configuration provides the most stable quality at all levels of aggregation to provide interpretable probability profiles by scientific specialty group.

5.3. Comparison of a graph-based semantic-ontological model with large language models

To evaluate the proposed model, we compared it with a number of large language models applied in classification mode without retraining. In all cases, the same task was established: to restore the specialty code or the closest passport based on the text of the document. Top 1 accuracy and top 3 accuracy were calculated on corpora of abstracts and scientific articles (Table 8).

According to the abstracts, the hybrid graph semantic-ontological model gives a top 1 accuracy of 0.69 and a top 3 accuracy of 0.90. The lexical TF-IDF model and semantic ICAN are inferior to it in both metrics (0.58 for TF-IDF and 0.74 for ICAN, 0.57 and 0.81, respectively). Among large language models, the ChatGPT 5.2 Thinking[24] configuration achieves the best values, with a top 1 accuracy of 0.79 and a top 3 accuracy of 0.84. The ChatGPT 4o[25] model performs at 0.71 and 0.73, DeepSeek[26] at 0.61 and 0.70, LLaMA[27] at 0.57 and 0.63, while Alice AI (YandexGPT)[28] lags significantly behind all options, scoring 0.46 and 0.52. Thus, while the large language model ChatGPT 5.2 Thinking outperforms the hybrid scheme in terms of first choice accuracy according to the abstracts, the graph semantic-ontological model provides higher completeness for the three closest specialties to form a fuzzy but substantively stable profile of similarities based on passports.

Table 8. Results of comparing a graph-based semantic ontology model with large language models

Authors' abstracts			Articles		
Method	Accuracy according to top 1	Accuracy according to top 3	Method	Accuracy according to top 1	Accuracy according to top 3
TF-IDF	0.58	0.74	TF-IDF	0.85	0.86
ICAN	0.57	0.81	ICAN	0.81	0.84
Hybrid	0.69	0.9	Hybrid	0.91	0.93
ChatGPT 4o	0.71	0.73	ChatGPT 4o	0.8	0.82
ChatGPT 5.2 Thinking	0.79	0.84	ChatGPT 5.2 Thinking	0.82	0.97
Alice AI (YandexGPT)	0.46	0.52	Alice AI (YandexGPT)	0.56	0.69
LLaMA	0.57	0.63	LLaMA	0.62	0.67
DeepSeek	0.61	0.7	DeepSeek	0.79	0.84

[24] https://openai.com/ru-RU/index/introducing-gpt-5-2/. Accessed April 04, 2026. (In Russ.).
[25] https://openai.com/ru-RU/index/hello-gpt-4o/. Accessed April 04, 2026. (In Russ.).
[26] https://www.deepseek.com/. Accessed April 04, 2026.
[27] https://www.llama.com/. Accessed April 04, 2026.
[28] https://alice.yandex.ru/. Accessed April 04, 2026. (In Russ.).

According to scientific articles from journals listed in the VAK catalog, the picture is different. The hybrid model achieves a top 1 accuracy of 0.91 and a top 3 accuracy of 0.93. The lexical TF-IDF model gives an accuracy of 0.85 and 0.86, while the corresponding ICAN scores are 0.81 and 0.84, respectively. Among large language models, ChatGPT 4o shows an accuracy of 0.80 and 0.82, while ChatGPT 5.2 Thinking gives 0.82 and 0.97, DeepSeek shows 0.79 and 0.84, LLaMA shows 0.62 and 0.67, and Alice AI shows 0.56 and 0.69. According to the top 1 metric, the hybrid model outperforms all language models by nine to fifteen percentage points. According to the top 3 metric, ChatGPT 5.2 Thinking shows the best result (0.97), while the hybrid model gives a slightly lower value of 0.93, but remaining significantly higher than the other configurations. This means that when classifying articles, the proposed graph semantic-ontological scheme better captures the main passport, while large language models more often include the correct specialty code in a wide set of closest candidates.

The differences in behavior are consistent with the nature of the approaches being compared. The graph-based semantic-ontological model, which is tightly bound to the texts of specialty passports, uses them as an ontological basis to factor solutions into interpretable TF-IDF and ICAN components.

This effect is particularly noticeable in articles from journals included in the VAK list, where the wording is closer to the normative language of passports. Large language models, which rely on generalized representations of scientific disciplines, often take into account context that is not reflected in passport texts, Therefore, in abstracts that contain detailed reviews, references to related fields, and less formalized presentation, the best ChatGPT configurations demonstrate higher accuracy in the first choice, but lose some of their interpretability and controllability.

The comparison shows that the proposed graph semantic-ontological model is comparable to modern language models in terms of classification quality, surpassing them in a number of metrics in scenarios where strict consistency with passport texts is important, while remaining significantly cheaper in terms of computational costs, as well as more transparent in terms of the structure of the decisions made.

Large language models should be viewed as an additional tool that complements rather than replacing ontology-oriented schemas. This is especially relevant in expert support tasks that require both quantitative scoring and explicit linking of results to normative descriptions of scientific disciplines.

6. DISCUSSION OF RESULTS AND CONCLUSION

The graph-based semantic-ontological model of scientific text analysis presented in this work is based on the VAK specialty passports, which are used both as textual descriptions of subject areas and as a source of ontological feature space. After forming a unified dictionary of lemmas and stable phrases based on the texts of the passports, the passports themselves, as well as the dissertation abstracts and scientific articles, can be described in this database.

The similarity of a document to the code of a scientific specialty is interpreted through the consistency of the text content with the normative description of the field.

At the same time, the semantic-ontological space has two coordinated layers. The lexical layer is built on the basis of TF-IDF vectors, which record the use of terms and terminological combinations selected from the passport chunk corpus. The semantic layer, which is based on the ICAN model, describes the text through a graph of co-occurring lemmas and the degree of its vertices. This supports paraphrasing, synonymy, the use of abbreviations, and terms scattered throughout the text, while maintaining a link to the same ontological basis. Thus, the hybrid metric based on lexical and semantic approaches combines local and global metrics in each layer.

Experimental results on abstracts show that the hybrid model systematically outperforms individual lexical and semantic approaches in terms of accuracy. The TF-IDF-based lexical model, which provides high accuracy where terminology is standardized, closely matches the wording of abstracts, while the semantic model is particularly useful in fields involving a freer scientific style, including active use of abbreviations and variable descriptions of the research subject.

Aggregated indicators for broad areas of knowledge confirm the observed pattern. For social sciences and humanities, the semantic layer plays a critical role in improving ranking quality, since thematic boundaries are often expressed through complex formulations and contextual markers. For a significant portion of technical and natural science specialties, the lexical layer already provides a high baseline level of quality, while the semantic layer refines the document profile in cases of similar codes and borderline topics. Thus, the model demonstrates behavior that can be explained in terms of content and is consistent with the peculiarities of the terminological structure of different scientific fields.

Further development of the model can take several directions. One of these involves the formation of lexemes and terminological combinations of passports with external ontologies, knowledge bases, and specialized terminological resources. This will permit a movement

from a textual ontological basis to a more formalized multi-domain ontology. In addition, it is possible to add contextual embeddings and neural network models as an additional layer on top of the existing TF-IDF and ICAN schemas while maintaining interpretability through decomposition on an ontological basis.

A separate task is the optimization of parameters on validation samples and analysis of their dependence on the field of knowledge and document type.

In conclusion, it can be confirmed that the graph-based semantic-ontological model based on the texts of the VAK specialty passports demonstrates the possibility of transforming the normative corpus into a working ontological space for quantitative analysis of scientific texts. The results obtained for abstracts and scientific articles show that this approach can serve as a basis both in the provision of automated support for expert decision-making, as well as for monitoring the structure of scientific knowledge within a given normative ontology.

Authors' contributions

N.S. Kurdyukov—conceptualization; methodology; formal analysis; writing the original draft. He has developed the core concept of information morphism within semantic-ontological and graph-based frameworks, designed the methodological approach, implemented the primary algorithms, conducted formal experiments and analysis, and prepared the initial manuscript draft.

V.N. Kalinin—methodology; validation; data curation; software; writing the original draft. He has contributed to the refinement of the methodological framework, developed computational tools, performed validation of the developed models and tools on experimental datasets, curated and structured the textual corpora used in the study, and prepared initial manuscript draft.

S.A. Kudzh—methodology supervision; validation; writing the review and editing. He has provided expert supervision of the methodological framework, ensured the scientific validity and rigor of the proposed models and analytical procedures, and contributed to critical review and refinement of the manuscript.

D.O. Zhukov—supervision; methodology oversight; validation; writing the review and editing. He has led the overall scientific supervision of the study, oversaw the development and consistency of the methodological approach, validated the research outcomes at a conceptual and theoretical level, and critically reviewed and approved the final manuscript.

REFERENCES

1. Altınel B., Ganiz M.C. Semantic text classification: A survey of past and recent advances. *Inf. Process. Management*. 2018;54(6):1129–1153. https://doi.org/10.1016/j.ipm.2018.08.001
2. Sikelis K., Tsekouras G.E., Kotis K.I. Ontology-based Feature Selection: A Survey. *arXiv* preprint arXiv:2104.07720 [cs.AI], 2021. https://doi.org/10.48550/arXiv.2104.07720
3. Ehring D., Ferraz-Doughty P., Luttmer J., Nagarajah A. A first step towards automatic identification and provision of user-specific knowledge: A verification of the feasibility of automatic text classification using the example of standards. *Procedia CIRP*. 2023;119:1103–1108. https://doi.org/10.1016/j.procir.2023.02.183
4. Layer M., Luttmer J., Nagarajah A., Stelzer R. Structured representation of pre-defined information backflow in standards and directives. *Standards*. 2024;4:262–285. https://doi.org/10.3390/standards4040013
5. Stănescu G., Oprea S.-V. Recent trends and insights in semantic web and ontology-driven knowledge representation across disciplines using topic modeling. *Electronics*. 2025;14(7):1313. https://doi.org/10.3390/electronics14071313
6. Touza I., Balama G., Lazarre W., Guidedi K., Kolyang. Ontology-driven text classification and data mining: Beyond keywords toward semantic intelligence. *Revue d'Intelligence Artificielle*. 2025;39(3):25–35. https://doi.org/10.18280/ria.390301
7. Pertsas V., Constantopoulos P. Ontology-driven extraction of contextualized information from research publications. In: *Proceedings of the 15th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management* (*IC3K 2023*). V. 2. KEOD. 2023. P. 108–118. https://doi.org/10.5220/0012254100003598
8. Mohd M., Javeed S., Nowsheena, Wani M.A., Khanday H.A. Sentiment analysis using lexico-semantic features. *J. Inform. Sci*. 2024;50(6):1449–1470. https://doi.org/10.1177/01655515221124016
9. Demidova L., Zhukov D., Andrianova E., Kalinin V. Model of lexico-semantic bonds between texts for creating their similarity metrics and developing statistical clustering algorithm. *Algorithms*. 2023;16:198. https://doi.org/10.3390/a16040198
10. Saeeda L., Med M., Ledvinka M., Blaško M., Křemen P. Entity linking and lexico-semantic patterns for ontology learning. In: Harth A., et al. *The Semantic Web*. Series: Lecture Notes in Computer Science. 2020. V. 12123. P. 138–153. https://doi.org/10.1007/978-3-030-49461-2_9
11. Yelmen I., Gunes A., Zontul M. Multi-class document classification using lexical ontology-based deep learning. *Appl. Sci*. 2023;13(10):6139. https://doi.org/10.3390/app13106139
12. Bugueño M., de Melo G. Connecting the dots: What graph-based text representations work best for text classification using graph neural networks? In: *Findings of the Association for Computational Linguistics*: *EMNLP 2023*. 2023. P. 8943–8960. https://doi.org/10.18653/v1/2023.findings-emnlp.600

13. Varella Ehrenfried H., Venturi Date V.T., Todt E. Exploring graph representation strategies for text classification. *Connect. Sci*. 2023;35(1):2289832. https://doi.org/10.1080/09540091.2023.2289832
14. Sánchez-Antonio C., Valdez-Rodríguez J.E., Calvo H. TTG-Text: A graph-based text representation framework enhanced by typical testors for improved classification. *Mathematics*. 2024;12:3576. https://doi.org/10.3390/math12223576
15. Onan A. Hierarchical graph-based text classification framework with contextual node embedding and BERT-based dynamic fusion. *Journal of King Saud University – Computer and Information Sciences.* 2023;35(7):101610. https://doi.org/10.1016/j.jksuci.2023.101610
16. Tsvetkov V.Ya., Kurdyukov N.S. Informational ontological modeling. *Russian Technological Journal.* 2025;13(2):18–26. https://doi.org/10.32362/2500-316X-2025-13-2-18-26
17. Nabhan A.R., Shaalan K. A graph-based approach to text genre analysis. *Computación y Sistemas*. 2016;20(3):527–539. https://doi.org/10.13053/CyS-20-3-2471
18. Ali I., Melton A. Semantic-based text document clustering using cognitive semantic learning and graph theory. In: *Proceedings of the 12th IEEE International Conference on Semantic Computing* (*ICSC 2018*). 2018. P. 243–247. https://doi.org/10.1109/ICSC.2018.00042
19. Lemaire B., Denhière G. Incremental construction of an associative network from a corpus. In: *Proceedings of the 26th Annual Meeting of the Cognitive Science Society*. 2004. V. 26. P. 825–830.

About the Authors

Nikita S. Kurdyukov, Postgraduate Student, Department of Instrumental and Applied Software, Institute of Information Technologies, MIREA – Russian Technological University (78, Vernadskogo pr., Moscow, 119454 Russia). E-mail: nskurdyukov@gmail.com. RSCI SPIN-code 8535-1612, https://orcid.org/0000-0001-6784-3369

Vladimir N. Kalinin, Assistant, Department of Telecommunications, Institute of Radio Electronics and Informatics, MIREA – Russian Technological University (78, Vernadskogo pr., Moscow, 119454 Russia). E-mail: kalinin_v@mirea.ru. Scopus Author ID: 57562579000, https://orcid.org/0000-0003-1365-4639

Stanislav A. Kudzh, Dr. Sci. (Eng.), Professor, Department of Instrumental and Applied Software, Institute of Information Technologies, MIREA – Russian Technological University (78, Vernadskogo pr., Moscow, 119454 Russia). E-mail: kudzh@mirea.ru. Scopus Author ID 56521711400, ResearcherID AAG-1319-2019, RSCI SPIN-code 8173-1572, https://orcid.org/0000-0003-1407-2788

Dmitry O. Zhukov, Dr. Sci. (Eng.), Professor, Department of Telecommunications, Institute of Radio Electronics and Informatics, MIREA – Russian Technological University (78, Vernadskogo pr., Moscow, 119454 Russia). E-mail: zhukov_do@mirea.ru. Scopus Author ID 57189660218, RSCI SPIN-code 1798-8891, https://orcid.org/0000-0002-1211-5214

Об авторах

Курдюков Никита Сергеевич, аспирант, кафедра инструментального и прикладного программного обеспечения, Институт информационных технологий, ФГБОУ ВО «МИРЭА – Российский технологический университет» (119454, Россия, Москва, пр-т Вернадского, д. 78). E-mail: nskurdyukov@gmail.com. SPIN-код РИНЦ 8535-1612, https://orcid.org/0000-0001-6784-3369

Калинин Владимир Николаевич, ассистент, кафедра телекоммуникаций, Институт радиоэлектроники и информатики, ФГБОУ ВО «МИРЭА – Российский технологический университет» (119454, Россия, Москва, пр-т Вернадского, д. 78). E-mail: kalinin_v@mirea.ru. Scopus Author ID 57562579000, https://orcid.org/0000-0003-1365-4639

Кудж Станислав Алексеевич, д.т.н., профессор, профессор кафедры инструментального и прикладного программного обеспечения, Институт информационных технологий, ФГБОУ ВО «МИРЭА – Российский технологический университет» (119454, Россия, Москва, пр-т Вернадского, д. 78). E-mail: kudzh@mirea.ru. Scopus Author ID 56521711400, ResearcherID AAG-1319-2019, SPIN-код РИНЦ 8173-1572, https://orcid.org/0000-0003-1407-2788

Жуков Дмитрий Олегович, д.т.н., профессор, профессор кафедры телекоммуникаций, Институт радиоэлектроники и информатики, ФГБОУ ВО «МИРЭА – Российский технологический университет» (119454, Россия, Москва, пр-т Вернадского, д. 78). E-mail: zhukov_do@mirea.ru. Scopus Author ID 57189660218, SPIN-код РИНЦ 1798-8891, https://orcid.org/0000-0002-1211-5214

Translated from Russian into English by L. Bychkova
Edited for English language and spelling by Thomas A. Beavitt

Multiple robots (robotic centers) and systems. Remote sensing and non-destructive testing

Роботизированные комплексы и системы.
Технологии дистанционного зондирования неразрушающего контроля

UDC 621.396.6
https://doi.org/10.32362/2500-316X-2026-14-3-43-59
EDN EKRNQE

RESEARCH ARTICLE

Localization of structural defects of printed circuit board assembly by vibration diagnostics

Saygid U. Uvaysov, Aleksey V. Dolmatov @, Vo The Hai, Nguyen Duc Hai, Pham Xuan Hanh, Ruslan M. Uvaysov

MIREA – Russian Technological University, Moscow, 119454 Russia
@ Corresponding author, e-mail: dolmatov@mirea.ru

• **Submitted:** 14.05.2025 • **Revised:** 30.12.2025 • **Accepted:** 24.03.2026

Abstract

Objectives. Structural defects may arise during the production and operation of printed circuit board assemblies (PCBAs) installed in radio-electronic assemblies (REAs) due to manufacturing imperfections or the influence of external factors. Of particular concern are latent defects that cannot be detected after the PCBA has been manufactured and which may lead to failures during operation. The typical installation in sealed enclosures of modern PCBAs, which are characterized by a high density of electronic components, significantly complicates the use of conventional inspection and diagnostic methods. This study aims to improve the reliability of detecting and classifying latent defects in PCBAs in sealed blocks based on their mechanical amplitude-frequency characteristics (AFCs) using a deep neural network.

Methods. The paper proposes a vibration diagnostics method based on the numerical modeling of mechanical processes, experimental vibration testing, and the application of an artificial neural network for technical condition classification. Diagnostics involves analyzing the obtained mechanical AFCs using an accelerometer mounted on the enclosure and comparing them with a database of calculated AFCs generated for various technical states of the PCBA.

Results. To verify the proposed method experimentally, a PCBA mock-up was fabricated and installed inside an enclosure with various structural defects introduced into its design. A database of calculated mechanical AFCs was formed for both healthy and faulty states. After obtaining experimental AFCs, they were compared with the calculated data, and the introduced defects were identified using a deep neural network.

Conclusions. The developed vibration diagnostics method enables the highly accurate detection and classification of latent defects arising in radio-electronic equipment during production and operation. This improves the reliability of technical condition assessment of PCBAs.

Keywords: diagnostics, vibrational loading, radio-electronic equipment, PCBA, DNN

For citation: Uvaysov S.U., Dolmatov A.V., Vo T.H., Nguyen D.H., Pham X.H., Uvaysov R.M. Localization of structural defects of printed circuit board assembly by vibration diagnostics. *Russian Technological Journal.* 2026;14(3):43–59. https://doi.org/10.32362/2500-316X-2026-14-3-43-59, https://www.elibrary.ru/EKRNQE

Financial disclosure: The authors have no financial or proprietary interest in any material or method mentioned.

The authors declare no conflicts of interest.

НАУЧНАЯ СТАТЬЯ

Локализация дефектов конструкций печатных узлов методом вибродиагностики

С.У. Увайсов, А.В. Долматов [@], Т.Х. Во, Д.Х. Нгуен, С.Х. Фам, Р.М. Увайсов

МИРЭА – Российский технологический университет», Москва, 119454 Россия
[@] Автор для переписки, e-mail: dolmatov@mirea.ru

• **Поступила:** 14.05.2025 • **Доработана:** 30.12.2025 • **Принята к опубликованию:** 24.03.2026

Резюме

Цели. В процессе производства и эксплуатации печатных узлов (ПУ), устанавливаемых в блоках радиоэлектронных средств (РЭС), могут возникать конструктивные дефекты, обусловленные технологическими погрешностями или воздействием внешних эксплуатационных факторов. Наибольшую опасность представляют латентные дефекты, которые невозможно выявить после изготовления ПУ, но которые способны привести к отказам в процессе эксплуатации. Современные ПУ характеризуются высокой плотностью компоновки электрорадиоэлементов и, как правило, размещаются в герметичных блоках, что существенно затрудняет применение традиционных методов контроля. Целью работы является повышение достоверности выявления и классификации латентных дефектов конструкций ПУ в герметичных блоках по механическим амплитудно-частотным характеристикам (АЧХ) с использованием глубокой нейронной сети.

Методы. Предложен метод вибродиагностики, основанный на численном моделировании механических процессов, экспериментальных вибрационных испытаниях и применении искусственной нейронной сети для классификации технического состояния. Диагностирование осуществляется путем анализа механических АЧХ, полученных с помощью акселерометра, установленного на корпусе блока, и их сравнения с базой расчетных АЧХ, сформированной для различных технических состояний ПУ.

Результаты. Для экспериментальной проверки метода изготовлен макет ПУ внутри блока, в конструкцию которого вносились различные дефекты, и сформирована база расчетных механических АЧХ для исправного и неисправных состояний. После получения экспериментальных АЧХ выполнено их сравнение с расчетными данными, и с использованием глубокой нейронной сети определены виды внесенных дефектов.

Выводы. Разработанный метод вибродиагностики ПУ позволяет с высокой точностью выявлять и классифицировать латентные дефекты, возникающие в РЭС в процессе производства и эксплуатации, и повышает достоверность оценки технического состояния ПУ.

Ключевые слова: диагностирование, вибрационные воздействия, радиоэлектронное средство, печатный узел, DNN

Для цитирования: Увайсов С.У., Долматов А.В., Во Т.Х., Нгуен Д.Х., Фам С.Х., Увайсов Р.М. Локализация дефектов конструкций печатных узлов методом вибродиагностики. *Russian Technological Journal*. 2026;14(3):43–59. https://doi.org/10.32362/2500-316X-2026-14-3-43-59, https://www.elibrary.ru/EKRNQE

Прозрачность финансовой деятельности: Авторы не имеют финансовой заинтересованности в представленных материалах или методах.

Авторы заявляют об отсутствии конфликта интересов.

INTRODUCTION

A typical diagnostic system comprises a unit under test (UUT), as well as the various tools and methods for determining its technical condition. The tools are defined as a set of hardware and software devices used to measure and analyze diagnostic data obtained from the UUT at all stages of its lifecycle. Depending on how the UUT interacts with the diagnostic tool, different functional and testing techniques may be employed. Functional techniques are used to diagnose the UUT during operation, employing sensors to gather information regarding its technical condition. In this case, the diagnostic tools do not interfere with the UUT. Testing techniques involve applying a signal or set of signals are applied to the UUT input to trigger processes that provide information about its technical state [1].

Diagnosis of the technical condition of radio-electronic assemblies (REAs) constructed as modular units with printed circuit board assemblies (PCBAs) involves analyzing their mechanical amplitude-frequency characteristics (AFCs) under vibration conditions. This technique has been widely used in technical diagnostics [2–4]. Design defects in a device can lead to changes in its resonant frequencies as compared to those of a well-functioning device [5]. Based on this, a methodology can be developed for diagnosing PCBA structures that detects and classifies defects in devices during manufacturing and operation without affecting their structure.

Within the scope of this study, the authors are investigating the development of a non-destructive diagnostic technique. This includes the creation of models and algorithms for automatically detecting and identifying types of latent defects in a structure. These defects can affect the mechanical AFC of a block design, but the research does not consider failures resulting from the electrical parameters of electronic components. To enhance the reliability of the technique, a mechanical simulation process is applied to the UUT, taking into account variations in structural parameters.

1. METHOD OF DIAGNOSING PCBA STRUCTURES IN CASE OF VIBRATION EFFECTS

Most design defects in PCBA that occur during manufacture and operation are caused by mechanical factors, such as shock, harmonic vibration, and acoustic noise. These include missing or deformed electronic components, cracks in the structure, loosening of fasteners, inhomogeneity of the PCBA material, and other issues. However, defects that occur during equipment operation cannot be directly detected when using the device for its intended purpose [6], as current methods require shutting down the equipment, or compromising the structural integrity of the unit containing the controller, which is undesirable for sealed units.

Such defects may occur during the manufacture or operation of the units. To detect and identify the type of these latent defects, we propose a diagnostic method that compares the experimentally obtained AFC on the external surface of the blocks at a designated control point with a set of calculated AFCs for different technical conditions of the PCBA structure at the same control point [7]. Based on this comparison, it is possible to determine the technical condition of the device and take an appropriate action. The use of this technique can reduce labor expenses, diagnostic time, and financial costs associated with the monitoring and diagnosis of PCBA structures [8, 9].

With the use of modern modeling techniques and AFC analysis, it is possible to accurately identify the relationship between a specific type of defect and a corresponding diagnostic feature. In order to collect diagnostic data and subsequently analyze it, the following series of steps should be taken:

- conversion of physical phenomena into diagnostic signals by measuring the AFC using an accelerometer;
- collection of characteristics from this signal and analysis thereof through modeling of various technical conditions;
- comparison of the analytical results with values defined in technical documentation;
- use of an artificial intelligence (AI) algorithm for decisions-making regarding the technical condition.

The structural diagram of the proposed method is presented in Fig. 1.

In order to diagnose the technical condition of the PCBA structure based on the design documentation (Block 1), it is necessary to develop a model that enables the analysis of mechanical processes (Block 2). At the same time, the design model for the block with PCBA must accurately describe its geometric structure along with the physical and mechanical properties of its materials, as well as the mechanical characteristics of each electronic component of the simulated PCBA. In other words, the model must accurately reflect the data specified in the product's design documentation.

A computer-aided design (CAD) system can be utilized to simulate mechanical processes within the unit to determine the AFC at specific points when the device is subjected to vibrational mechanical forces. The described technique uses the *Solidworks*[1] software tool.

[1] https://www.solidworks.com/. Accessed March 23, 2026.

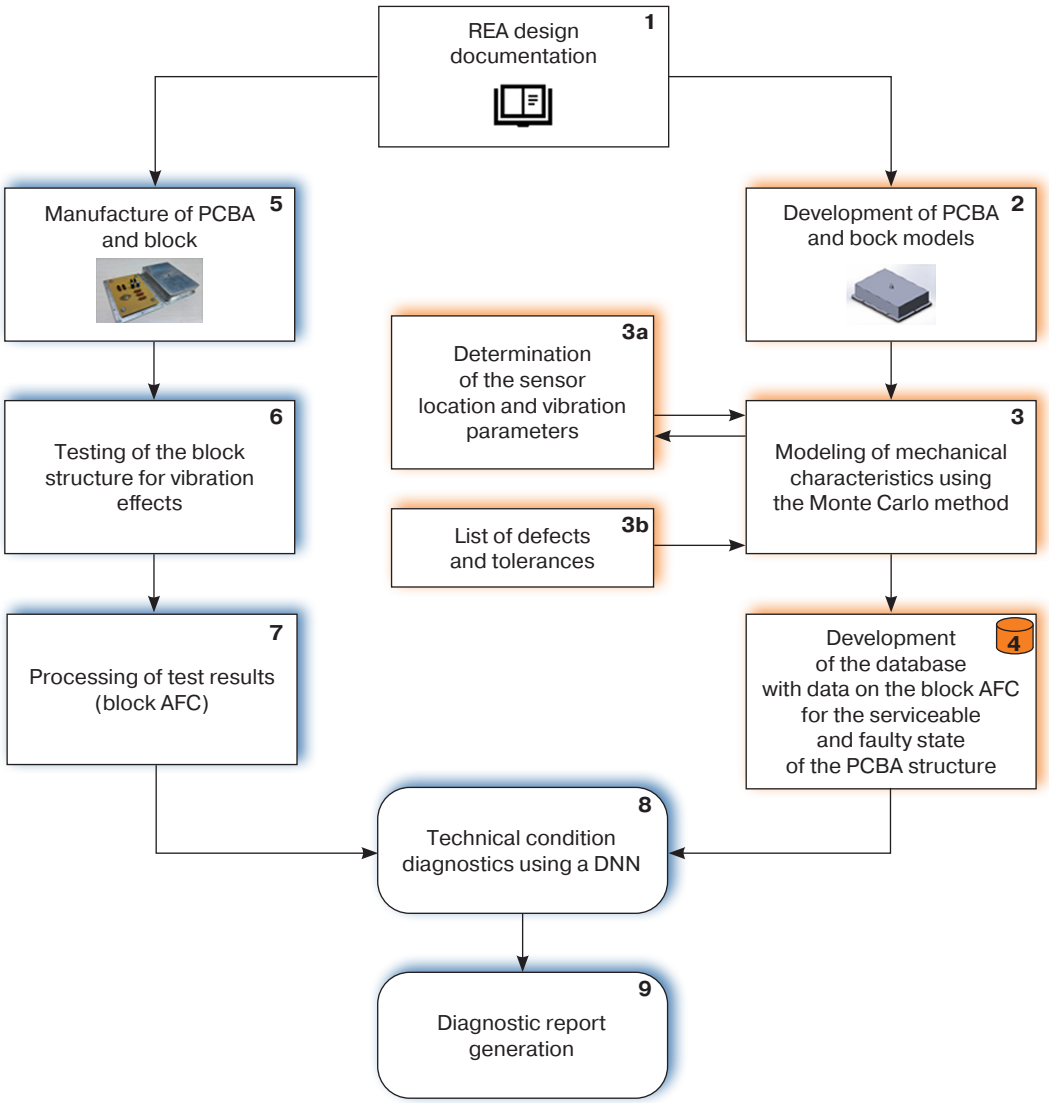


Fig. 1. Method for vibration-based diagnostics of PCBA

The AFC of the block, which is derived from the analysis of mechanical processes, depends on the selection of a reference point for calculating the AFC, as well as on the parameters of the applied mechanical vibrations, such as the amplitude and frequency of harmonic vibrations (Block 3a). The position of the reference point used to determine the AFC is chosen based on the peak amplitude of the vibrations; in this case, the center of the enclosure cover is the preferred option. During diagnostics, the amplitude of the harmonic vibrations should not exceed the theoretical maximum value of $A_{\text{max theor}}$ at which mechanical failure of the block and PCBA may occur, nor should it be significantly lower than the actual value of $A_{\text{min actual}}$ that can be detected by the sensor. Therefore, the minimum and maximum values of the harmonic oscillation amplitude at a monitoring point, denoted as $A_{\text{min effect}}$ and $A_{\text{max effect}}$, respectively, must meet the following condition[2]:

$$[A_{\text{min effect}}; A_{\text{max effect}}] \in [A_{\text{min actual}}; A_{\text{max theor}}].$$

In order to obtain information on the mechanical characteristics of the block structure under different PCBA technical conditions, it is necessary to simulate mechanical processes in both good PCBA conditions and in the presence of various defects. A list of potential defects is provided in Block 3b. The greater the number of defects considered in the simulation process, the more accurate the results of the technical condition diagnosis will be.

[2] Lu Ngoc Tien. *Acoustic Emission Method of Control of Multilayer Printed Circuit Boards of Radio-Electronic Devices*: Cand. Sci. Thesis (Eng.). Moscow: 2024.

The physical and mechanical properties (γ_i) of real PCBA structures often vary within a certain range, typically between 3% and 10%. To account for this variation, statistical Monte Carlo simulations are performed (Block 3) [10]. Based on this analysis, a set of AFCs is calculated for various combinations of the physical and mechanical PCBA design parameters that fall within acceptable ranges. In this scenario, it is assumed that the distribution of these parameter values follows a normal distribution:

$$f(\gamma_i) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(\gamma_i - \mu)^2}{2\sigma^2}\right),$$

where $f(\gamma_i)$ is the probability density function for the distribution of the γ_i physical and mechanical parameter; μ is the expected (or nominal) value of this parameter; σ is its standard deviation.

For the *k*th iteration of the Monte Carlo simulation, the *i*th physical and mechanical parameter is calculated using the following formula:

$$\gamma_i = \gamma_i^{\text{nom}}(1 + \xi_k \delta_i),$$

where γ_i is the value of the *i*th parameter in the *k*th iteration of the Monte Carlo simulation; γ_i^{nom} is the nominal value of the *i*th physical and mechanical parameter; δ_i is a relative tolerance for the *i*th physical and mechanical parameter; $\xi_k \sim N(-1, 1)$ is a random variable following a standard normal distribution [11, 12].

A database of calculated AFCs (Block 4) for various PCBA design conditions was developed using *Microsoft Excel*. The database is used to make a final conclusion about the PCBA's technical condition by comparing the AFCs obtained from field testing.

To determine the technical condition of the PCBA during the production or operation of devices manufactured in accordance with the design documentation (Block 5), a mechanical AFC test is carried out on the device using a vibration testing apparatus (Block 6). In order to obtain AFC data during these tests, a sensor is installed at the same reference point at which the calculated AFC was obtained during simulation. In this study, the VP-32 model manufactured by IMV Corporation (Japan) is used as the vibration sensor. This highly sensitive piezoelectric accelerometer, which converts vibration acceleration into an electrical charge, has a wide operating frequency range (up to 10 kHz) and is highly resistant to shock loads, making it suitable for use in vibration analysis under conditions of intense mechanical stress. The main element of the sensor is a quartz crystal that converts vibration into a proportional electrical voltage. The sensitivity of the sensor is expressed in pC/g, where pC is the unit of charge known as the picocoulomb, and g is the acceleration due to gravity, expressed in m/s^2. The sensor's measurement range falls within the linear portion of its AFC curve [13].

The experimentally obtained mechanical AFC of the block design (Block 7) is converted into a data table containing amplitude and frequency values. This data is then compared with the stored mechanical AFC in the database (Block 4) using an artificial neural network (Block 8) to detect and classify any defects. In this process, a deep neural network (DNN) is employed. Based on the comparison between the experimental and simulated mechanical AFCs, a report is generated regarding the PCBA technical condition (Block 9).

2. OBTAINING CALCULATED MECHANICAL AFC

When designing a block with PCBA, it is essential to simulate mechanical processes under the influence of vibration-induced mechanical stresses. This simulation, which is a critical component of the proposed approach, significantly impacts the accuracy of the diagnostic outcomes. The simulation is conducted for different technical conditions of the device. The more diverse the technical states considered for the structure, the more accurate the final conclusions based on the diagnostic results will be.

When diagnosing the PCBA structure, a list of potential latent defects is compiled and the mechanical characteristics of the device are simulated to obtain the calculated AFC at a control point of the unit. This corresponds to each identified defect. The experimental AFC is then measured on a physical prototype to identify potential defects by comparing the experimental and calculated AFCs for each possible technical condition. This comparison establishes whether the PCBA is functioning properly or has a defect.[3]

An algorithm has been developed to identify a list of potential defects in a PCBA. Its scheme is presented in Fig. 2 below.

This algorithm can be used to determine the likelihood of detecting a particular defect in the PCBA structure. Based on the AFC analysis control panel, a decision is made as to whether to include the relevant defect in the list of detected defects (Block 3b in Fig. 1).

The *SolidWorks Simulation* package is used to calculate mechanical characteristics under vibrational influences for the analysis of mechanical processes. The resonant frequencies of the structural block are calculated for input effects in the form of harmonic vibrations ranging from 0 to 2000 Hz.

[3] Malpass L. *SolidWorks 2009 API – Advanced Product Development*. 2009, 246 p.

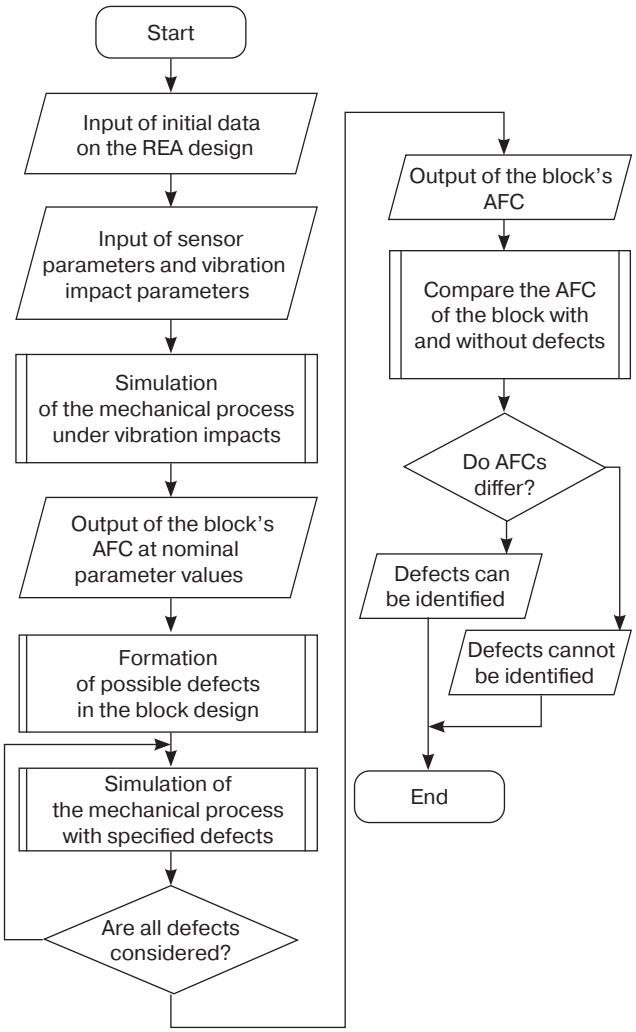


Fig. 2. Algorithm for identifying potential defects in a PCBA

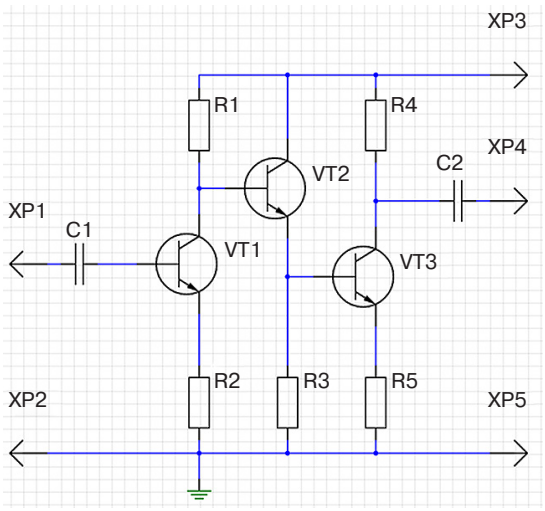
The UUT is a power amplifier, whose electrical circuit and PCBA design model are presented in Fig. 3. The prototype PCBA is as depicted in Fig. 4. This prototype device for simulating the distribution of electronic components on a printed circuit has been designed solely for this study, which includes mechanical and vibration testing of the device's design. However, this model does not take into account the electrical parameters of the electronic components, since this study focuses on the mechanical characteristics and dynamic response of the structure, rather than the electrical performance of the device.

The mechanical AFC of a tested prototype in good working condition (Fig. 5) exhibits resonant frequencies at f_1 = 878.1 Hz, f_2 = 1089.5 Hz, and f_3 = 1405.7 Hz.

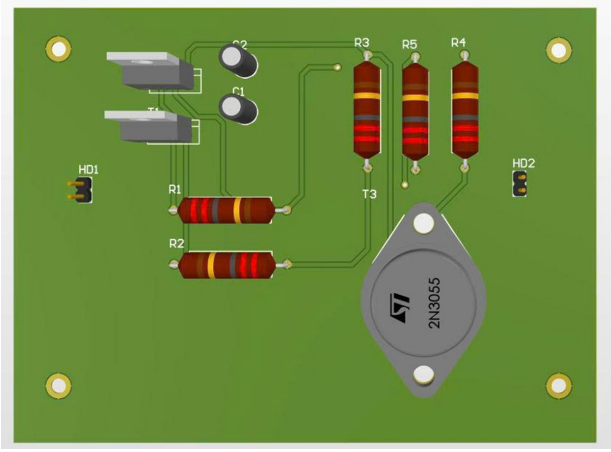
In cases where the tested prototype exhibits one or more of the identified defects, the resonant frequencies of the device vary as follows:

- Defect A (absent electronic component): f_1 = 878.3 Hz, f_2 = 1092.6 Hz, and f_3 = 1409.5 Hz (Fig. 6).
- Defect B (PCBA fastener loosening): f_1 = 878.2 Hz, f_2 = 1090.6 Hz, f_3 = 1428.8 Hz (Fig. 7).
- Defect C (electronic component deformation): f_1 = 878.1 Hz, f_2 = 1091.5 Hz, f_3 = 1403.6 Hz (Fig. 8).
- Defect D (layout and placement violations of an electronic component on the printed circuit): f_1 = 878.4 Hz, f_2 = 1090.2 Hz, f_3 = 1424.2 Hz (Fig. 9).
- Defect E (installation of an electronic component that does not correspond to the design documentation (Fig. 10).
- Defect F (heterogeneity of the printed circuit material structure): f_1 = 842.4 Hz, f_2 = 878.4 Hz, f_3 = 951.5 Hz (Fig. 11).

The overall AFC for all defects listed above is shown in Fig. 12.



(a)



(b)

Fig. 3. Electrical circuit (a) and design model of control unit (b) for the studied object. Designations of circuit elements correspond to those adopted in GOST 2.710-81[4]

[4] GOST 2.710-81. Interstate Standard. *Unified system for design documentation. Alpha-numerical designations in electrical diagrams*. Moscow: Izd. Standartov; 1985 (in Russ.).

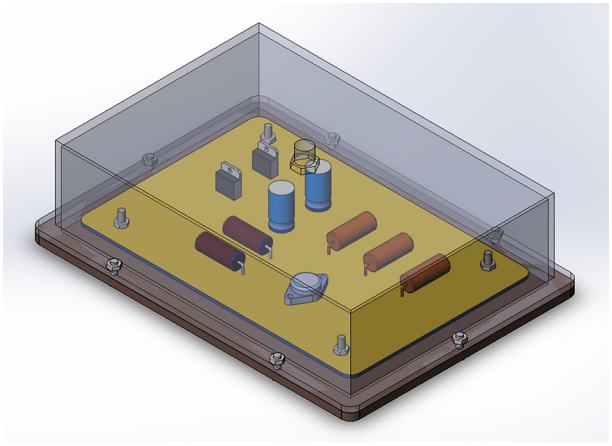


Fig. 4. Design model of a block with PCBA

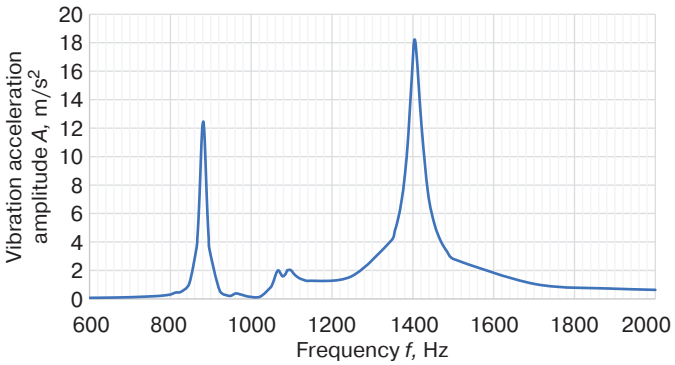
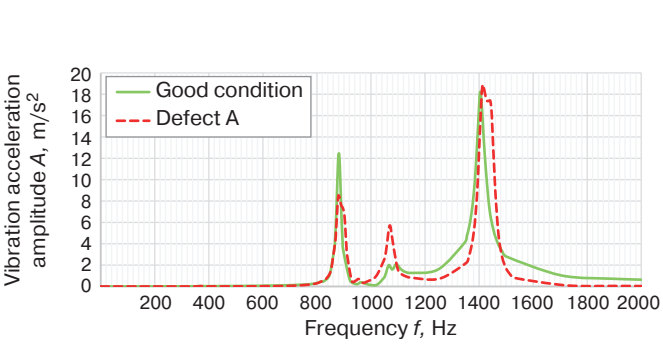


Fig. 5. AFC of the prototype in good operating condition



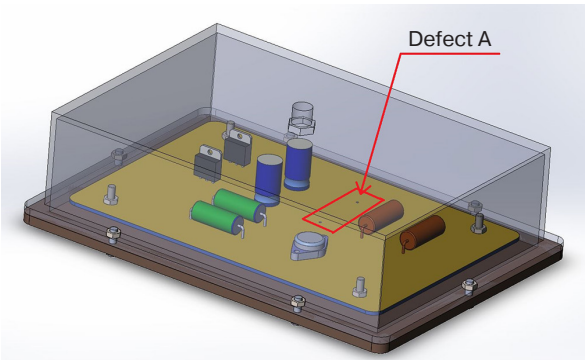
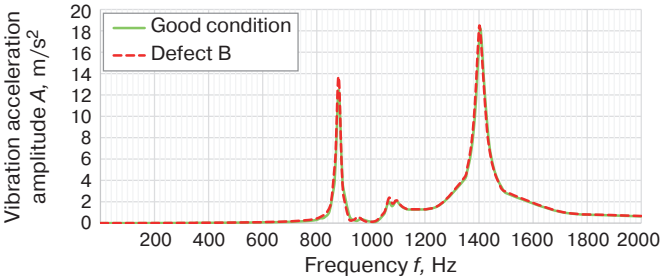


Fig. 6. AFC of the prototype with defect A



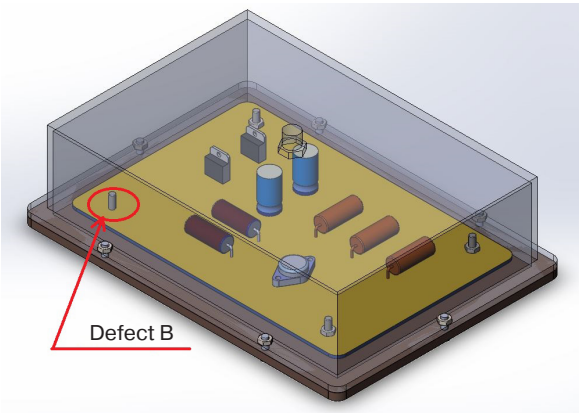
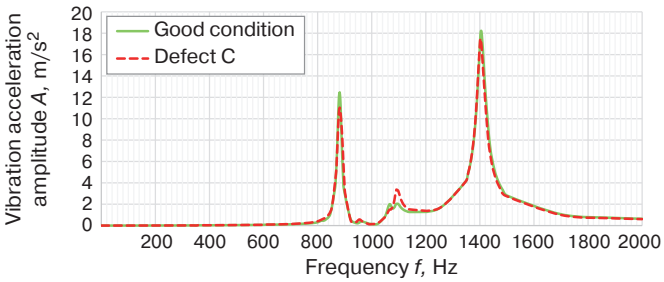


Fig. 7. AFC of the prototype with defect B



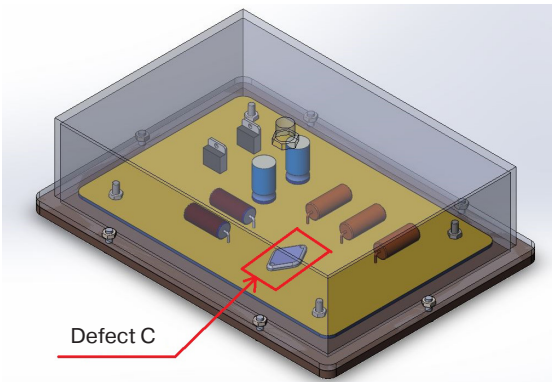


Fig. 8. AFC of the prototype with defect C

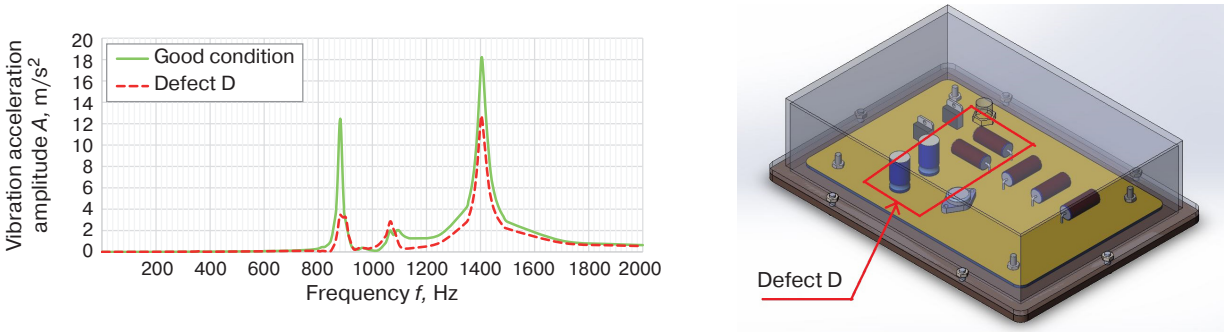


Fig. 9. AFC of the prototype with defect D

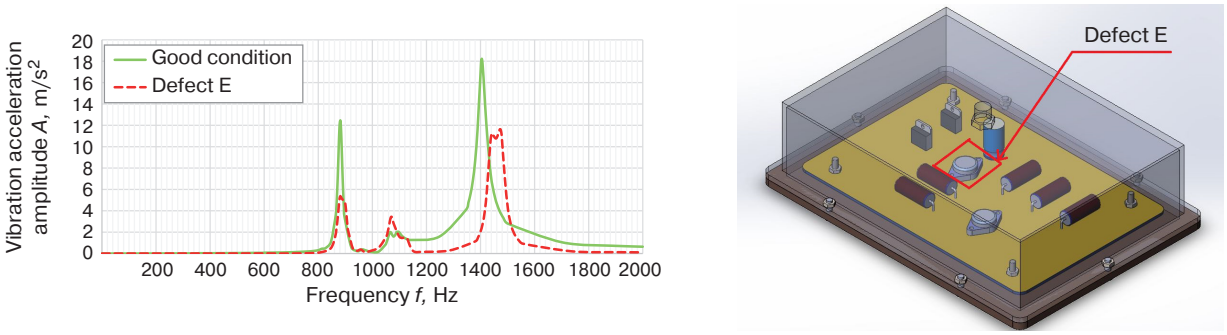


Fig. 10. AFC of the prototype with defect E

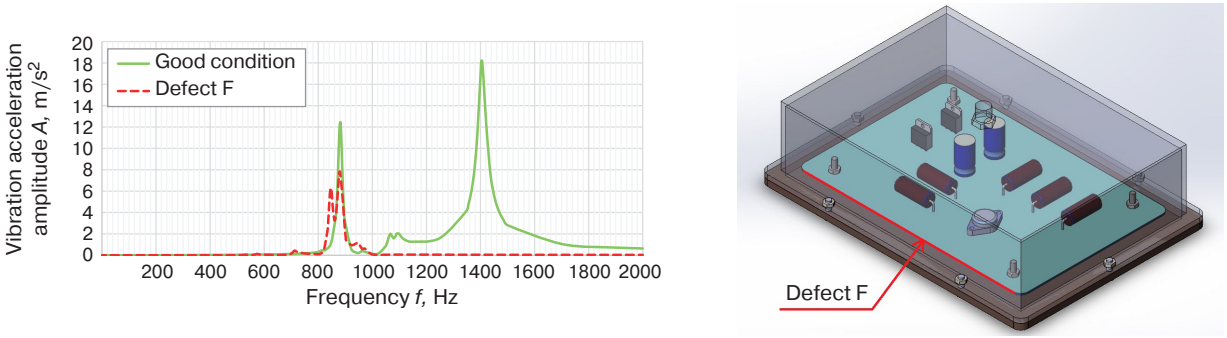


Fig. 11. AFC of the prototype with defect F

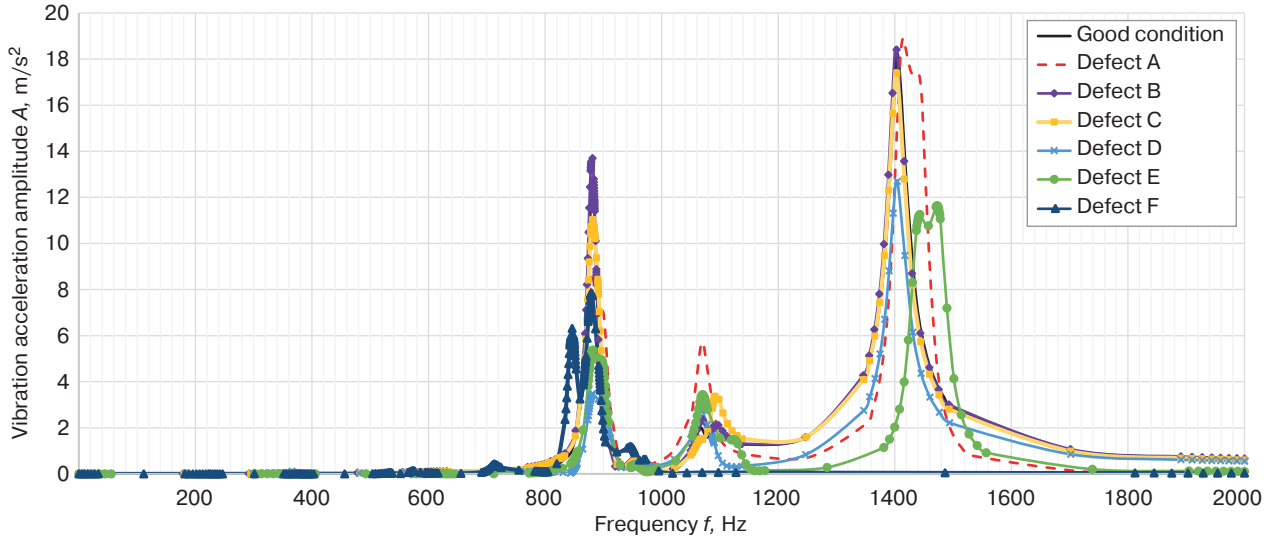


Fig. 12. The overall AFC of the tested prototype with various defects

Table. Frequency and amplitude values of vibration acceleration for various technical conditions of PCBA design during the modeling process

Good condition		Defect A		Defect B		Defect C		Defect D		Defect E		Defect F	
Frequency, Hz	Amplitude, m/s^2	Frequency, Hz	Amplitude, m/s^2	Frequency, Hz	Amplitude, m/s^2	Frequency, Hz	Amplitude, m/s^2	Frequency, Hz	Amplitude, m/s^2	Frequency, Hz	Amplitude, m/s^2	Frequency, Hz	Amplitude, m/s^2
$1.59\cdot10^{-6}$	$3.68\cdot10^{-19}$	$1.59\cdot10^{-6}$	$1.02\cdot10^{-20}$	$1.59\cdot10^{-6}$	$4.42\cdot10^{-19}$	$1.59\cdot10^{-6}$	$3.76\cdot10^{-19}$	$1.59\cdot10^{-6}$	$4.31\cdot10^{-19}$	$1.59\cdot10^{-6}$	$7.53\cdot10^{-20}$	$1.59\cdot10^{-6}$	$6.72\cdot10^{-20}$
9.5	$1.32\cdot10^{-5}$	8.1	$2.62\cdot10^{-7}$	7.8	$1.06\cdot10^{-5}$	7.9	$9.19\cdot10^{-6}$	8.0	$1.08\cdot10^{-5}$	8.0	$1.92\cdot10^{-6}$	4.9	$6.26\cdot10^{-7}$
19.1	$5.27\cdot10^{-5}$	16.5	$1.10\cdot10^{-6}$	15.9	$4.43\cdot10^{-5}$	16.1	$3.85\cdot10^{-5}$	16.3	$4.54\cdot10^{-5}$	16.5	$8.05\cdot10^{-6}$	9.9	$2.62\cdot10^{-6}$
28.6	0.0001	25.4	$2.60\cdot10^{-6}$	24.5	0.000105	24.8	$9.13\cdot10^{-5}$	25.1	$1.08\cdot10^{-4}$	25.3	$1.91\cdot10^{-6}$	15.3	$6.21\cdot10^{-6}$
……	……	……	……	……	……	……	……	……	……	……	……	……	……
1952.5	0.6712	1962.0	0.0426	1959.0	0.6948	1959.0	0.6303	1959.1	0.5700	1964.1	0.1149	1929.4	0.03589
1968.3	0.6591	1975.3	0.0414	1973.3	0.6839	1973.4	0.6202	1973.4	0.5618	1976.7	0.1127	1954.1	0.0356
1984.2	0.6476	1987.9	0.0403	1987.0	0.6741	1987.0	0.6111	1987.0	0.5550	1988.6	0.1107	1977.6	0.0354
2000	0.6368	2000	0.0393	2000	0.6652	2000	0.6028	2000	0.5478	2000	0.1090	2000	0.0352

The presence of any defect in the PCBA design can cause changes in its AFC. Therefore, defects are detected by comparing the AFC of a given prototype with those of previous responses corresponding to various possible defects in the PCBA design. This is an effective method of monitoring and determining the technical condition of the PCBA [14, 15].

In order to compare AFCs, a table is created containing vibration acceleration amplitude values for different technical conditions of the block design at a specific point, measured at various vibration frequencies ranging from 0 Hz to 2000 Hz.

3. EXPERIMENTAL VERIFICATION OF THE PCBA DEFECT DETECTION USING A VIBRATION TESTING APPARATUS

In order to obtain the experimental AFC of the studied prototype in both good and defective PCBA conditions, a benchmark test is performed using a testing setup consisting of a vibration testing apparatus and a computer equipped with specialized software for signal analysis installed. The setup comprises:

- A vibration testing apparatus M030/MA1 manufactured by IMV Corporation (Japan), which is used to generate mechanical harmonic vibrations at a specific amplitude and frequency that are set for the experiment. In this case, the UUT is subjected to harmonic vibration with an acceleration amplitude of 10 m/s^2 at frequencies ranging from 0 to 2000 Hz [15].
- The accelerometer comprises a VP-32 sensor that measures vibration acceleration on the UUT enclosure cover (maximum recorded acceleration of 500 m/s^2 over the operating frequency range of 2 to 10000 Hz).
- A personal computer (PC) operates a vibration testing apparatus to generate harmonic vibrations, and then records and displays the AFC measured using an accelerometer.

The UUT design is a prototype block (Fig. 13) having an aluminum body measuring 3 mm in thickness and a base measuring 5 mm. A hole has been created in the enclosure cover for the to be installation of the accelerometer. Inside the enclosure, there is a printed circuit prototype containing electronic components, which is fastened with screws 10 m away from the base.

Fig. 13. The design of the prototype block

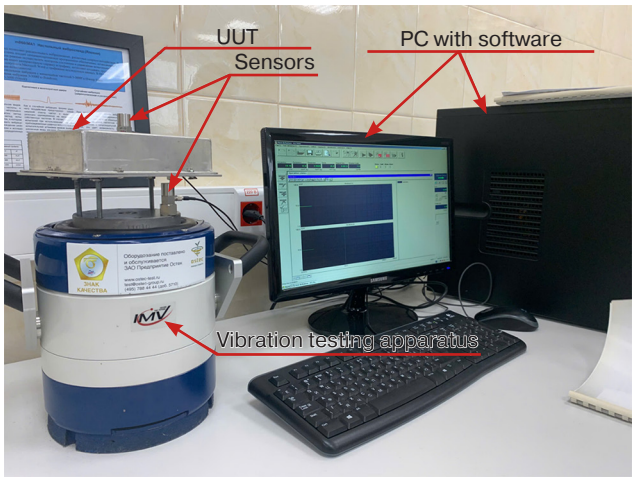


Fig. 14. The laboratory and experimental setup for vibration diagnostics

The laboratory and experimental setup for analyzing the mechanical AFC to vibration is shown in Fig. 14.

Figure 15 shows the mechanical AFC of the prototype obtained through field testing under vibration conditions.

Based on the experimental measurements, the prototype is found to have a satisfactory mechanical AFC. This response is measured in the presence of the aforementioned defects in both good and faulty conditions. In future, this AFC will be converted into a suitable format for comparison with the results of simulations. This will enable the technical condition of the prototype block to be assessed and the defects to be classified.

4. APPLICATION OF DEEP MACHINE LEARNING FOR THE CLASSIFICATION OF DEFECTS IN PCBA

Deep machine learning is a subfield of machine learning that uses artificial neural networks and algorithms to mimic the way the human brain works. This approach enables the independent learning by systems from large amounts of data to identify hidden patterns within them. Although the concept of deep learning dates back to the 1960s, it has only become widely used in practice with the development of computing technologies and big-data analysis methods, which have significantly boosted the efficiency of neural networks [16].

DNNs are used to recognize the complex, nonlinear relationships between the shape and parameters of the AFC, as well as the type of defect. This is not possible with simple perceptron models. The use of a DNN enables stable convergence and high classification accuracy.

A DNN is a type of artificial neural network consisting of multiple interconnected layers. The depth of the network is determined by the number of hidden layers and the complexity of its structure [17, 18].

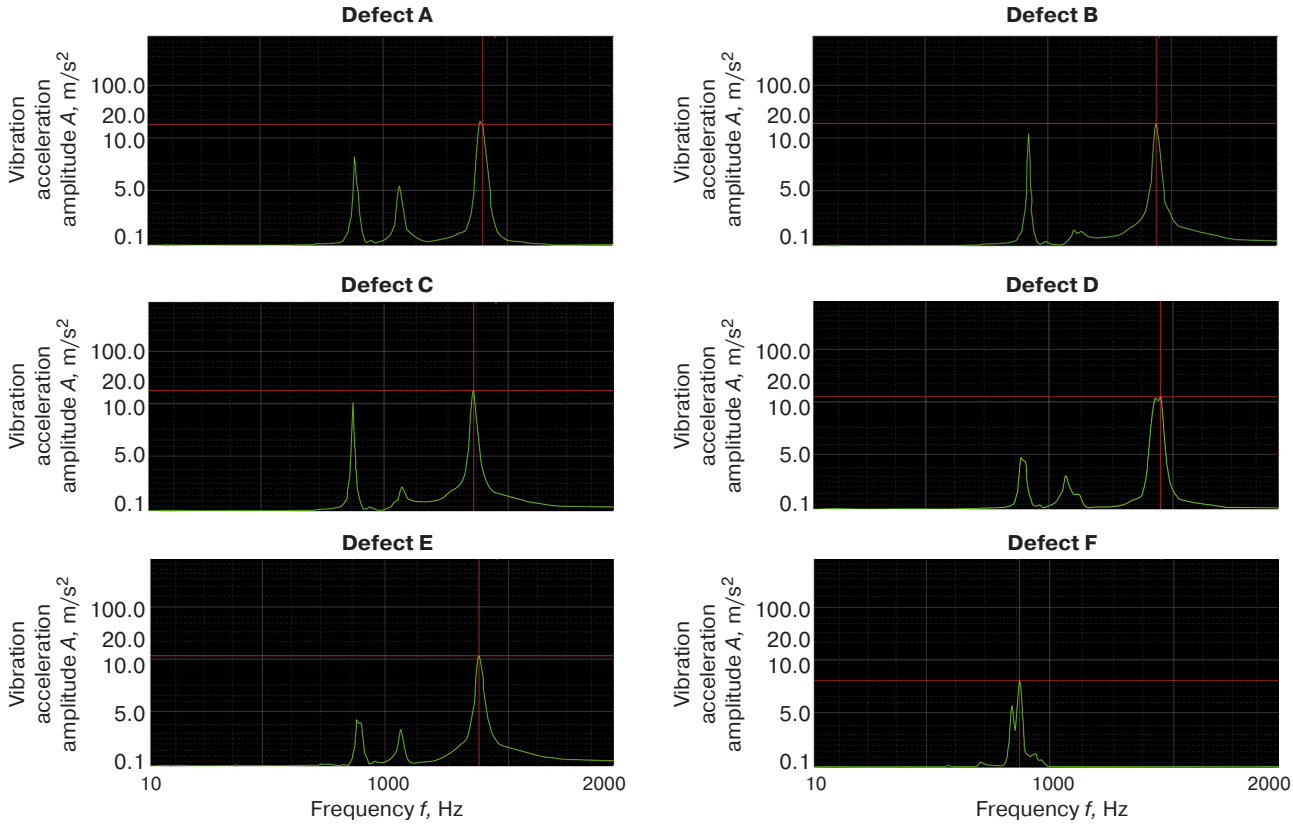


Fig. 15. Mechanical AFC of the prototype design obtained during field testing for vibration effects

Each layer processes data and passes the results to the next layer, enabling problems of high computational complexity to be solved.

The main components of the DNN are:

- the input layer, which receives the initial model data in the form of a two-dimensional numerical matrix containing vibration acceleration amplitude values and their corresponding frequencies for different technical conditions;
- the hidden layer, which processes the data and transfers information between layers;
- the output layer, which makes the final decision about the condition of the unit (good or defective).

The activation function provides a nonlinear approximation of the dependencies between the AFC characteristics and the types of defects in the PCBA design.

Weight coefficients and biases are adjustable network parameters that allow for its adjustment to the training data.

Figure 16 illustrates the structure of an artificial neuron used in a DNN. Here, x_i represents the neuron's input signals, w_{ki} represents the weight coefficients of the corresponding inputs; b_k represents the bias; Σ represents the summation operation; u_j is the linear combination of the neuron's inputs (also called the weighted sum or net input); j is the number of network input signals, depending on the frequency range and φ(.) represents the activation function.

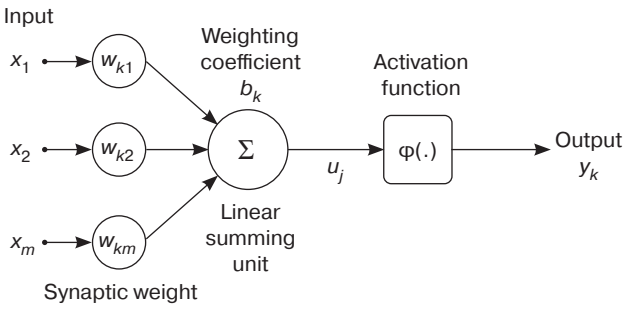


Fig. 16. Structure of an artificial neuron in a DNN

The study involves creating a database of design defects for the UUT, with a sample size of n = 1000. This corresponds to 1000 different implementations of each state, as obtained using the Monte Carlo simulation technique. The output layer consists of seven neurons, with each neuron corresponding to one functional state and six types of potential design defect in the studied PCBA.

The network architecture implemented during modeling is as follows (Fig. 17):

- input layer: combined frequencies and amplitudes;
- hidden layer: neurons with a rectified linear unit (ReLU) activation function:

$$f(x) = \max(0, x);$$

- output layer: nodes with a softmax activation function:

$$p_i = \frac{e^{z_j}}{\sum_{j=1}^{7} e^{z_j}}, \; i = \overline{1,7}; \qquad (1)$$

where z_j is a linear combination of the inputs at the output layer and p_i is the probability of belonging to class i.

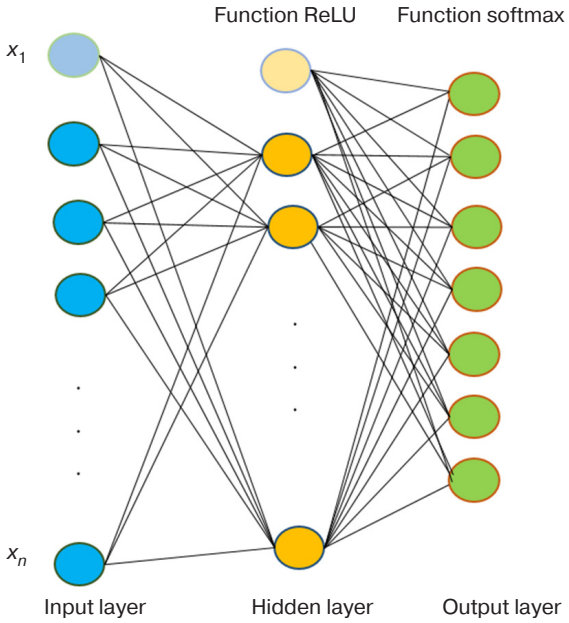


Fig. 17. Architecture of the developed DNN model

The activation function plays a key role in the operation of a neural network by ensuring its nonlinearity and modeling the process of signal transmission between neurons.

Due to its straightforward implementation and high training efficiency, ReLU is one of the most common activation functions. In addition, it has several advantages over sigmoid and tanh functions:

- provides faster convergence (training speed is six times higher than tanh);
- requires less computational effort;
- prevents the gradient saturation effect, thereby increasing training stability.

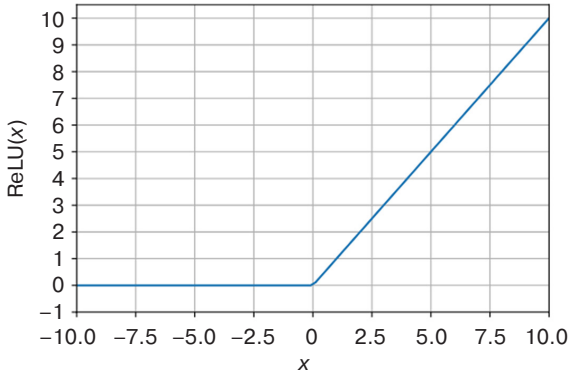


Fig. 18. Graph representing the ReLU function

By using the ReLU function in the defect classification task, the model's convergence rate can be significantly increased, as has been observed during training with *epoch* = 10. Here, *epoch* refers to the number of complete training cycles of the model on the entire dataset, with the weights of the neurons updated each time.

The technical condition is classified using a DNN implemented in Python with the Keras library. The model input comprises data arrays representing the AFC in the form of "frequency–amplitude" numerical pairs, which are formed based on the modeling and field test results. Training uses the cross-entropy loss and the Adam optimizer with 10 epochs.

The accuracy value of the model, as calculated when executing the DNN code in Python, is shown in Fig. 19.

Based on the training results, the model has an accuracy of 99.57%, indicating the high effectiveness of the proposed approach.

The classification and identification results obtained are presented in the form of histograms (Figs. 20–25), which reflect the percentage distribution of defect detection probability, calculated using formula (1).

```
loss, acc = model.evaluate(X_test, y_test)
print("Model accuracy: {:5.2f}%".format(100*acc))

22/22 ———————————————————————————— 1s 30ms/step - accuracy: 0.9957 - loss: 0.6735
Model accuracy: 99.57%
```

Fig. 19. Screenshot of the DNN model accuracy calculation results in the *Google Colab*[5] environment

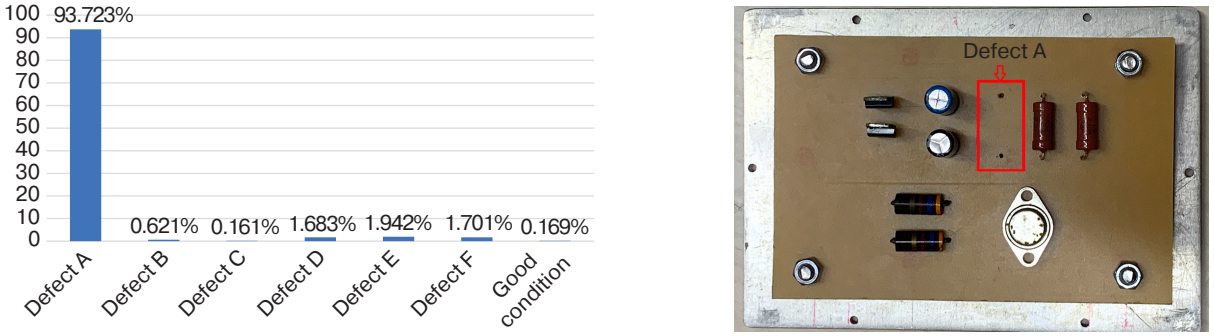


Fig. 20. Report on the technical condition of the tested model with defect A

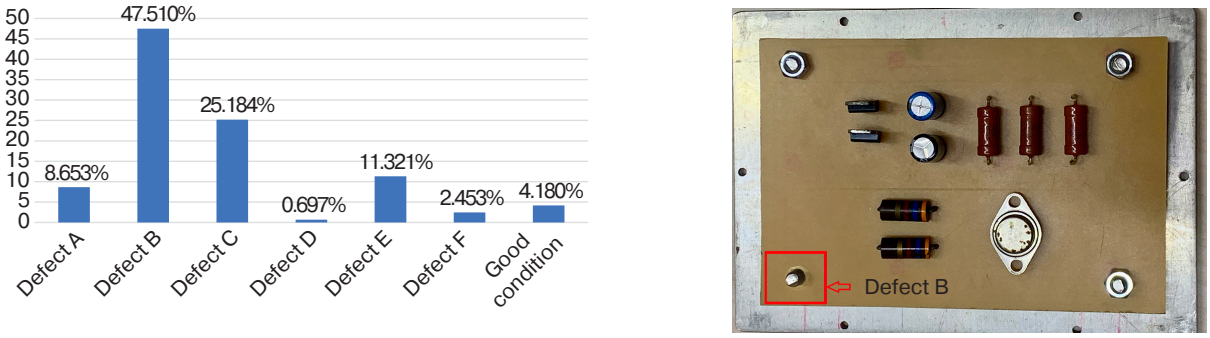


Fig. 21. Report on the technical condition of the tested model with defect B

[5] https://colab.google/. Accessed March 23, 2026.

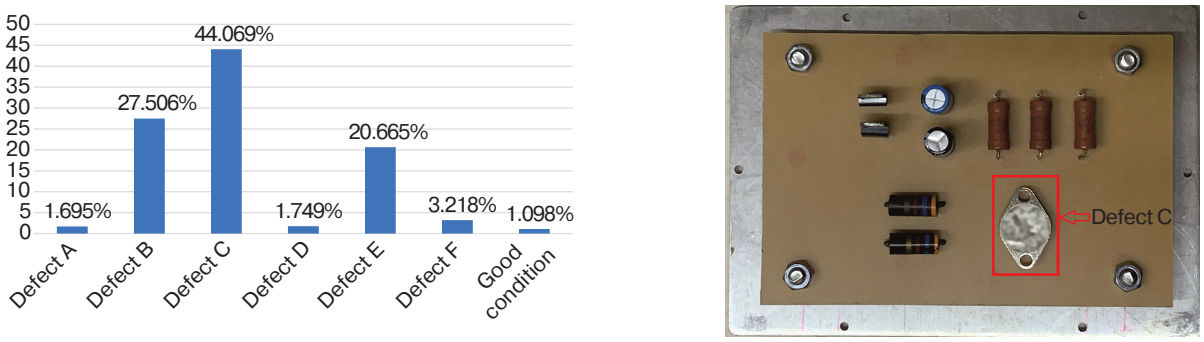


Fig. 22. Report on the technical condition of the tested model with defect C

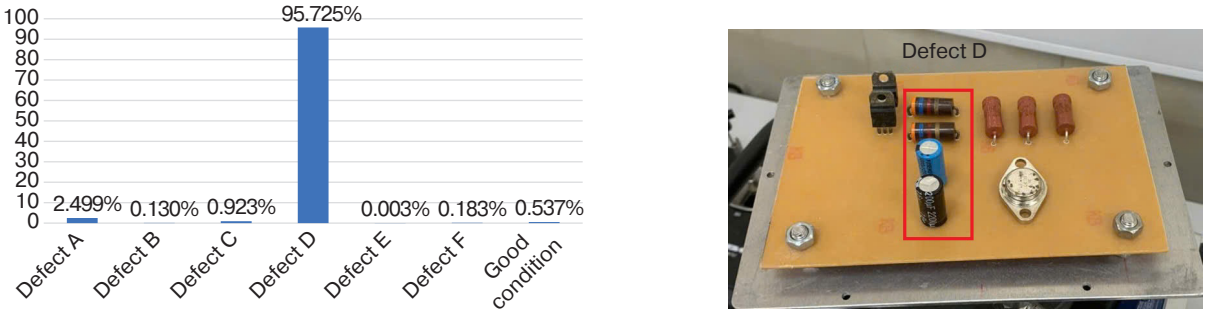


Fig. 23. Report on the technical condition of the tested model with defect D

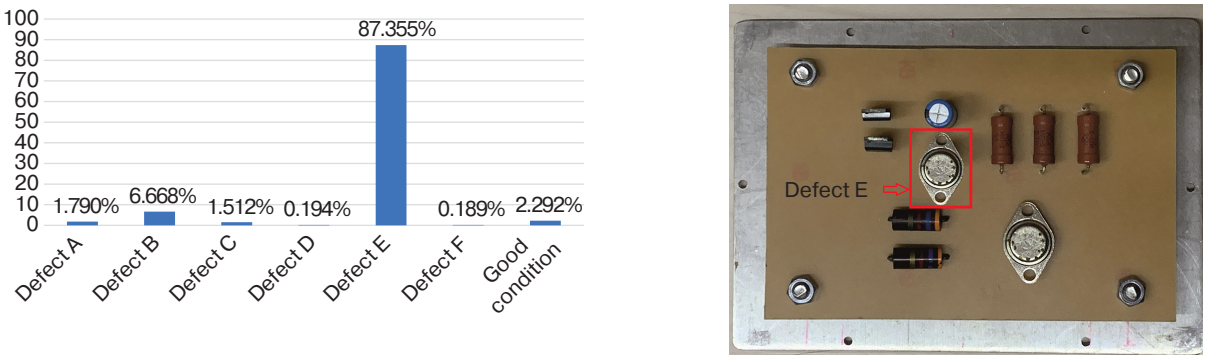


Fig. 24. Report on the technical condition of the tested model with defect E

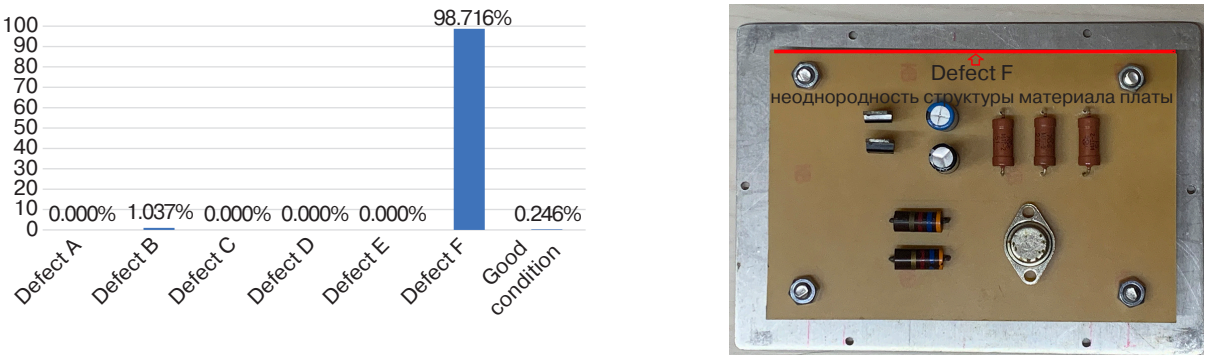


Fig. 25. Report on the technical condition of the tested model with defect F

The defect classification results show that the DNN model can recognize each type of defect in the PCBA design with a fairly high probability:

- defect A: 93.7%;
- defect B: 47.5%;
- defect C: 44.1%;
- defect D: 95.7%;
- defect E: 87.4%;
- defect F: 98.7%.

The developed DNN model has therefore confirmed the possibility of accurately identifying hidden defects in the PCBA design through analysis of the mechanical AFC of the block obtained during vibration testing.

CONCLUSIONS

The paper presents a vibration-based diagnostic method for PCBA structures that has been developed and experimentally tested. This approach involves analyzing the mechanical AFC of the block and using DNN to classify defects.

The key findings of the study are as follows:

- Numerical and field experiments have confirmed that the mechanical AFC of block structures under vibration stress is affected by the design defects listed in Section 2.
- A vibration diagnostics method for PCBA structures has been developed that integrates numerical modeling of mechanical processes with field testing. This enables latent defects in PCBA to be detected and classified with high accuracy.
- A DNN model has been implemented that achieves 99.6% accuracy in defect classification.
- The feasibility of applying the proposed method for PCBA diagnostics during the manufacturing and operational stages, without opening the sealed enclosures, has been demonstrated.

Thus, the proposed method for diagnosing the effects of vibration on PCBA can be used to make the diagnostic process more efficient by improving the accuracy with which the technical condition is determined. Using numerical modeling and artificial neural networks automates the procedure for analyzing mechanical AFC characteristics. The results obtained in this study can form the basis for further scientific research and development into vibration diagnostic methods for PCBA structures.

Authors' contribution. All authors contributed equally to the research work.

REFERENCES

1. Uvaysov S.U., Dolmatov A.V., Vo T.H., Luu N.T., Nguyen C.D. Diagnostics of structural integrity violations of avionics during impact tests. *Russian Technological Journal.* 2024;12(2):28–38. https://doi.org/10.32362/2500-316X-2024-12-2-28-38
2. Dolmatov A.V., Suleimanov S.P., Uvaisov R.I. Diagnosing the integrity of the electronic equipment design. *Problemy kachestva, bezopasnosti i diagnostiki v usloviyakh informatsionnogo obshchestva = Problems of Quality, Safety and Diagnostics in the Information Society*. 2004;1:99–100 (in Russ.). https://elibrary.ru/pfgsrx
3. Kostyukov A.S., Bashkirov A.V., Gostev M.Yu., Demikhova A.S., Pirogova Yu.A. Comparative analysis of software complexes for determination of mechanical characteristics of RED. *Vestnik Voronezhskogo gosudarstvennogo tekhnicheskogo universiteta = Bulletin of the Voronezh State Technical University*. 2020;16(4):117–126 (in Russ.). https://doi.org/10.25987/VSTU.2020.16.4.016
4. Shalumov A.S. ASONIKA – Russian CAD electronics in terms of virtual tests. *Elektronika: Nauka, tekhnologiya, biznes = Electronics: Science, Technology, Business*. 2022;3:82–83 (in Russ.). https://doi.org/10.22184/1992-4178.2022.214.3.82.83
5. Rezchikova E.V., Luong Q.L. Vibration protection of on-board electronic equipment from external influences. *Kontrol'. Diagnostika = Testing. Diagnostics*. 2021;24(11):22–30 (in Russ.). https://doi.org/10.14489/td.2021.11.pp.022-030
6. Belyaev A.A., Kononov D.P., Krotov S.V. Problems of diagnostics of modern diesel locomotive engines. *Byulleten' rezul'tatov nauchnykh issledovanii = Bulletin of Scientific Research Results*. 2024;1:7–20 (in Russ.). https://doi.org/10.20295/2223-9987-2024-01-7-20
7. Bityukov V.K., Dolmatov A.V., Zadernovsky A.A., Starikovsky A.I., Uvaysov R.M. Calculation permissible deviations of vibration accelerations of printed circuit assemblies by simulation modeling. *Russian Technological Journal.* 2023;11(6):28–38. https://doi.org/10.32362/2500-316X-2023-11-6-28-38
8. Zang V.T., Dao An.Q., Pham L.Q.H., Nguyen V.D., Nguyen V.T., Uvaysova A.S. Feature of application of JTAG technology in diagnostics of printed units. In: *Innovative, Information and Communication Technologies:* Proceedings of the 17th International Scientific and Practical Conference. 2020. P. 427–431 (in Russ.). https://elibrary.ru/xpkskk
9. Nekhoroshkov D.P., Yusipov M.R., Sotnikova S.Y. Modeling of thermal and mechanical processes of the meteorological satellite unit. *Innovatsionnye, informatsionnye i kommunikatsionnye tekhnologii = Innovative, Information and Communication Technologies.* 2019;1:309–313 (in Russ.). https://elibrary.ru/ydlitd
10. Isabekova T.I., Savzikhanova S.E. Monte Carlo simulation of reliability of electronic components of very large-scale integrated circuits. *Vestnik Dagestanskogo gosudarstvennogo tekhnicheskogo universiteta. Tekhnicheskie nauki = Herald of Dagestan State Technical University. Technical Sciences*. 2025;52(3):49–60 (in Russ.). https://doi.org/10.21822/2073-6185-2025-52-3-49-60

11. Bach Phi Duong, Jong-Myon Kim. Prognosis of remaining bearing life with vibration signals using a sequential Monte Carlo framework. *J. Acoust. Soc. Am*. 2019;146(4):EL358–EL363. https://doi.org/10.1121/1.5129076
12. Kashnikova A.P., Belyaeva M.B. The Monte Carlo method in the problems of modeling processes and systems. *Modern Science*. 2021;1-2:358–362 (in Russ.). https://elibrary.ru/ywleen
13. Kofanov Yu.N., Sotnikova S.Yu. Improving the accuracy of modeling based on identification. *Innovatsii na osnove informatsionnykh i kommunikatsionnykh tekhnologii = Innovations Based on Information and Communication Technologies*. 2015;1:176–178 (in Russ.). https://elibrary.ru/ukehyt
14. Zhuravlev A.O., Polyakov A.O., Andrikov D.A. Vibration Diagnostic Methods from Methods of Obtaining Data to Processing It Using Modern Means. *Vestnik Rossiiskogo universiteta druzhby narodov. Seriya: Inzhenernye issledovaniya = RUDN Journal of Engineering Research*. 2024;25(4):380–396 (in Russ.). https://doi.org/10.22363/2312-8143-2024-25-4-380-396
15. Irzaev G.H. Provision and evaluation of testability of the design of complex radio-electronic meansdevices at the stages of their design. *Kontrol'. Diagnostika = Testing. Diagnostics*. 2023;26(5):34–41 (in Russ.). https://doi.org/10.14489/td.2023.05.pp.034-041
16. Rakhmanov P.A., Garryev S.S., Redzhepgeldiev S.S. Deep learning and machine learning: key differences and impact on modern technologies. *Vestnik Nauki*. 2024;4(73):651–654 (in Russ.). https://elibrary.ru/rnvyuu
17. Poletaev A.A., Yablokov A.E. Neural network method for technical diagnostics of food equipment based on vibration acceleration and vibration velocity spectra. *Izvestiya Tul'skogo gosudarstvennogo universiteta. Tekhnicheskie nauki = News of the Tula State University. Technical Sciences*. 2024;5:476–481 (in Russ.). https://www.elibrary.ru/nyfisf
18. Klyachkin V.N., Kuvaiskova Yu.E., Lomovtseva N.A. Diagnosting the condition of a technical object using classification machine learning classification. *Programmnye produkty i sistemy = Software & Systems*. 2021;4:572–578 (in Russ.). http://doi.org/10.15827/0236-235X.136.572-578

СПИСОК ЛИТЕРАТУРЫ

1. Увайсов С.У., Долматов А.В., Во Т.Х., Лыу Н.Т., Нгуен К.Д. Диагностика нарушений целостности конструкций бортовых радиоэлектронных средств при испытаниях на ударные воздействия. *Russian Technological Journal.* 2024;12(2):28–38. https://doi.org/10.32362/2500-316X-2024-12-2-28-38
2. Долматов А.В., Сулейманов С.П., Увайсов Р.И. Диагностирование целостности конструкции электронной аппаратуры. *Проблемы качества, безопасности и диагностики в условиях информационного общества.* 2004;1:99–100. https://elibrary.ru/pfgsrx
3. Костюков А.С., Башкиров А.В., Гостев М.Ю., Демихова А.С., Пирогова Ю.А. Сравнительный анализ программных комплексов для определения механических характеристик РЭС. *Вестник Воронежского государственного технического университета*. 2020;16(4):117–126. https://doi.org/10.25987/VSTU.2020.16.4.016
4. Шалумов А.С. АСОНИКА – российская САПР электроники в части виртуальных испытаний. *Электроника*: *Наука, технология, бизнес*. 2022;3:82–83. https://doi.org/10.22184/1992-4178.2022.214.3.82.83
5. Резчикова Е.В., Лыонг К.Л. Виброзащита бортовой электронной аппаратуры от внешних воздействий. *Контроль. Диагностика*. 2021;24(11):22–30. https://doi.org/10.14489/td.2021.11.pp.022-030
6. Беляев А.А., Кононов Д.П., Кротов С.В. Проблемы диагностики современных тепловозных двигателей. *Бюллетень результатов научных исследований.* 2024;1:7–20. https://doi.org/10.20295/2223-9987-2024-01-7-20
7. Битюков В.К., Долматов А.В., Задерновский А.А., Стариковский А.И., Увайсов Р.М. Расчет допустимых отклонений виброускорений печатных узлов методом имитационного моделирования. *Russian Technological Journal*. 2023;11(6):28–38. https://doi.org/10.32362/2500-316X-2023-11-6-28-38
8. Занг В.Т., Дао А.К., Фам Л.К.Х., Нгуен В.Д., Нгуен В.Т., Увайсова А.С. Особенность применения технологии JTAG в диагностике печатных узлов. В сб.: *Инновационные, информационные и коммуникационные технологии:* сборник трудов XVII Международной научно-практической конференции. 2020. С. 427–431. https://elibrary.ru/xpkskk
9. Нехорошков Д.П., Юсипов М.Р., Сотникова С.Ю. Моделирование тепловых и механических процессов блока метеорологического спутника. *Инновационные, информационные и коммуникационные технологии*. 2019;1:309–313. https://elibrary.ru/ydlitd
10. Исабекова Т.И., Савзиханова С.Э. Моделирование надежности электронных компонентов сверхбольших интегральных схем методом Монте-Карло. *Вестник Дагестанского государственного технического университета. Технические науки*. 2025;52(3):49–60. https://doi.org/10.21822/2073-6185-2025-52-3-49-60
11. Bach Phi Duong, Jong-Myon Kim. Prognosis of remaining bearing life with vibration signals using a sequential Monte Carlo framework. *J. Acoust. Soc. Am*. 2019;146(4):EL358–EL363. https://doi.org/10.1121/1.5129076
12. Кашникова А.П., Беляева М.Б. Метод Монте-Карло в задачах моделирования процессов и систем. *Modern Science*. 2021;1-2:358–362. https://elibrary.ru/ywleen
13. Кофанов Ю.Н., Сотникова С.Ю. Повышение точности моделирования на основе идентификации. *Инновации на основе информационных и коммуникационных технологий*. 2015;1:176–178. https://elibrary.ru/ukehyt
14. Журавлёв А.О., Поляков А.О., Андриков Д.А. *Методы вибродиагностики: от способов получения данных до их обработки современными средствами. Вестник РУДН. Серия*: *Инженерные исследования*. 2024;25(4):380–396. https://doi.org/10.22363/2312-8143-2024-25-4-380-396
15. Ирзаев Г.Х. Обеспечение и оценка контролепригодности конструкции сложных радиоэлектронных средств на этапах их проектирования. *Контроль. Диагностика*. 2023;26(5):34–41. https://doi.org/10.14489/td.2023.05.pp.034-041

16. Рахманов П.А., Гаррыев С.С., Реджепгелдиев С.С. Глубокое обучение и машинное обучение: ключевые отличия и влияние на современные технологии. *Вестник науки*. 2024;4(73):651–654. https://elibrary.ru/rnvyuu
17. Полетаев А.А., Яблоков А.Е. Нейросетевой метод технической диагностики оборудования по спектрам виброскорости и виброускорений. *Известия Тульского государственного университета. Технические науки*. 2024;5:476–481. https://www.elibrary.ru/nyfisf
18. Клячкин В.Н., Кувайскова Ю.Е., Ломовцева Н.А. Диагностика состояния технического объекта с помощью классификации методами машинного обучения. *Программные продукты и системы*. 2021;4:572–578. http://doi.org/10.15827/0236-235X.136.572-578

About the Authors

Saygid U. Uvaysov, Dr. Sci. (Eng.), Professor, Head of the Department of Design and Production of Radioelectronic Devices, Institute of Radio Electronics and Informatics, MIREA – Russian Technological University (78, Vernadskogo pr., Moscow, 119454 Russia). E-mail: uvajsov@mirea.ru. Scopus Author ID 55931417100, ResearcherID H-6746-2015, RSCI SPIN-code 3801-4816, https://orcid.org/0000-0003-1943-6819

Aleksey V. Dolmatov, Cand. Sci. (Eng.), Associate Professor, Department of Design and Production of Radioelectronic Devices, Institute of Radio Electronics and Informatics, MIREA – Russian Technological University (78, Vernadskogo pr., Moscow, 119454 Russia). E-mail: dolmatov@mirea.ru, RSCI SPIN-code 3887-2405, https://orcid.org/0000-0003-2969-2971

Vo The Hai, Postgraduate Student, Department of Design and Production of Radioelectronic Devices, Institute of Radio Electronics and Informatics, MIREA – Russian Technological University (78, Vernadskogo pr., Moscow, 119454 Russia). E-mail: thehai.ttrd@gmail.com. RSCI SPIN-code 5210-3214, https://orcid.org/0009-0009-7240-4374

Nguyen Duc Hai, Postgraduate Student, Department of Design and Production of Radioelectronic Devices, Institute of Radio Electronics and Informatics, MIREA – Russian Technological University (78, Vernadskogo pr., Moscow, 119454 Russia). E-mail: duchai09011997@gmail.com. RSCI SPIN-code 5614-9517, https://orcid.org/0009-0002-8109-9729

Pham Xuan Hanh, Postgraduate Student, Department of Design and Production of Radioelectronic Devices, Institute of Radio Electronics and Informatics, MIREA – Russian Technological University (78, Vernadskogo pr., Moscow, 119454 Russia). E-mail: phamxuanhanhld161@gmail.com. RSCI SPIN-code 8103-7664, https://orcid.org/0009-0002-6485-8404

Ruslan M. Uvaysov, Postgraduate Student, Department of Design and Production of Radioelectronic Devices, Institute of Radio Electronics and Informatics, MIREA – Russian Technological University (78, Vernadskogo pr., Moscow, 119454 Russia). E-mail: ruslan27398@yandex.ru. RSCI SPIN-code 5906-9978, https://orcid.org/0000-0003-0619-649X

Об авторах

Увайсов Сайгид Увайсович, д.т.н., профессор, заведующий кафедрой конструирования и производства радиоэлектронных средств, Институт радиоэлектроники и информатики, ФГБОУ ВО «МИРЭА – Российский технологический университет» (119454, Россия, Москва, пр-т Вернадского, д. 78). E-mail: uvajsov@mirea.ru. Scopus Author ID 55931417100, ResearcherID H-6746-2015, SPIN-код РИНЦ 3801-4816, https://orcid.org/0000-0003-1943-6819

Долматов Алексей Вячеславович, к.т.н., доцент, кафедра конструирования и производства радиоэлектронных средств, Институт радиоэлектроники и информатики, ФГБОУ ВО «МИРЭА – Российский технологический университет» (119454, Россия, Москва, пр-т Вернадского, д. 78). E-mail: dolmatov@mirea.ru. SPIN-код РИНЦ 3887-2405, https://orcid.org/0000-0003-2969-2971

Во Тхе Хай, аспирант, кафедра конструирования и производства радиоэлектронных средств, Институт радиоэлектроники и информатики, ФГБОУ ВО «МИРЭА – Российский технологический университет» (119454, Россия, Москва, пр-т Вернадского, д. 78). E-mail: thehai.ttrd@gmail.com. SPIN-код РИНЦ 5210-3214, https://orcid.org/0009-0009-7240-4374

Нгуен Дык Хай, аспирант, кафедра конструирования и производства радиоэлектронных средств, Институт радиоэлектроники и информатики, ФГБОУ ВО «МИРЭА – Российский технологический университет» (119454, Россия, Москва, пр-т Вернадского, д. 78). E-mail: duchai09011997@gmail.com. SPIN-код РИНЦ 5614-9517, https://orcid.org/0009-0002-8109-9729

Фам Суан Хань, аспирант, кафедра конструирования и производства радиоэлектронных средств, Институт радиоэлектроники и информатики, ФГБОУ ВО «МИРЭА – Российский технологический университет» (119454, Россия, Москва, пр-т Вернадского, д. 78). E-mail: phamxuanhanhld161@gmail.com. SPIN-код РИНЦ 8103-7664, https://orcid.org/0009-0002-6485-8404

Увайсов Руслан Магомедович, аспирант, кафедра конструирования и производства радиоэлектронных средств, Институт радиоэлектроники и информатики, ФГБОУ ВО «МИРЭА – Российский технологический университет» (119454, Россия, Москва, пр-т Вернадского, д. 78). E-mail: ruslan27398@yandex.ru. SPIN-код РИНЦ 5906-9978, https://orcid.org/0000-0003-0619-649X

Translated from Russian into English by K. Nazarov
Edited for English language and spelling by Thomas A. Beavitt

Modern radio engineering and telecommunication systems

Современные радиотехнические и телекоммуникационные системы

UDC 004.032.26+537.874.4+550.837.8
https://doi.org/10.32362/2500-316X-2026-14-3-60-71
EDN LBUPEG

RESEARCH ARTICLE

Neurovisual recognition of signal radio images

Vladislav A. Kozhemyako [@], Alexey D. Yarlykov

MIREA – Russian Technological University, Moscow, 119454 Russia
[@] Corresponding author, e-mail: kozhemyako@mirea.ru

• **Submitted:** 03.07.2025 • **Revised:** 04.12.2025 • **Accepted:** 23.03.2026

Abstract

Objectives. The study set out to solve the problem of radiovision classification of objects based on identified features by developing a combined neurovision algorithm for real-time recognition of signal radio images of objects using machine learning (ML) technologies and a fully connected neural network with data augmentation, as well as to improve the probability of correct classification in neurovision signal processing.

Methods. In the study, several methods were used: electrodynamic modeling, machine learning (linear regression, classification, and Random Forest), and deep learning (fully connected neural networks). The bootstrap aggregating (bagging) technique was also employed. An assessment of object classification accuracy metrics and statistical criteria for the reproducibility of radio images was carried out.

Results. A combined neurovision object recognition method was developed that demonstrated a probability of correct classification of at least 0.97 for any of the objects transmitted for training with specified form factors when using augmented data. Data augmentation was shown to increase the neural network's probability of correct classification by 0.04. The obtained results confirm the adequacy of neural network approaches compared to classical ML methods for neurovision object recognition, particularly when dealing with a limited base dataset of objects for neural network training. The proposed method was tested for basic classification of spherical and cubic object models in the centimeter radio frequency range.

Conclusions. Neural networks with data augmentation demonstrate a probability of correct classification exceeding 0.97 for neurovision recognition of radio images as compared to neural networks without data augmentation (0.04 lower) and traditional ML methods (0.13 lower). Although ML methods are inferior to neural networks in radio image reproducibility, they remain indispensable in cases where computational resources are limited. For real-world applications, database expansion through field experiments and the implementation of hybrid neural network architectures are required.

Keywords: neuroimaging method, signal radio image, neural network, machine learning, signature classification, linear regression, random forest, electrodynamic modeling

For citation: Kozhemyako V.A., Yarlykov A.D. Neurovisual recognition of signal radio images. *Russian Technological Journal.* 2026;14(3):60–71. https://doi.org/10.32362/2500-316X-2026-14-3-60-71, https://www.elibrary.ru/LBUPEG

Financial disclosure: The authors have no financial or proprietary interest in any material or method mentioned.

The authors declare no conflicts of interest.

НАУЧНАЯ СТАТЬЯ

Нейровизионное распознавание сигнальных радиоизображений

В.А. Кожемяко [@], **А.Д. Ярлыков**

МИРЭА – Российский технологический университет, Москва, 119454 Россия
[@] *Автор для переписки, e-mail: kozhemyako@mirea.ru*

• **Поступила:** 03.07.2025 • **Доработана:** 04.12.2025 • **Принята к опубликованию:** 23.03.2026

Резюме

Цели. Целями работы являются: создание комбинированного нейровизионного алгоритма распознавания сигнальных радиоизображений объектов в режиме реального времени с использованием технологий машинного обучения и нейронной сети с полносвязной архитектурой и аугментацией данных; повышение вероятности правильной классификации при нейровизионной обработке сигналов.

Методы. В работе применены методы электродинамического моделирования, машинного обучения (линейная регрессия, классификация, случайный лес) и глубокого обучения (полносвязные нейронные сети). Применена техника бэггинга. Проведена оценка показателей точности классификации объектов и статистических критериев воспроизводимости радиоизображений.

Результаты. Разработан комбинированный нейровизионный метод распознавания объектов, показавший вероятность правильной классификации любого из переданных к обучению объектов с заданными форм-факторами не менее 0.97 при использовании аугментированных данных. Показано, что аугментация данных повышает вероятность правильной классификации нейронной сетью на 0.04. Полученные результаты подтвердили адекватность нейросетевых методов для задач нейровизионного распознавания объектов по сравнению с методами машинного обучения, прежде всего, при ограниченной базовой выборке объектов для обучения нейронной сети. Предложенный метод исследован для базисной классификации сферических и кубических моделей объектов в сантиметровом радиочастотном диапазоне частот.

Выводы. Нейронные сети с аугментацией данных демонстрируют вероятность правильной классификации свыше 0.97 в задачах нейровизионного распознавания радиоизображений в сравнении с нейронными сетями без аугментации данных (ниже на 0.04) и методами машинного обучения (ниже на 0.13). Методы машинного обучения уступают нейросетям в воспроизводимости радиоизображений, однако являются незаменимыми при ограниченных ресурсах вычислительной мощности. Для применения в реальных условиях требуются расширение базы данных за счет натурных экспериментов и применение гибридных архитектур нейронных сетей.

Ключевые слова: нейровизионный метод, сигнальное радиоизображение, нейронная сеть, машинное обучение, классификация радиоизображений, линейная регрессия, метод случайного леса, электродинамическое моделирование

Для цитирования: Кожемяко В.А., Ярлыков А.Д. Нейровизионное распознавание сигнальных радиоизображений. *Russian Technological Journal*. 2026;14(3):60–71. https://doi.org/10.32362/2500-316X-2026-14-3-60-71, https://www.elibrary.ru/LBUPEG

Прозрачность финансовой деятельности: Авторы не имеют финансовой заинтересованности в представленных материалах или методах.

Авторы заявляют об отсутствии конфликта интересов.

INTRODUCTION

One of the key tasks in the field of radiovision consists in the detection and classification of objects in space. It is well known that recognizing the shape, size and radiophysical identifiers of objects in real time using neural network algorithms requires a combined approach to signal processing that harnesses the complementary strengths of neural network and machine learning (ML) approaches. In some cases, these technologies offer clear advantages over optimal processing methods [1]. The correct classification of objects based on their radio images, which exhibit a gradient, sharp transition, and heterogeneous surface structure, is critical in a number of scientific and practical areas. One example is the creation of synthetic aperture radars for small satellite constellations for Earth observation, such as the Kondor, which are designed for obtaining high- and medium-resolution radio images.[1]

Unlike the agreed methods of wavelet filtering of ultra-wideband signals in the frequency-time domain [2], modern solutions for radio image recognition focus more on processing data using ML technologies and neural network methods based on classifiable features. These methods do not require assumptions about the probability distribution of data [1]. Neural networks can be used to analyze multiple signal parameters simultaneously, thus significantly increasing processing speed and reducing the probability of error by 5–15%.

The paper explores the use of ML techniques, such as linear regression, classification and random forest, along with a neural network with a fully connected architecture, to recognize radio images of objects of various radiophysical natures, which are described by the superposition of basic stereometric shapes. The main focus is on comparing these methods and exploring the potential for combining them.

The neurovision study consists of three stages. The first stage involves creating a database of radio signal images using the *Ansys HFSS* electrodynamic simulation environment.[2] The second stage involves training ML models and creating a neurovision algorithm, as well as further training of the neural network. The third stage involves analyzing the accuracy of object classification and statistical criteria for radio image reproducibility.

The combined neurovision method proposed in this paper is of practical interest in the development of digital processing technologies for secondary signals from aerospace and ground-based radio visors. In such scenarios, given the limited computing resources available on board, it is essential to minimize the time taken by the algorithm for object recognition [3].

1. CREATING A DATABASE OF RADIO IMAGES FOR NEUROVISION TRAINING

In the context of ML methodologies or neural networks, the establishment of a database is imperative for training a particular model or architecture. The initial step therefore involves the creation and population of such a database with signal responses from designated objects. To conduct a software-numerical experiment involving the classification of objects by shape based on signal responses, a database of radio responses from basic shapes was prepared using the *Ansys HFSS* electrodynamic simulation environment [4].

The experimental model created in the *Ansys HFSS* environment to generate radio images is depicted in (Fig. 1). Horn antennas were synthesized to radiate and receive at frequencies between 0.5 and 2 GHz. These antennas are located equidistantly through a radio-absorbing partition, with the aperture rotated at an angle of 60° to the Y-axis. The object under study, represented by a basic stereometric figure, is placed on the line of sight (LOS) passing through the plane of the radio-absorbing partition at a distance corresponding to the Fraunhofer zone. Meanwhile, the receiving antenna detects the electrical component of the alternating field scattered by the basic figure's signature—that is to say, the radio signal image. Two types of scalable objects are selected for the experiment: five cube variants and five sphere variants. The cubes have edge lengths of 0.35, 0.4, 0.45, 0.5, and 0.55 m, respectively. The spheres have diameters of 0.35, 0.4, 0.45, 0.5, and 0.55 m, respectively. The base objects are stereometric figures with perfectly conductive surfaces.

The object signature is irradiated by a radiovision signal using a Gaussian monocycle model (Fig. 2) [4]. The monocycle generates an ultra-wideband spectrum ranging from 0.7 to 1.0 GHz, equivalent to wavelengths between 42 and 30 cm. The objects under study have scalable dimensions ranging from 35 to 55 cm.

The experimental model shown in Fig. 1 is used to register the signal scattered by the signature of a radio and television object. To do this, a Gaussian monocycle radiation pattern must be generated as shown in Fig. 2. The specified amplitude spectrum of the Gaussian monocycle is calculated in the *Ansys HFSS* electrodynamic simulation environment. Upon successful completion of the calculation, *Ansys HFSS* displays animations of frequency-time nomograms of the near and far fields, which demonstrate the propagation of radiovision

[1] User's Guide to Earth Remote Sensing Data Obtained by the Kondor-FKA Space System – 2023. *Nauchnyi Tsentr Operativnogo Monitoringa Zemli* (Scientific Center for Operational Monitoring of the Earth). https://ntsomz.ru/wp-content/uploads/2023/05/2023.02.17.rukovodstvo.pol_.zovatela.kondor-fka.dla_.saita_.pdf. Accessed August 14, 2025. (In Russ.).

[2] https://www.ansys.com/products/. Accessed August 14, 2025.

probing and scattered signals [5]. In the simulation, the animation calculation duration is set to 25 ns. Figure 3 shows freeze frames of four nomograms displaying the distribution of the electrical component of the scattered signal signature of a radiovision pulse field from:

- a cube with an edge length of 0.55 m at 4950 ps (Fig. 3a) and at 11010 ps (Fig. 3b);
- a sphere with a diameter of 0.35 m at 5520 ps (Fig. 3c) and at 12030 ps (Fig. 3d).

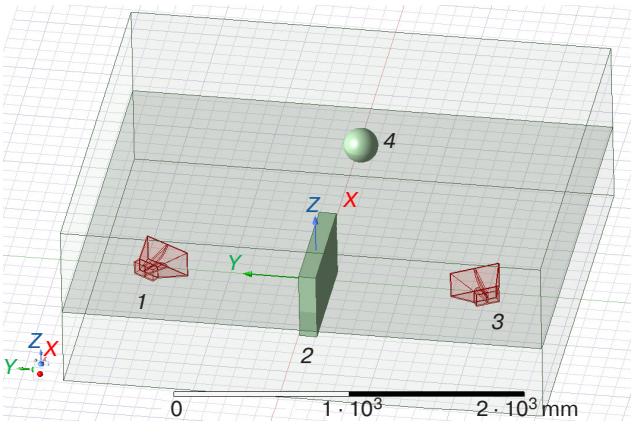


Fig. 1. Software-numerical electrodynamic model of the experiment on the formation of radio images of objects: (*1*) receiving antenna, (*2*) radio-absorbing partition, (*3*) transmitting antenna, and (*4*) radiovision object specified by the signature of the base figure

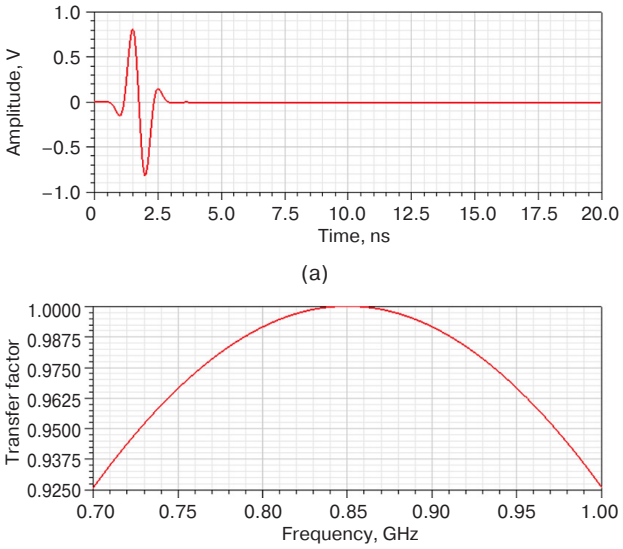


Fig. 2. Representation of a Gaussian monocycle in the time (a) and frequency (b) domains

In addition to animating nomograms of the propagation of incident and scattered components of electromagnetic fields, the model allows for the construction of radio profiles of the distribution of the scattered signature of an object's electric field in a selected direction. Figure 4 shows radio images of objects obtained through diffuse scattering by lateral surface inhomogeneities, with a cube having an edge length of 0.55 m and a sphere having a radius of 0.35 m [6].

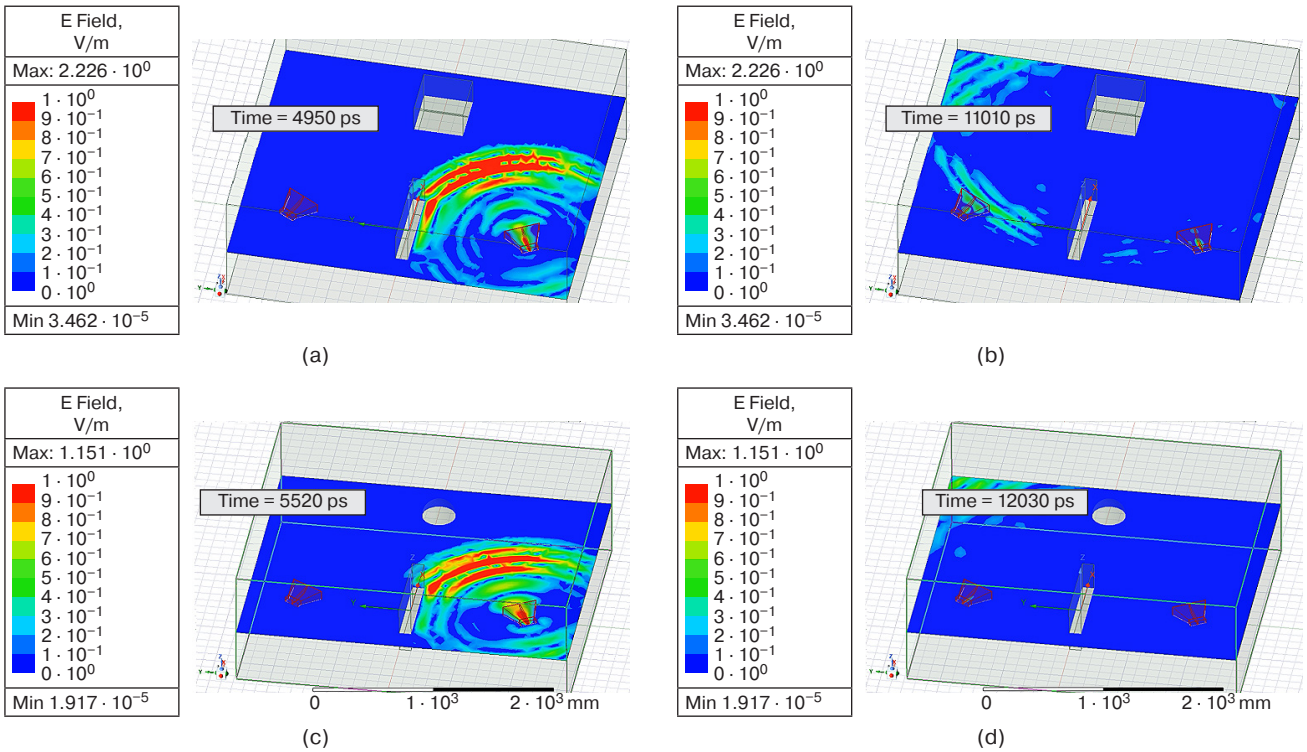


Fig. 3. Freeze frames of distribution nomograms scattered in the far zone of the field from a cube and a sphere

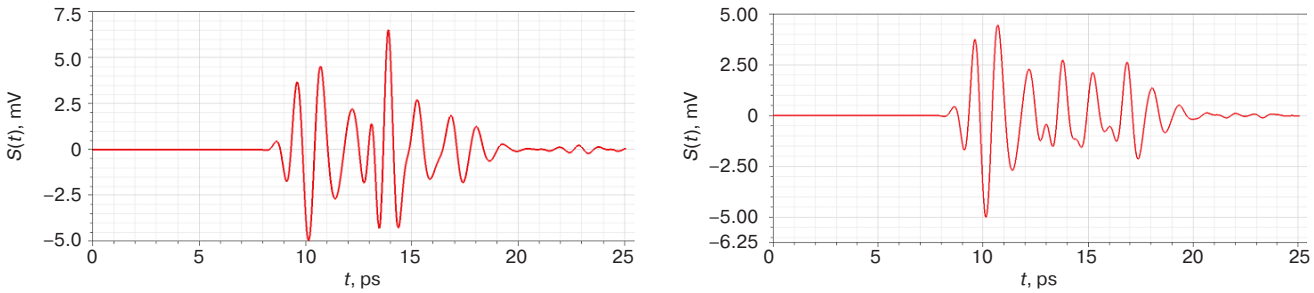


Fig. 4. Radio images of cubic (a) and spherical (b) objects, obtained from a given scattering direction determined by Snell's law

The radio images obtained can be exported from *Ansys HFSS* as a data sample comprising 965 instantaneous values of the reduced field strength in millivolts (mV), which are sampled at intervals of 0.025 ns. The arrays of software-numerical experimental values that make up the neurovision database, obtained for five spherical and five cubic objects with different form factors, are given in Table 1.

Table 1. Neurovision data of radio images with established form factors

Object type	Sphere (form factor D, m)					Cube (form factor L, m)				
Form factor	0.35	0.4	0.45	0.5	0.55	0.35	0.4	0.45	0.5	0.55
Time index with a step of 25 ps	Instantaneous dynamic values of impulse radio images, mV									
340	−0.0195	−0.0514	0.0945	0.1146	−1.7058	−1.5478	−0.8098	1.7553	0.3806	0.3287
341	−0.1693	−0.2056	−0.0405	−0.0189	−1.5988	1.3636	−0.4782	1.1756	−0.2834	−0.3413
342	−0.3451	−0.3844	−0.1967	−0.1740	−1.4304	−1.1189	−0.0748	0.5183	−0.9570	−1.0179
343	−0.5208	−0.5632	−0.3781	−0.3554	−1.2007	−0.7913	0.3640	0.5183	−1.6125	−1.6716
344	−0.7138	−0.7572	−0.5595	−0.5367	−0.8847	−0.4636	0.8264	−0.1699	−2.2679	−2.3252
345	−0.9105	−0.9534	−0.7567	−0.7349	−0.5687	−0.0658	1.2975	−0.8719	−2.8737	−2.9239
346	−1.1041	−1.1445	−0.9564	−0.9362	−0.1779	0.3665	1.7685	−1.5597	−3.4262	−3.4663
347	−1.2774	−1.3120	−1.1510	−1.1332	0.2514	0.8218	2.2189	−2.2475	−3.9123	−3.9395
348	−1.4506	−1.4796	−1.3222	−1.3076	0.7083	1.2854	2.6359	−2.8865	−4.2756	−4.2872
349	−1.5870	−1.6069	−1.4934	−1.4819	1.1808	1.7489	3.0048	−3.4709	−4.6389	−4.6349
350	−1.6879	−1.6966	−1.6241	−1.6166	1.6533	2.1925	3.2758	−3.9862	−4.8615	−4.8405
351	−1.7456	−1.7412	−1.7168	−1.7135	2.1131	2.6034	3.5468	−4.3722	−4.9819	−4.4429
…	…	…	…	…	…	…	…	…	…	…
963	−0.0575	−0.0021	−0.0202	−0.0125	−0.0124	−0.0505	0.0309	−0.0059	−0.0279	0.0012
964	−0.0501	−0.0085	−0.0173	−0.0177	−0.0102	−0.0434	0.0329	−0.0094	−0.0274	0.0008
965	−0.0501	−0.0134	−0.0173	−0.0222	−0.0102	−0.0434	0.0347	−0.0119	−0.0255	0.0008

The resulting neurovision database is comprised of a data frame in which the categories for the neural network model are radio images of objects, while the indices are instantaneous dynamic values of the signal radio profile. A special Python program was developed to represent and process the neurovision database of responses in a specified format using ML and neural network methods. The resulting database is entered into the program as a separate DataFrame data type using the Pandas module (a library for processing and analyzing structured data). This allows two-dimensional arrays of information containing both numerical and string data to be stored. The Pandas module additionally provides researchers with access to all stored data. Since it works with a copy of the original radio and television data, this module uses the computing device's memory resources to store and process information. This means that the information obtained in the form of a DataFrame can be used for ML and neural network methods [7].

2. ANALYSIS OF THE EFFECTIVENESS OF ML METHODS IN RADIO IMAGE RECOGNITION

As part of the conducted neurovision research, the prediction results obtained using three different ML methods are compared. The choice of algorithms is determined by the aim of covering classical linear models, as well as more complex non-linear and ensemble approaches. The selected algorithms are linear regression, a classification method based on the support vector machine, and a random forest method. Linear regression is chosen as the basic, interpretable method, while the classification method comprises a powerful algorithm for working with high-dimensional data, and the random forest method functions as a benchmark ensemble method that is resistant to overfitting [8]. The ML algorithm models are built using the scikit-learn Python library (an open-source library for predictive data analysis).

We consider the semantic characteristics of each algorithm, which are as follows:

A. In the case of the linear regression method, the equation $Y = aX + b$ is used, where a and b are coefficients that are determined during training [9]. However, since the output parameter must be a numerical integer, the values of the category names should be replaced; for example, "sphere" and "cube" are replaced with the features "1" and "2," respectively.

B. The classification method differs from regression mainly insofar as regression has no classes, so it is the possible value of the output parameter that is predicted. In contrast, classification has a strictly limited number of classes or categories, as does the sample used to provide an answer [10].

C. Random Forest is a more complex version of simple ensemble models (Fig. 5). The main difference from previous methods consists in the random selection of object features for training from the general database set and distributed into smaller subsamples (datasets). Furthermore, features of the same object may appear in several new subsamples [9]. Thus, unlike classical ensembles, which use a system of all possible features, the random forest method enables the selection of a feature on which the neural network model will be strictly based.

As shown in Fig. 5, the random forest method is notable for its ability to artificially boost the database of signal responses used for training, thereby minimizing the likelihood of neurovision recognition errors.

Five additional radio images of objects are obtained for comparison of the considered algorithms: three from spherical bases with form factors of 0.35, 0.375, and 0.4 m, and two from cubic bases with form factors of 0.45 and 0.475 m. The data is then loaded into a database of trained software models for evaluation of neurovision recognition accuracy and identification of error probabilities. Table 2 shows the results of the ML-based neurovision model for recognizing objects by form factor in the absence of interference.

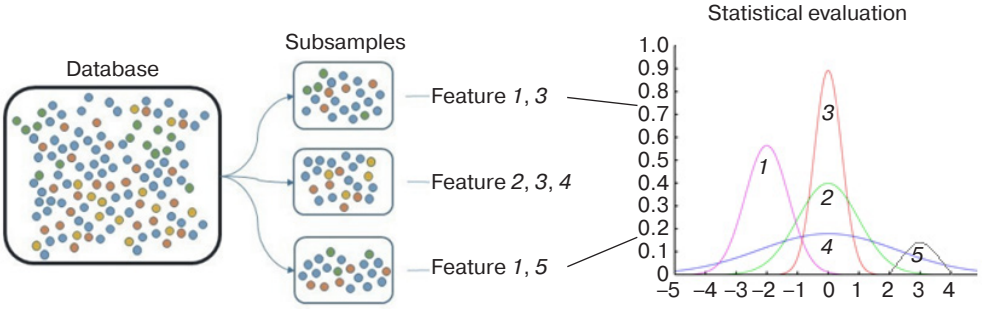


Fig. 5. Formation of datasets using the random forest method based on strictly selected spectral-temporal radiovision features

Table 2. Probability of correct classification for trained models of neurovision recognition of objects with a given form factor

Method	Linear progression	Classification	Random forest
Probability of correct classification	0.3	0.72	0.85

The study employs two complementary metrics to evaluate the effectiveness of ML models and neural networks: the correlation coefficient and the probability of correct classification.

The correlation coefficient (r) measures the similarity between the model's prediction and the reference sample, and is used as a threshold criterion to determine reliability. The established threshold value of $r = 0.9$ is decisive: if the calculated coefficient exceeds this value, the classification result is considered reliable and accepted as correct. Conversely, if $r < 0.9$, the result is deemed insufficiently reliable, suggesting a potential model error.

The probability of correct classification is defined as the share of test cases for which the correlation coefficient exceeds the threshold of 0.9. This metric, which reflects the model's overall ability to perform correct recognitions, is calculated as the ratio of the number of successfully classified examples (with $r \geq 0.9$) to the total test sample size, N_{total}:

$$P = N_{r \geq 0.9}/N_{total}.$$

The highest probability of correct classification achieved using the above ML methods is 0.85 for the random forest method. However, in the absence of sufficient training data, this value can lead to unsatisfactory results compared to wavelet analysis methods [11].

3. ANALYSIS OF THE EFFECTIVENESS OF NEURAL NETWORK APPLICATION IN RADIO IMAGE RECOGNITION

The developed neural network architecture is a fully connected, multilayer neural network consisting of three layers: two hidden layers and one output layer (Fig. 6) [12]. The network receives 965 instantaneous values of the reduced field strength in millivolts (mV) at a sampling rate of 25 picoseconds (ps). The first hidden layer, which consists of 16 neurons, converts the input signal from the received dimension into 16 features. The second hidden layer, consisting of eight neurons, converts the input signal from the 16 features into eight features. The output layer consists of one neuron that converts the input signal from eight features into one output value.

A three-layer fully connected neural network architecture is chosen to strike a balance between computational efficiency and model stability against overfitting with a limited amount of initial training data. Given the small size of the dataset, increasing the number of layers would significantly increase the number of trainable parameters and reduce the model's ability to generalize. Thus, the adopted architecture provides sufficient expressive power to extract feature hierarchies from radio signals while maintaining the computational efficiency required for real-time tasks.

The neural network is implemented using object-oriented programming principles.[3] A snippet of the code listing describing the neural network is shown in Fig. 7. This demonstrates the definition of a neural network class (Fig. 6) with three fully connected layers [13].

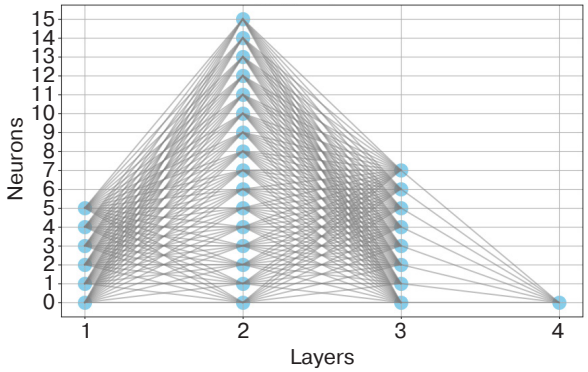


Fig. 6. Architecture of a neurovision network with three fully connected layers

The code initializes the weight matrices (**W1**, **W2**, and **W3**) and the bias vectors (**b1**, **b2**, and **b3**) with their respective dimensions. The forward function, which is a method of the neural network (NN) class, implements direct signal propagation through layers with rectified linear unit (ReLU) activation functions and a sigmoid function at the output.

The $\mathrm{ReLU} = \begin{cases} x, & x > 0, \\ 0, & x < 0 \end{cases}$ operator is selected as the activation function for the neurons of the 1st and 2nd layers [14].

[3] Bosenko T.M. *Fundamentals of Object-Oriented Analysis and Programming in Python*: A study guide. Moscow: Moscow City Pedagogical University; 2023, 80 p. (In Russ.).

```
class NN:  1 usage
    def __init__(self, input_size):
        self.W1 = tf.Variable(tf.random.normal([input_size, 16]), name='weight1')
        self.b1 = tf.Variable(tf.zeros([16]), name='bias1')
        self.W2 = tf.Variable(tf.random.normal([16, 8]), name='weight2')
        self.b2 = tf.Variable(tf.zeros([8]), name='bias2')
        self.W3 = tf.Variable(tf.random.normal([8, 1]), name='weight3')
        self.b3 = tf.Variable(tf.zeros([1]), name='bias3')

    def forward(self, x):  3 usages (1 dynamic)
        z1 = tf.matmul(x, self.W1) + self.b1
        a1 = tf.nn.relu(z1)
        z2 = tf.matmul(a1, self.W2) + self.b2
        a2 = tf.nn.relu(z2)
        z3 = tf.matmul(a2, self.W3) + self.b3
        return tf.sigmoid(z3)

    @property  2 usages (2 dynamic)
    def trainable_variables(self):
        return [self.W1, self.b1, self.W2, self.b2, self.W3, self.b3]
```

Fig. 7. Code listing snippet describing a neural network

First and foremost, this is because the ReLU operator enables the processing of data arrays, offers computational efficiency, eliminates the issue of vanishing gradients, and encourages sparse activations.

The weight matrices **W1**, **W2**, and **W3** are initialized with random values and the bias vectors **b1**, **b2**, and **b3** with zeros.

The data is processed through the neural network in a sequential manner through three layers. In the first layer, the input data is transformed linearly by multiplying it by the weight matrix **W1** and adding the bias vector **b1**. After this, the ReLU activation function is applied to the resulting value. In the second layer, the output of the previous layer is again multiplied by a weight matrix **W2** and a bias vector **b2**. The ReLU activation is also applied. Finally, in the output layer, the features obtained are multiplied by a third weight matrix **W3** and a bias **b3**. The final value is then normalized using the sigmoid activation function to produce a probability value indicating the likelihood of belonging to a specific class [12].

The network is trained using the backpropagation error algorithm. Binary cross-entropy is used as the loss function to measure the difference between predicted probabilities and true labels. The trainable_variables function returns a list of all parameters that can be modified by the optimizer. The Adam (short for Adaptive Moment Estimation) optimizer is used, which is a hybrid optimization algorithm that adjusts the learning rate for each parameter based on estimates of the first and second moment gradients. This algorithm is effective for handling large datasets [15]. The training dataset has previously been augmented by 30% using the bagging ML method in the program [9].

The learning process involves the following stages:

- Forward pass: Calculating predictions for input values.
- Error calculation: Comparing predictions with true values.
- Backward pass: Calculating gradients, i.e., directions of parameter change [12].
- Updating weights: Adjusting network parameters to reduce error.

To remove the additional dimension from the predictions, a squeeze function is applied, which is essential for calculating losses correctly. This allows coordinating the dimensions of the tensors of predictions and ground truth values in order to calculate the loss function [16].

The evaluation of the model's prediction quality is carried out using the binary_crossentropy function. This function calculates the binary cross-entropy between the actual labels and the predicted values, which provides a measure of the discrepancy between two probability distributions. This metric can be used to evaluate the effectiveness of the model in binary classification tasks by quantifying the difference between the predicted probability distribution and the true label distribution.

The database of neurovision data derived from radio imaging of objects (Table 1) was used as the training

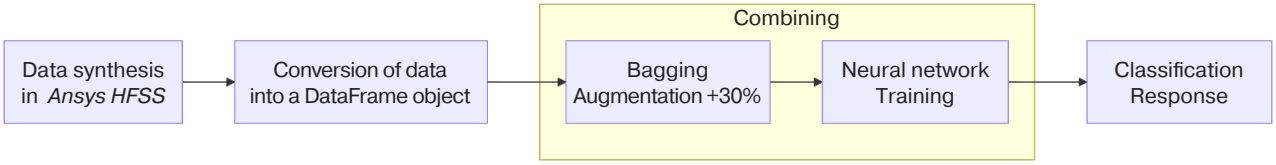


Fig. 8. Diagram of the combined algorithm for radio image recognition

dataset for the developed neural network model. To verify the accuracy of the calculations and access the generalizability of the model, a separate testing set of 5 radio images has been generated. The three images are from spherical objects with form factors of 0.35, 0.375, and 0.4 m; two images are from cubic objects with 0.45 and 0.475 form factors. In order to ensure the objectivity in the validation process, it is crucial that the testing objects with intermediate form factors of 0.375 and 0.475 m are not included in the training set. Although all data is generated in *Ansys HFSS*, the testing set consists of objects that are novel to the model. The neural network demonstrates a 0.97 accuracy rate on the independent testing set. This result is likely due to the fact that the training dataset is smaller than the desired. However, if the data sample is not artificially augmented using bagging, the same neural network's probability of correct classification would be only 0.93. This is a reduction of 0.04 compared to the original probability. This significantly improves the resulting accuracy of the model and demonstrates the efficacy of bagging in mitigating overfitting with small sample sizes.

Thus, the developed integrated algorithm implements a synergistic approach to radio imaging data processing by combining ML techniques with a neural network in a consistent manner. As shown in the diagram (Fig. 8), the procedure starts with the synthesis of source data in *Ansys HFSS*, which is then transformed into a DataFrame object. The key subsequent stage involves combining the bagging method for data augmentation with a fully connected neural network for classification. This allows the system to overcome limitations of a small initial sample size by artificially expanding the dataset and simultaneously utilizing the high recognition capabilities of the neural network model, resulting in high classification accuracy.

CONCLUSIONS

In the paper, a novel neurovision method has been developed and validated for recognizing radio signals from objects based on ML and neural networks. The method combines a fully-connected artificial neural network with a bagging ML approach, allowing for the real-time classification of basic stereometric shapes from radio images acquired in the centimeter-wavelength range. Experimental validation of the algorithm is conducted using a dataset of synthetic radio image signals generated by a custom software-based electrodynamic simulator in the *Ansys HFSS* environment. This enables the creation of a comprehensive dataset containing radio images of various objects with defined geometric properties. Using this database, the developed algorithm is able to recognize the objects utilizing a neurovision network consisting of three fully-connected layers and data augmentation techniques.

Based on a comparative analysis of ML techniques and neural networks, experimental findings have been obtained that confirm the effectiveness of the proposed integrated approach for radiovision tasks:

- The developed three-layer fully connected neural network with ReLU and sigmoid activation functions achieves a 0.97 probability of correct classification for spherical and cubic objects.
- It is shown that using data augmentation techniques such as bagging to increase the training dataset by 30% enhances the neural network's accuracy by 0.04, from 0.93 to 0.97, effectively addressing overfitting in small datasets.
- It is demonstrated that, out of the ML techniques under consideration, the random forest algorithm yields the highest probability of correct classification (0.85), which is 0.55 greater than that of linear regression (0.3) and 0.12 greater than the classification technique (0.72), although it is inferior to neural network approaches.
- The combined algorithm that combines the bagging approach with a neural network using data augmentation for classification can be used to achieve a 0.97 probability of correct classification.

The proposed combined approach, which demonstrates efficacy through the synergistic application of the ML algorithm (bagging) for data augmentation and a neural network for classification, successfully overcomes the limitations of each technique when applied individually. At the same time, this combined approach has several limitations, including dependence on the accuracy of synthetic data and the necessity to validate objects with intricate geometries under interference. These limitations determine the avenues for future research.

Authors' contributions

V.A. Kozhemyako—formulation of aims and objectives, practical research of the combined neurovision method and the algorithm for recognizing signal radio images.

A.D. Yarlykov—formulation of the research plan, formulation of conclusions.

REFERENCES

1. Terletskii A.S., Terletskaya E.S. *Neironnye seti i iskusstvennyi intellekt*: *Osnovy neironnykh setei na yazyke Python* (*Neural Networks and Artificial Intelligence*: *Fundamentals of Neural Networks in Python*). Lipetsk: Lipetsk State Pedagogical P. Semenov-Tyan-Shansky University; 2023, 76 p. (In Russ.). https://www.elibrary.ru/ugipee
2. Torgaev S.N., Lezhnina I.A., Shul'ga I.D. *Prakticheskoe rukovodstvo po tsifrovoi obrabotke signalov*: *tsifrovye fil'try i obrabotka EHKG signalov* (*A Practical Guide to Digital Signal Processing*: *Digital Filters and ECG Signal Processing*). Tomsk: STG; 2020, 112 p. (In Russ.). https://www.elibrary.ru/rupfct
3. Tsaregorodtsev M.A. Multithreaded numerical implementation of cryptographic algorithms of parallel action for protecting confidential information in defense-industrial complexes during its processing, storage and transmission to the cloud storage of big data. *Al'manakh Permskogo voennogo instituta voisk natsional'noi gvardii = Almanac of the Perm Military Institute of the National Guard Troops*. 2023;4(12):104–112 (in Russ.). https://www.elibrary.ru/vvtbii
4. Kurushin A.A. *Gibridnoe modelirovanie v HFSS ANSYS* (*Hybrid Modeling in HFSS ANSYS*): A tutorial. Moscow: SOLON-Press; 2023, 292 p. (In Russ.).
5. Bankov S.E., Kurushin A.A. *Raschet antenna i SVCh struktur s pomoshch'yu HFSS Ansoft* (*Calculation of Antennas and Microwave Structures using HFSS Ansoft*). Moscow: RODNIK; 2009, 256 p. (In Russ.).
6. Kozhemyako V.A. Getting a signal response from an object in Ansys CAD. In: *Actual Problems and Prospects for the Development of Radio Engineering and Infocommunication Systems* ("*Radioinfocom-2024*"): Collection of scientific articles based on the materials of the 8th International Scientific and Practical Conference. Moscow: RTU MIREA; 2024. P. 381–384 (in Russ.). https://www.elibrary.ru/mwzoxe
7. Ivanova V.Yu., Solovyev D.O. Overeview of Big Data Processing Methods Using Apache Spark, Pandas Library, and SQL. *Naukosfera = Naukosphere*. 2024;5(1):43–47 (in Russ.). https://doi.org/10.5281/zenodo.11241367, https://www.elibrary.ru/uljwcm
8. Habib J.M.T., Poguda A.A. Comparison of Deep Learning Sentiment Analysis Methods, Including LSTM and Machine Learning. *Otkrytoe Obrazovanie = Open Education*. 2023;27(4):60–71 (in Russ.). https://doi.org/10.21686/1818-4243-2023-4-60-71
9. Kazantsev T. *Iskusstvennyi intellekt i mashinnoe obuchenie. Osnovy programmirovaniya na Python* (*Artificial Intelligence and Machine Learning. Fundamentals of Python Programming*). LitRes: Samizdat; 2020, 123 p. (In Russ.).
10. Bolshakov N.I., Sidorova E.V. Comparative Analysis of Machine Learning Methods for Problems of Data Classification. *Matematicheskie metody v tekhnologiyakh i tekhnike = Mathematical Methods in Technology and Engineering*. 2023;8: 66–71 (in Russ.). https://elibrary.ru/zdgmrk
11. Kouemou G., Opitz F. Impact of Wavelet-Based Signal Processing Methods on Radar Classification Systems Using Hidden Markov Models. In: *2008 International Radar Symposium*. Wroclaw, Poland. 2008. https://doi.org/10.1109/IRS.2008.4585763
12. Shevchenko A.S., Samarin V.A. *Neironnye seti* (*Neural Networks*): A Tutorial. Moscow: Ai Pi Ar Media; 2025, 181 p. (In Russ.).
13. Smirnov E.E., Kostyleva V.V., Murtazina A.R., Razin I.B. Comparison of Convolutional and Fully Connected Neural Networks in Relation to Image Recognition Tasks. *Izvestiya vysshikh uchebnykh zavedenii. Tekhnologiya tekstil'noi promyshlennosti = Textile Industry Technology. Series: Proceedings of Higher Educational Institutions*. 2023;5(407): 236–242 (in Russ.). https://elibrary.ru/gvpmpa
14. Paramonov A.A., Nguyen V.M., Nguyen M.T. Multi-task neural network for solving the problem of recognizing the type of QAM and PSK modulation under parametric a priori uncertainty. *Russian Technological Journal*. 2023;11(4):49–58. https://doi.org/10.32362/2500-316X-2023-11-4-49-58
15. Abdulkadirov R.I., Alikhanov A.A., Aidamirov N.O., Babayan R.A., Dadalyan A.A., Davydov N.M. Comparative analysis of optimization algorithms using various test functions. In: *High-Performance Computing for Solving Applied Problems*: Collection of materials of the 12th (69th) Annual Scientific and Practical Conference of Students, Teachers, and Young Scientists of the North Caucasus Federal University. Stavropol; 2025. P. 17–21 (in Russ.). https://elibrary.ru/dqdcfh
16. Gaunov S.R., Baimuradov U.G., Sitnikov S.Yu. Machine learning in Python: using Tensorflow and Scikit-Learn libraries. *Ehkonomika i upravlenie*: *problemy, resheniya = Economics and Management*: *Problems, Solutions*. 2024;8(12-153):72–81 (in Russ.). https://doi.org/10.36871/ek.up.p.r.2024.12.08.009

СПИСОК ЛИТЕРАТУРЫ

1. Терлецкий А.С., Терлецкая Е.С. *Нейронные сети и искусственный интеллект*: *Основы нейронных сетей на языке Python.* Липецк: Липецкий государственный педагогический университет имени П.П. Семенова-Тян-Шанского; 2023, 76 с. https://www.elibrary.ru/ugipee
2. Торгаев С.Н., Лежнина И.А., Шульга И.Д. *Практическое руководство по цифровой обработке сигналов*: *цифровые фильтры и обработка ЭКГ сигналов*. Томск: ООО «СТТ»; 2020, 112 с. https://www.elibrary.ru/rupfct

3. Царегородцев М.А. Многопотоковая численная реализация криптографических алгоритмов параллельного действия для защиты конфиденциальной информации в оборонно-промышленных комплексах при ее обработке, хранении и передачи в облачное хранилище больших данных. *Альманах Пермского военного института войск национальной гвардии.* 2023;4(12):104–112. https://www.elibrary.ru/vvtbii
4. Курушин А.А. *Гибридное моделирование в HFSS ANSYS:* учебное пособие. М.: СОЛОН-Пресс; 2023, 292 с.
5. Банков С.Е., Курушин А.А. *Расчет антенн и СВЧ структур с помощью HFSS Ansoft*. М.: ЗАО «НПП «РОДНИК»; 2009, 256 с.
6. Кожемяко В.А. Получение отклика сигнала от объекта в САПР Ansys. В сб.: *Актуальные проблемы и перспективы развития радиотехнических и инфокоммуникационных систем* («*Радиоинфоком-2024*»): Сборник научных статей по материалам VIII Международной научно-практической конференции. Москва, 18–22 ноября 2024 г. М.: РТУ МИРЭА; 2024. С. 381–384. https://www.elibrary.ru/mwzoxe
7. Иванова В.Ю., Соловьев Д.О. Обзор методов обработки больших данных с использованием Apache Spark, библиотеки Pandas и SQL. *Наукосфера*. 2024;5(1):43–47. https://doi.org/10.5281/zenodo.11241367, https://www.elibrary.ru/uljwcm
8. Хабиб Ж.М.Т., Погуда А.А. Сравнение методов анализа настроений глубокого обучения, включая LSTM и машинное обучение. *Открытое образование*. 2023;27(4):60–71. https://doi.org/10.21686/1818-4243-2023-4-60-71
9. Казанцев Т. *Искусственный интеллект и машинное обучение. Основы программирования на Python*. ЛитРес: Самиздат; 2020, 123 с.
10. Большаков Н.И., Сидорова Е.В. Сравнительный анализ методов машинного обучения для задач классификации данных. *Математические методы в технологиях и технике*. 2023;8:66–71. https://elibrary.ru/zdgmrk
11. Kouemou G., Opitz F. Impact of Wavelet-Based Signal Processing Methods on Radar Classification Systems Using Hidden Markov Models. In: *2008 International Radar Symposium*. Wroclaw, Poland. 2008. https://doi.org/10.1109/IRS.2008.4585763
12. Шевченко А.С., Самарин В.А. *Нейронные сети:* учебное пособие. М.: Ай Пи Ар Медиа; 2025, 181 с.
13. Смирнов Е.Е., Костылева В.В., Муртазина А.Р., Разин И.Б. Сравнение сверточных и полносвязных нейронных сетей применительно к задачам распознавания изображений. *Известия высших учебных заведений. Технология текстильной промышленности.* 2023;5(407):236–242. https://elibrary.ru/gvpmpa
14. Парамонов А.А., Нгуен В., Нгуен М. Многозадачная нейронная сеть в задаче распознавания вида QAM- и PSK-модуляции в условиях параметрической априорной неопределенности. *Russian Technological Journal*. 2023;11(4):49–58. https://doi.org/10.32362/2500-316X-2023-11-4-49-58
15. Абдулкадиров Р.И., Алиханов А.А., Айдамиров Н.О., Бабаян Р.А., Дадалян А.А., Давыдов Н.М. Сравнительный анализ алгоритмов оптимизации на примере различных тестовых функций. В сб.: *Высокопроизводительные вычисления для решения прикладных задач*: Сборник материалов XII (69-й) ежегодной научно-практической конференции студентов, преподавателей и молодых ученых Северо-Кавказского федерального университета, Ставрополь, 14–27 апреля 2025 года. Ставрополь; 2025. С. 17–21. https://elibrary.ru/dqdcfh
16. Гаунов С.Р., Баймурадов У.Г., Ситников С.Ю. Машинное обучение на Python: использование библиотек Tensorflow и Scikit-Learn. *Экономика и управление*: *проблемы, решения*. 2024;8(12-153):72–81. https://doi.org/10.36871/ek.up.p.r.2024.12.08.009

About the Authors

Vladislav A. Kozhemyako, Assistant, Department of Radio Wave Processes and Technologies, Institute of Radio Electronics and Informatics, MIREA – Russian Technological University (78, Vernadskogo pr., Moscow, 119454 Russia). E-mail: kozhemyako@mirea.ru. SPIN-code RSCI 4210-0353, http://orcid.org/0009-0009-8512-2288

Alexey D. Yarlykov, Cand. Sci. (Eng.), Senior Lecturer, Department of Radio Wave Processes and Technologies, Institute of Radio Electronics and Informatics, MIREA – Russian Technological University (78, Vernadskogo pr., Moscow, 119454 Russia). E-mail: yarlykov@mirea.ru. Scopus Author ID 57290652000, RSCI SPIN-code 3450-1587, https://orcid.org/0000-0002-7232-8588

Об авторах

Кожемяко Владислав Александрович, ассистент, кафедра радиоволновых процессов и технологий, Институт радиоэлектроники и информатики, ФГБОУ ВО «МИРЭА – Российский технологический университет» (119454, Россия, Москва, пр-т Вернадского, д. 78). E-mail: kozhemyako@mirea.ru. SPIN-код РИНЦ 4210-0353, http://orcid.org/0009-0009-8512-2288

Ярлыков Алексей Дмитриевич, к.т.н., старший преподаватель, кафедра радиоволновых процессов и технологий, Институт радиоэлектроники и информатики, ФГБОУ ВО «МИРЭА – Российский технологический университет» (119454, Россия, Москва, пр-т Вернадского, д. 78). E-mail: yarlykov@mirea.ru. Scopus Author ID 57290652000, SPIN-код РИНЦ 3450-1587, https://orcid.org/0000-0002-7232-8588

Translated from Russian into English by K. Nazarov
Edited for English language and spelling by Thomas A. Beavitt

Micro- and nanoelectronics. Condensed matter physics

Микро- и наноэлектроника. Физика конденсированного состояния

UDC 537.632
https://doi.org/10.32362/2500-316X-2026-14-3-72-82
EDN JXIQMS

RESEARCH ARTICLE

Application of the Berreman formalism for modeling magneto-optical Kerr effects in multilayered structures

Igor V. Gladyshev, Alexey N. Yurasov, Maxim M. Yashin @

MIREA – Russian Technological University, Moscow, 119454 Russia
@ *Corresponding author, e-mail: yashin@mirea.ru*

• **Submitted:** 22.04.2025 • **Revised:** 05.05.2025 • **Accepted:** 24.03.2026

Abstract

Objectives. Materials composed of numerous ultrathin layers, each having a thickness on the order of several nanometers, constitute an advanced class of composite structures exhibiting unique physical properties not typically found in conventional materials. These materials are of significant interest in both scientific and industrial sectors due to their adaptability and broad potential for application. Researchers are particularly intrigued by structures incorporating both magnetic and non-magnetic layers. The investigation of magneto-optical phenomena—particularly the Kerr effect—within these structures contributes to a deeper understanding of their physical characteristics, as well as enhancing prospects for their practical implementation. Since, to ensure the accurate interpretation of experimental data, it is imperative to consider potential interference effects, it becomes necessary to develop a mathematical model of the structure for comparing experimental findings with theoretical calculations. The purpose of this study is to analyze one of the modeling methods for multilayer structures in which magneto-optical Kerr effects can manifest themselves in individual or all layers.

Methods. The Berreman method, which is based on the matrix representation of Maxwell's differential equations, is used to model all three magneto-optical Kerr effects (polar, longitudinal, transverse) in multilayer thin-film structures.

Results. For optically isotropic materials, Berreman matrices have been derived for experimental configurations required to observe the transverse, polar, and longitudinal Kerr effects. A method is additionally proposed to account for the influence of thick layers within the investigated structure.

Conclusions. For the matrices presented in this paper, the Berreman method was used to analyze magneto-optical Kerr effects in an isotropic medium. As well as allowing us to obtain accurate formulas for magneto-optical effects, this provided more accurate modeling of complex multilayer structures, as well as contributing to an in-depth understanding of their physical characteristics, which provides new opportunities for analyzing and searching a wide range of materials.

Keywords: magneto-optical Kerr effects, Berreman method, Berreman matrix, dielectric constant tensor, multilayer structures

For citation: Gladyshev I.V., Yurasov A.N., Yashin M.M. Application of the Berreman formalism for modeling magneto-optical Kerr effects in multilayered structures. *Russian Technological Journal.* 2026;14(3):72–82. https://doi.org/10.32362/2500-316X-2026-14-3-72-82, https://www.elibrary.ru/JXIQMS

Financial disclosure: The authors have no financial or proprietary interest in any material or method mentioned.

The authors declare no conflicts of interest.

НАУЧНАЯ СТАТЬЯ

Применение метода Берремана при моделировании магнитооптических эффектов Керра в многослойных структурах

И.В. Гладышев, А.Н. Юрасов, М.М. Яшин [@]

МИРЭА – Российский технологический университет, Москва, 119454 Россия
[@] *Автор для переписки, e-mail: yashin@mirea.ru*

• **Поступила:** 22.04.2025 • **Доработана:** 05.05.2025 • **Принята к опубликованию:** 24.03.2026

Резюме

Цели. Материалы, состоящие из множества ультратонких слоев, каждый из которых имеет толщину порядка нескольких нанометров, являются перспективным классом композитных структур с уникальными физическими характеристиками, не присущими традиционным материалам. Они представляют значительный интерес в научной и промышленной сферах благодаря своей многофункциональности и широким возможностям применения. Особое внимание исследователей привлекают структуры, включающие как магнитные, так и немагнитные слои. Исследование магнитооптических явлений, в частности эффекта Керра, в данных структурах способствует углублению понимания их физических свойств и расширению возможностей их практического применения. Для корректной интерпретации экспериментальных данных необходимо учитывать возможные интерференционные эффекты. В связи с этим возникает потребность в разработке математической модели структуры и сопоставлении экспериментальных результатов с теоретическими расчетами. Целью настоящего исследования является анализ одного из методов моделирования многослойных структур, в которых всесторонне теоретически рассматриваются все три магнитооптических эффекта Керра (полярный, меридиональный, экваториальный) с получением универсальных формул.

Методы. Для моделирования всех трех магнитооптических эффектов Керра в многослойных тонкопленочных структурах применяется метод Берремана, основанный на матричном представлении дифференциальных уравнений Максвелла.

Результаты. Для оптически изотропных материалов получены матрицы Берремана, соответствующие экспериментальным геометриям, необходимым для наблюдения экваториального, полярного и меридионального эффектов Керра. Предложен метод учета толстых слоев в исследуемой структуре.

Выводы. Использование метода Берремана с применением матриц, представленных в данной работе, для анализа магнитооптических эффектов Керра в изотропной среде позволило получить точные формулы магнитооптических эффектов и обеспечило более точное моделирование сложных многослойных структур, а также способствует углубленному пониманию их физических характеристик, открывая возможности для анализа и поиска широкого спектра материалов.

Ключевые слова: магнитооптические эффекты Керра, метод Берремана, матрица Берремана, тензор диэлектрической проницаемости, многослойные структуры

Для цитирования: Гладышев И.В., Юрасов А.Н., Яшин М.М. Применение метода Берремана при моделировании магнитооптических эффектов Керра в многослойных структурах. *Russian Technological Journal*. 2026;14(3):72–82. https://doi.org/10.32362/2500-316X-2026-14-3-72-82, https://www.elibrary.ru/JXIQMS

Прозрачность финансовой деятельности: Авторы не имеют финансовой заинтересованности в представленных материалах или методах.

Авторы заявляют об отсутствии конфликта интересов.

INTRODUCTION

Michael Faraday's 1845 discovery of the interaction of light with magnetized matter subsequently led to the discovery of a multitude of linear and nonlinear, direct and inverse magneto-optical effects in the visible, infrared, ultraviolet, and X-ray regions of the spectrum. This field of physics is known as magneto-optics or magneto-photonics. Kerr effects, including magneto-optical reflection effects, provide the fundamental basis for magneto-optical Kerr spectroscopy method. This approach is used to analyze the magnetic state of local areas of a sample at the depth of formation of the magneto-optical signal. By studying the spectral dependencies of the magneto-optical Kerr effect at different wavelengths of radiation, it becomes possible to obtain information about magneto-optical transitions that reflect the electronic, crystalline, and magnetic structure of the local area of the sample under study [1–4]. Thus, magneto-optical spectroscopy is an important tool for studying micro- and nanostructures, including multilayer systems.

Multilayer magnetic structures are a subject of increased interest among researchers due to their significant potential for application in various fields of science and technology. The growing scientific interest in the study and development of these structures in recent years is accompanied by an increase in the number of scientific publications devoted to their research. Particular attention is paid to systems consisting of strongly magnetic (ferromagnets and ferrimagnets) and weakly magnetic materials [5–8].

However, modeling structures that include layers of materials with magneto-optical properties is a more complex task than calculating optical systems based on isotropic media. While, when obtaining analytical expressions for a certain number of magnetic layers, such as in [9], any increase in their number adds already represents a laborious task, with a significant number of magnetic layers, it becomes unfeasible. As a rule, the case of normal incidence of light on the structure is considered within the framework of the Jones matrix method.

BERREMAN METHOD

More than half a century ago, Berreman proposed a method for calculating the propagation of electromagnetic waves using 4 × 4 complex matrices based on the matrix representation of Maxwell's equations [10]. This highly accurate and versatile method is used to accounting for the peculiarities of light propagation in complex anisotropic structures, including multilayer systems. Despite its advantages, Berreman's method has not been widely used due to its high computational complexity and the difficulty of calculating matrices in general. Nevertheless, when solving problems related to optical anisotropy (uniaxial, biaxial) and the need to take into account magnetic anisotropy or optical activity of the medium, the advantages of the Berman method may turn out to be significant, thus justifying its application under the appropriate conditions.

When applying the Berman method, one of the most challenging problems arises when calculating the matrix exponent. However, a number of techniques have been developed to calculate this exponent with varying degrees of accuracy [11]. Moreover, if the accuracy is insufficient, the calculated layer can be divided into smaller sublayers. A fairly accurate and effective algorithm for calculating Berreman matrices for layers of considerable thickness can be found in the scientific literature [12]. In certain cases, such as a uniaxial medium, the Berreman matrix can be calculated analytically.

The present paper also provides accurate matrices for homogeneous media with induced optical activity for geometries corresponding to the transverse, polar, and longitudinal Kerr effects. The Berreman method is used to analyze the propagation of a plane monochromatic wave through a homogeneous medium. A one-dimensional inhomogeneous medium can be described by a system of plane-parallel layers, each of which can be considered homogeneous.

Let us consider the essence of the method using the coordinate system shown in Fig. 1. Let us also assume that there is air on both sides of the layer under investigation. Consequently, the medium from which light falls on the layer of material and into which it passes after passing through this layer is homogeneous, non-absorbing, and has a refractive index equal to one.

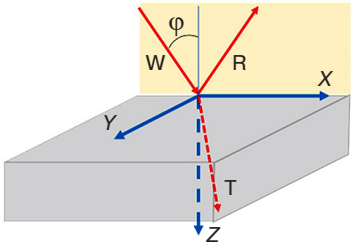


Fig. 1. Coordinate system used:
X, *Y*, *Z* are coordinate axes;
φ is the angle of incidence;
W is the incident (wave),
R is reflected, and
T is transmitted light beams

Since the wave is monochromatic, the time dependence of all components of the electric (E) and magnetic (H) fields has the following form: $e^{-i\omega t}$. Within the geometry under consideration, the projection of

the wave vector onto the x-axis, denoted as k_x, is the same for all waves and takes the following value:

$$k_x = \frac{\omega}{c}\sin\varphi, \quad (1)$$

where ω is the frequency, φ is the angle of incidence, and c is the speed of light in a vacuum.

Then Maxwell's equations can be written as:

$$\mathbf{R}\Psi = -i\omega\mathbf{M}\Psi, \quad (2)$$

where **R, M** are 6 × 6 block matrices, while Ψ is a column matrix containing the following elements:

$$\mathbf{R} = \begin{pmatrix} \mathbf{O} & \mathbf{rot} \\ -\mathbf{rot} & \mathbf{O} \end{pmatrix}, \mathbf{M} = \begin{pmatrix} \hat{\varepsilon} & \mathbf{O} \\ \mathbf{O} & \bar{\mu} \end{pmatrix}, \Psi = \begin{pmatrix} E_x \\ E_y \\ E_z \\ H_x \\ H_y \\ H_z \end{pmatrix}.$$

Here **O** is a zero matrix of dimension 3 × 3 and, for our case,

$$\mathbf{rot} = \begin{pmatrix} 0 & -\frac{\partial}{\partial z} & 0 \\ \frac{\partial}{\partial z} & 0 & -ik_x \\ 0 & ik_x & 0 \end{pmatrix},$$

$$\hat{\varepsilon} = \varepsilon_0 \begin{pmatrix} \varepsilon_{11} & \varepsilon_{12} & \varepsilon_{13} \\ \varepsilon_{21} & \varepsilon_{22} & \varepsilon_{23} \\ \varepsilon_{31} & \varepsilon_{32} & \varepsilon_{33} \end{pmatrix},$$

$$\hat{\mu} = \mu_0 \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

ε_0 and μ_0 are the electric and magnetic constants, respectively.

In the process of solving Eq. (2), a system is formed that includes two linear homogeneous algebraic equations and four differential equations. The algebraic equations are solved with respect to the field components E_z and H_z, after which the resulting expressions are substituted into the differential equations. The result is a system of four linear homogeneous differential equations of the first order containing the unknown field components E_x, E_y, H_x, and H_y:

$$\frac{\partial}{\partial z}\xi = i\omega\Delta\xi, \quad (3)$$

where ξ is the matrix-column of the form:

$$\xi = \begin{pmatrix} E_x \\ H_y \\ E_y \\ -H_x \end{pmatrix},$$

and Δ is the differential propagation matrix for a given medium, with dimensions 4 × 4.

As mentioned earlier, the medium is divided into layers within which the components of the matrix Δ do not depend on the coordinate z. In each layer of thickness h the solution of the corresponding homogeneous first-order differential Eq. (3) has the form:

$$\xi(z + h) = e^{i\omega h\Delta}\,\xi(z) \equiv \mathbf{P}(h)\xi(z).$$

Thus, the **P** matrix, representing the Berman matrix of this layer, appears as follows:

$$\mathbf{P}(h) = e^{i\omega h\Delta}. \quad (4)$$

When the optical parameters of the medium depend on z, for example, for a structure consisting of homogeneous layers within which the optical parameters are considered constant, integrating Eq. (3) reduces to multiplying the corresponding matrices for individual layers:

$$\mathbf{P} = \prod_{j=1}^{n} \mathbf{P}_j.$$

The matrix of the next layer is multiplied by the previous ones on the left:

$$\xi(h) = \{\mathbf{P}_n \cdot \mathbf{P}_{n-1} \cdot \mathbf{P}_{n-2} \cdot \ldots \cdot \mathbf{P}_3 \cdot \mathbf{P}_2 \cdot \mathbf{P}_1\}\xi(0).$$

The electromagnetic field on one side of the structure is determined by the superposition of incident and reflected waves, while on the other side there is only the transmitted wave. Then:

$$\xi_T = \mathbf{P}(\xi_W + \xi_R). \quad (5)$$

According to the designations shown in Fig. 1, the indices W, R, and T denote incident, reflected, and transmitted waves, respectively. Multiplying both sides of Eq. (5) on the left by the matrix $\mathbf{F} = \mathbf{P}^{-1}$, representing the inverse of the Berman matrix of the layer, we obtain a system of linear equations that can be solved for the components of the reflected and transmitted waves. To identify the belonging to the corresponding wave, we assign upper indices to the field components corresponding to the designations in Fig. 1. The lower indices of the field components indicate the direction of the projection of the component under consideration onto the corresponding axis. Let us list them so that readers can apply them in their work.

$$E_y^{\mathrm{T}} = \frac{2\left(\chi_1 E_y^{\mathrm{W}} - \alpha\chi_2 E_x^{\mathrm{W}}\right)}{d - ag},\ E_x^{\mathrm{T}} = \frac{2\chi_2 E_x^{\mathrm{W}} - gE_y^{\mathrm{T}}}{b}, \quad (6)$$
$$E_y^{\mathrm{R}} = \beta_3 E_x^{\mathrm{T}} + \gamma_3 E_y^{\mathrm{T}} - E_y^{\mathrm{W}},\ E_x^{\mathrm{R}} = \beta_1 E_x^{\mathrm{T}} + \gamma_1 E_y^{\mathrm{T}} - E_x^{\mathrm{W}},$$

where

$$a = \frac{\beta_4 + \chi_1\beta_3}{\beta_2 + \chi_2\beta_1},\ b = \beta_2 + \chi_2\beta_1,\ g = \gamma_2 + \chi_2\gamma_1,\ d = \gamma_4 + \chi_1\gamma_3,$$
$$\beta_1 = f_{11} + \chi_2 f_{12},\ \gamma_1 = f_{13} + \chi_1 f_{14},$$
$$\beta_2 = f_{21} + \chi_2 f_{22},\ \gamma_2 = f_{23} + \chi_1 f_{24},$$
$$\beta_3 = f_{31} + \chi_2 f_{32},\ \gamma_3 = f_{33} + \chi_1 f_{34},$$
$$\beta_4 = f_{41} + \chi_2 f_{42},\ \gamma_4 = f_{43} + \chi_1 f_{44},$$
$$\chi_1 = \sqrt{\frac{\varepsilon_0}{\mu_0}}\cos\varphi,\ \chi_2 = \sqrt{\frac{\varepsilon_0}{\mu_0}}\cdot\frac{1}{\cos\varphi},$$

$f_{11}, f_{12}, \ldots, f_{44}$ are elements of the matrix $\mathbf{F} = \mathbf{P}^{-1}$, the inverse of the Berreman matrix.

Knowing the values of the light wave field components, it is possible to calculate the polarization rotation, as well as the reflection coefficient K_{R} and transmission coefficient K_{T}:

$$K_{\mathrm{R}} = \frac{|E_x^{\mathrm{R}}/\cos\varphi|^2 + |E_y^{\mathrm{R}}|^2}{|E_x^{\mathrm{W}}/\cos\varphi|^2 + |E_y^{\mathrm{W}}|^2},$$
$$K_{\mathrm{T}} = \frac{|E_x^{\mathrm{T}}/\cos\varphi|^2 + |E_y^{\mathrm{T}}|^2}{|E_x^{\mathrm{W}}/\cos\varphi|^2 + |E_y^{\mathrm{W}}|^2}, \quad (7)$$

but first it is necessary to determine the Berreman matrices for the modeled structure and its individual layers.

BERREMAN MATRICES FOR MODELING MAGNETO-OPTICAL KERR EFFECTS

Magneto-optical phenomena manifest themselves in changes in the optical properties of a film (structure) depending on the presence or absence of a magnetic field. In this regard, their modeling requires determining the Berreman matrices for both scenarios. In the cases considered here, they can all be obtained analytically.

The method for finding the matrix for an isotropic layer in the absence of a magnetic field was demonstrated by Berreman in one of his early works [10]. Substituting the dielectric permeability tensor in the form:

$$\hat{\varepsilon} = \varepsilon_0\varepsilon\begin{pmatrix}1 & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & 1\end{pmatrix},$$

we obtain a differential, 4 × 4 dimensional, propagation matrix Δ_0 for a given medium:

$$\Delta_0 = \begin{pmatrix}0 & U_a & 0 & 0\\ U_b & 0 & 0 & 0\\ 0 & 0 & 0 & U_s\\ 0 & 0 & U_d & 0\end{pmatrix}, \quad (8)$$

where

$$U_a = \frac{\mu_0}{\varepsilon}(\varepsilon - \sin^2\varphi),\ U_b = \varepsilon_0\varepsilon,\ U_s = \mu_0,\ U_d = \varepsilon_0(\varepsilon - \sin^2\varphi). \quad (9)$$

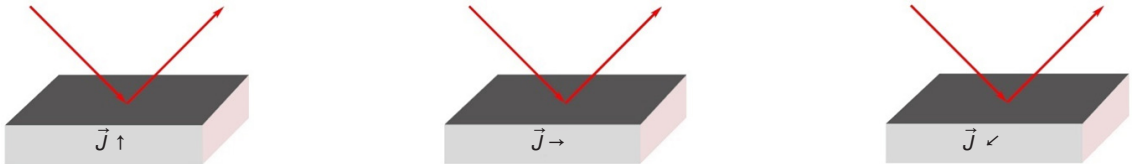
According to Eq. (4), to obtain the Berman matrix, it is necessary to take the exponential of matrix (8) multiplied by $i\omega h$. The simplest and most intuitive way to do this is to expand the exponential into a Taylor series. Summing the terms of the series, we can see that the elements of the resulting matrix also represent Taylor series expansions of certain functions. As a result, for the Berman matrix of an isotropic homogeneous medium $\mathbf{P_0}$, we can write the following:

$$\mathbf{P_0}(h) = \begin{pmatrix}\cos(\omega h\chi_0) & i\beta_0\sin(\omega h\chi_0) & 0 & 0\\ i\beta_0^{-1}\sin(\omega h\chi_0) & \cos(\omega h\chi_0) & 0 & 0\\ 0 & 0 & \cos(\omega h\chi_0) & i\delta_0^{-1}\sin(\omega h\chi_0)\\ 0 & 0 & i\delta_0\sin(\omega h\chi_0) & \cos(\omega h\chi_0)\end{pmatrix}. \quad (10)$$

Here

$$\chi_0 = \frac{1}{c}\sqrt{\varepsilon - \sin^2\varphi},\ \beta_0 = \sqrt{\frac{\mu_0}{\varepsilon_0}}\cdot\frac{\sqrt{\varepsilon - \sin^2\varphi}}{\varepsilon},\ \delta_0 = \sqrt{\frac{\varepsilon_0}{\mu_0}}\sqrt{\varepsilon - \sin^2\varphi}. \quad (11)$$

Depending on the configuration of the magnetic field and the incidence of light on the film or structure, three magneto-optical Kerr effects are distinguished: polar, longitudinal, and transverse. The polar Kerr effect occurs when the magnetization vector J, created in particular by an external magnetic field, is oriented perpendicular to the plane of the film and parallel to the plane of incidence of light (Fig. 2a). The longitudinal Kerr effect occurs when the magnetization vector is oriented parallel to the film surface and lies in the plane of light incidence (Fig. 2b). Conversely, the transverse Kerr effect is observed when the magnetization vector is oriented perpendicular to the plane of incidence of light and parallel to the film structure (Fig. 2c).



(a) (b) (c)

Fig. 2. Geometry of observation of magneto-optical Kerr effects: (a) polar, (b) longitudinal, (c) transverse. The directions of the coordinate axes are shown in Fig. 1

It should be noted that the *s*- and *p*-components of the electric field of a light wave in this geometry are related to the projections on the axes as follows: $E_s = E_y$, $E_p = E_x/\cos\varphi$. Knowing E_s and E_p, we can calculate the angle of rotation of the polarization plane for a plane-polarized wave $\theta_0 = \text{arctg}(|E_s|/|E_p|)$ or, if there is a non-zero phase difference ϕ between E_s and E_p, the azimuth angle θ_ϕ, which forms the main axis of the polarization ellipse with the plane of incidence $\theta_\phi = \text{arctg}(|E_s|/|E_p|)\cos\phi$. The ellipticity e can be calculated using the formula: $e = \text{tg}[0.5\arcsin(-2\text{Im}(\Xi)/(1-|\Xi|^2))]$, where $\Xi = (|E_s|/|E_p|)e^{i\phi}$ [13, 14]. These are precisely the values measured in the longitudinal and polar Kerr effects, while in the transverse effect, the main factor is the change in the intensity of the reflected electromagnetic wave.

Let us begin our consideration with the transverse Kerr effect (Fig. 2c). In this case, the dielectric permeability tensor of the medium in the chosen coordinate system can be written as:

$$\hat{\varepsilon} = \varepsilon_0\varepsilon\begin{pmatrix} 1 & 0 & iQ \\ 0 & 1 & 0 \\ -iQ & 0 & 1 \end{pmatrix},$$

where Q is the magneto-optical parameter.

In this case, solving Eq. (2) yields the following propagation matrix Δ_T (T is transverse Kerr effect):

$$\Delta_T = \begin{pmatrix} i\eta & U_a & 0 & 0 \\ U_b^* & -i\eta & 0 & 0 \\ 0 & 0 & 0 & U_s \\ 0 & 0 & U_d & 0 \end{pmatrix}. \qquad (12)$$

Here

$$U_b^* = U_b(1-Q^2),\ \eta = \frac{\sin\varphi}{c}Q.$$

The matrix (12) can be represented as the sum of matrices:

$$\Delta_T = \Delta_T^J + \Delta_T^* = \begin{pmatrix} i\varsigma & 0 & 0 & 0 \\ \vartheta & \theta i & 0 & \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} + \begin{pmatrix} 0 & U_a & 0 & 0 \\ U_b^* & 0 & 0 & 0 \\ 0 & 0 & 0 & U_s \\ 0 & 0 & U_d & 0 \end{pmatrix}.$$

Since the structure of the matrix Δ_T^* is identical to the matrix structure Δ_0, the matrix Δ_T^* will have a form similar to (10). We will also find the exponent of the matrix Δ_T^J, multiplied by $i\omega h$, using Taylor series expansion. Fortunately, in this case, the elements of the resulting matrix also represent series expansions. As a result, we obtain:

$$e^{i\omega h\Delta_T^J} = \begin{pmatrix} e^{-\omega h\eta} & 0 & 0 & 0 \\ 0 & e^{\omega h\eta} & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

Finally, the Berreman matrix in this case will have the form:

$$\mathbf{P}_{\mathrm{T}}(h) = \mathrm{e}^{i\omega h\Delta_{\mathrm{T}}^{J}}\mathrm{e}^{i\omega h\Delta_{\mathrm{T}}^{*}}, \tag{13}$$

$$\mathbf{P}_{\mathrm{T}}(h) = \begin{pmatrix} \mathrm{e}^{-\omega h\eta}\cos(\omega h\chi_{\mathrm{T}}) & i\,\mathrm{e}^{-\omega h\eta}\beta_{\mathrm{T}}\sin(\omega h\chi_{\mathrm{T}}) & 0 & 0 \\ i\,\mathrm{e}^{\omega h\eta}\beta_{\mathrm{T}}^{-1}\sin(\omega h\chi_{\mathrm{T}}) & \mathrm{e}^{\omega h\eta}\cos(\omega h\chi_{\mathrm{T}}) & 0 & 0 \\ 0 & 0 & \cos(\omega h\chi_0) & i\delta_0^{-1}\sin(\omega h\chi_0) \\ 0 & 0 & i\delta_0\sin(\omega h\chi_0) & \cos(\omega h\chi_0) \end{pmatrix}, \tag{14}$$

where

$$\chi_{\mathrm{T}} = \chi_0\sqrt{1-Q^2},\ \beta_{\mathrm{T}} = \beta_0\Big/\sqrt{1-Q^2}.$$

In the process of calculating matrix (14), a transition was made from the operation of adding matrices to the operation of multiplying their exponents, which are also matrices. It should be noted that, unlike addition, matrix multiplication is generally not commutative. If the reverse order of multiplication is used, matrix (14) will have a different form: the changes will affect elements P_{12} and P_{21}, resulting in the exponents included in them being swapped.

The multiplication order shown in (13) was chosen for the following reasons. First, based on general considerations, modifications caused by the magneto-optical effect manifest themselves after light radiation passes through the film (structure), rather than before it interacts with it. Second, the results of calculations [15] of the transverse Kerr effect for a cobalt film on a silicon substrate (Fig. 3) demonstrate that using the multiplication order (13) yields a result that is consistent with the experiments in terms of the nature of the spectral dependence and the order of magnitude. At the same time, the use of the reverse multiplication order leads to simulation results that do not fully correspond to the experimental data.

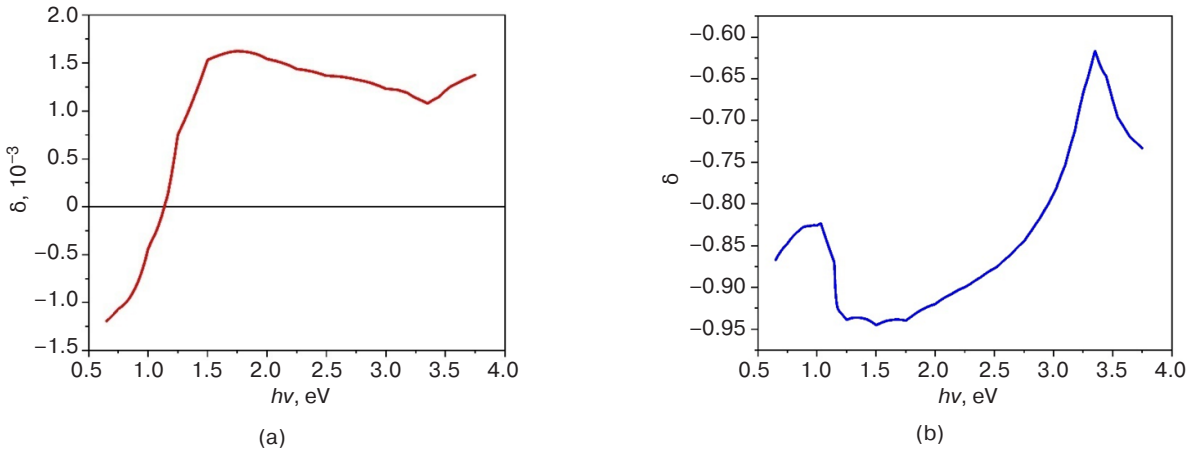


Fig. 3. Results of modeling the spectral dependence of the transverse Kerr effect (δ) of a cobalt film on a silicon substrate, taking into account the order of matrix multiplication: (a) presented in (13), (b) reverse. *hv* is photon energy

For the case of the polar Kerr effect (Fig. 2a), the permittivity tensor in the coordinates used can be represented as

$$\hat{\varepsilon} = \varepsilon_0\varepsilon\begin{pmatrix} 1 & iQ & 0 \\ -iQ & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

From (2) we obtain:

$$\Delta_{\mathrm{P}} = \begin{pmatrix} 0 & U_a & 0 & 0 \\ U_b & 0 & i\zeta & 0 \\ 0 & 0 & 0 & U_s \\ -i\zeta & 0 & U_d & 0 \end{pmatrix}, \tag{15}$$

where $\zeta = \varepsilon_0\varepsilon Q$.

Here we can also consider the sum of the matrix (7) and

$$\Delta_{\mathrm{P}}^{J}=\begin{pmatrix}0&0&0&0\\0&0&i\zeta&0\\0&0&0&0\\-i\zeta&0&0&0\end{pmatrix}. \tag{16}$$

Due to the fact that the second degree of the matrix (16) is already equal to zero,

$$e^{i\omega h\Delta_{\mathrm{P}}^{J}}=\begin{pmatrix}1&0&0&0\\0&1&-\omega h\zeta&0\\0&0&1&0\\\omega h\zeta&0&0&1\end{pmatrix}.$$

The order of matrix multiplication is considered when deriving the expression for the transverse effect. As a result, the Berreman matrix for the polar effect takes the form:

$$\mathbf{P}_{\mathrm{P}}(h)=\begin{pmatrix}\cos(\omega h\chi_0)&i\beta_0\sin(\omega h\chi_0)&0&0\\i\beta_0^{-1}\sin(\omega h\chi_0)&\cos(\omega h\chi_0)&-\omega h\zeta\cos(\omega h\chi_0)&-i\omega h\zeta\delta_0^{-1}\sin(\omega h\chi_0)\\0&0&\cos(\omega h\chi_0)&i\delta_0^{-1}\sin(\omega h\chi_0)\\\omega h\zeta\cos(\omega h\chi_0)&i\omega h\zeta\beta_0\sin(\omega h\chi_0)&i\delta_0\sin(\omega h\chi_0)&\cos(\omega h\chi_0)\end{pmatrix}. \tag{17}$$

Using similar reasoning, we can derive the Berreman matrix for the case of the longitudinal Kerr effect (Fig. 2b). In this case:

$$\hat{\varepsilon}=\varepsilon_0\varepsilon\begin{pmatrix}1&0&0\\0&1&iQ\\0&-iQ&1\end{pmatrix}$$

and

$$\mathbf{P}_{\mathrm{M}}(h)=\begin{pmatrix}\cos(\omega h\chi_0)&i\beta_0\sin(\omega h\chi_0)&-\rho\cos(\omega h\chi_{\mathrm{M}})&-i\rho\delta_{\mathrm{M}}^{-1}\sin(\omega h\chi_{\mathrm{M}})\\i\beta_0^{-1}\sin(\omega h\chi_0)&\cos(\omega h\chi_0)&0&0\\0&0&\cos(\omega h\chi_{\mathrm{M}})&i\delta_{\mathrm{M}}^{-1}\sin(\omega h\chi_{\mathrm{M}})\\i\rho\beta_0^{-1}\sin(\omega h\chi_0)&\rho\cos(\omega h\chi_0)&i\delta_{\mathrm{M}}\sin(\omega h\chi_{\mathrm{M}})&\cos(\omega h\chi_{\mathrm{M}})\end{pmatrix}. \tag{18}$$

Here

$$\rho=\frac{\omega}{c}Q\sin\varphi,\ \chi_{\mathrm{M}}=\frac{1}{c}\sqrt{\varepsilon(1+Q^2)-\sin^2\varphi},\ \delta_{\mathrm{M}}=\sqrt{\frac{\varepsilon_0}{\mu_0}}\sqrt{\varepsilon(1+Q^2)-\sin^2\varphi}.$$

THICK LAYERS CONSIDERATION

The Berreman method takes into account interference effects. However, the simulated structures may contain fairly thick layers, the thickness of which exceeds the coherence length l_{coh}. of the light source. While such layers typically represent substrates onto which films and complex structures are deposited, the structures under study may also include a certain number of thick layers. The principle of taking such layers into account is based on the averaging method, in which deviations from coherence are considered as random variables obeying a normal distribution [16]. Although this approach requires additional computing power, it enables the inclusion of layers of intermediate thickness, in which a partial violation of the coherence condition is observed. In the elements of the Berreman matrix of a thick layer, the frequency ω is replaced by the sum (ω + w), where w is a random variable distributed according to the normal law. All frequency-dependent characteristics of the medium remain unchanged. Thus, the introduction of the variable w affects only the phase change. The mathematical expectation M_w = 0. The standard deviation σ_w depends on the layer thickness and is taken from the condition that with a layer thickness equal to the coherence length of the radiation, a phase shift π will occur:

$$\sigma_w=\frac{h}{l_{\mathrm{coh}}}\pi.$$

The magnitude of the calculated effect for the entire structure is averaged. Consider the case where a thin-film structure is deposited on a thick substrate with a thickness h_{sub}. Then, the reflectance from such a structure is calculated as:

$$\langle R \rangle = \frac{1}{\sqrt{2\pi^3}} \cdot \frac{l_{coh}}{h_{sub}} \int_{-\infty}^{\infty} R'(w) e^{-\frac{l_{coh}^2}{2\pi 2 h_{sub}^2} w^2} dw.$$

The values of $R'(w)$ are determined in accordance with Eq. (7) for the radiation frequency ω_0. In this case, in the matrices of all layers, except for the layer being averaged, $\omega = \omega_0$ is taken. In the averaged layer $\omega = \omega_0 + w$ is taken. All material characteristics for all layers are taken for the frequency ω_0. If there are several thick layers, multiple integration is performed.

Another problem with thick layers is that, with high absorption, an overflow situation may occur during intermediate calculations for the matrix or its inverse. To address this issue, it is proposed to introduce a certain error into the calculations. Accordingly, the maximum allowable thickness of the absorbing layers will be determined based on the condition that, when passing through these layers, the light wave amplitude decreases by 5–7 orders of magnitude. This will avoid data overflow, and the introduced error will be clearly smaller than the experimental errors.

CONCLUSIONS

This study proposes using the Berreman method to model all three magneto-optical Kerr effects: polar, longitudinal, and transverse. The Berreman matrix method presented in this paper enables a systematic analysis of magneto-optical Kerr effects in an isotropic medium using the obtained precise formulas for magneto-optical effects. As well as providing a sound basis for more accurate modeling of complex multilayer structures, this facilitates a deeper understanding of their physical characteristics, opening up possibilities for exploring a wide range of promising magnetic materials. The approaches for accounting for the presence of films in the modeled structure whose thickness exceeds the coherence length of the light source in this paper expand the range of micro- and nanostructures under consideration.

ACKNOWLEDGMENTS

The work was supported by the Ministry of Science and Higher Education of the Russian Federation (State Assignment for Universities No. FGFZ-2023-0005).

Authors' contributions

I.V. Gladyshev—methodology development, modeling, discussion of the results, drafting an article and its editing.

A.N. Yurasov, M.M. Yashin—discussion of the results, preparation of the article and its editing.

REFERENCES

1. Rylkov V.V., Emelyanov A.V., Nikolaev S.N., et al. Transport Properties of Magnetic Nanogranular Composites with Dispersed Ions in an Insulating Matrix. *J. Exp. Theor. Phys.* 2020;131(1):160–176. https://doi.org/10.1134/S1063776120070109
[Original Russian Text: Rylkov V.V., Emelianov A.V., Nikolaev S.N., Nikiruy K.E., Sitnikov A.V., Fadeev E.A., Demin V.A., Granovskii A.B. Transport Properties of Magnetic Nanogranular Composites with Dispersed Ions in an Insulating Matrix. *Zhurnal Eksperimental'noi i Teorieticheskoi Fiziki* (*ZhETF*). 2020;158(1):164–183 (in Russ.). https://doi.org/10.31857/S0044451020070159]
2. Martyshov M.N., Emelyanov A.V., Demin V.A., et al. Multifilamentary Character of Anticorrelated Capacitive and Resistive Switching in Memristive Structures Based on $(Co{-}Fe{-}B)_x(LiNbO_3)_{100-x}$ Nanocomposite. *Phys. Rev. Appl*. 2020;14:034016. https://doi.org/10.1103/PhysRevApplied.14.034016
3. Gao C.N., Yang Y.X., Xiong Y.Q., et al. Low critical current density for spin-transfer torque in Fe-MgO granular film at room temperature. *J. Phys. D*. 2014;47(4):045003. https://doi.org/10.1088/0022-3727/47/4/045003
4. Gan'shina E.A., Garshin V.V., Perova N.N., et al. Magnetooptical Kerr spectroscopy of nanocomposites. *J. Exp. Theor. Phys*. 2023;137(4):572–581. https://doi.org/10.1134/S1063776123100151
[Original Russian Text: Gan'shina E.A., Garshin V.V., Perova N.N, Pripechenkov I.M., Yurasov A.N., Yashin M.M., Ryl'kov V.V., Granovskii A.B. Magnetooptical Kerr spectroscopy of nanocomposites. *Zhurnal Eksperimental'noi i Teorieticheskoi Fiziki* (*ZhETF*). 2023;164 (4):662–672 (in Russ.).]
5. Pavlov V.V., Usachev P.A., Nefedov S.G., et al. Optical and magneto-optical properties of multilayer nanosized $[Co/TiO_2]_n$ films. *Phys. Solid State.* 2018;60(11):2244–2253. https://doi.org/10.1134/S1063783418110239
[Original Russian Text: Pavlov V.V., Usachev P.A., Nefedov S.G., Stognii A.I., Novitskii N.N., Pisarev R.V. Optical and magneto-optical properties of multilayer nanosized $[Co/TiO_2]_n$ films. *Fizika tverdogo tela* (*FTT*). 2018;60(11):2203–2212 (in Russ.). https://doi.org/10.21883/FTT.2018.11.46664.26NN]

6. Gladyshev I.V., Ganshina E.A., Simdyanova M.A., et al. Optical and Magneto-Optical Properties of Multilayer Magnetic Structures Based on Permalloy. *Bull. Russ. Acad. Sci. Phys.* 2024;88(1):S76–S79. https://doi.org/10.1134/S1062873824708821
7. Yashin M.M., Ryabukhin V.E., Yurasov A.N. Magneto-optical transverse Kerr effect in $Co_x(CoO)_{1-x}$ nanocomposites. *Russian Technological Journal.* 2025;13(1):115–121. https://doi.org/10.32362/2500-316X-2025-13-1-115-121
8. Makarov D.G., Danilov V.V., Kovalenko V.F. Multilayer structures with light transmission controlled by a magnetic field. *Zhurnal Tekhnicheskoi Fiziki*. 2004;74(5):77–82 (in Russ.).
9. Maevskii V.M. Theory of magneto-optical effects in multilayer systems with arbitrary orientation of magnetization. *Fizika metallov i metallovedenie = Physics of Metals and Metallography*. 1985;59(2):213–219 (in Russ.).
10. Berreman D.W. Optics in Stratified and Anisotropic Media: 4x4-Matrix Formulation. *J. Opt. Soc. Am*. 1972;62(4):502–510. https://doi.org/10.1364/JOSA.62.000502
11. Moler C., Van Loan C. Nineteen Dubious Ways to Compute the Exponential of a Matrix, Twenty-Five Years Later. *SIAM Review*. 2003;45(1):3–49. https://doi.org/10.1137/s00361445024180
12. Palto S.P. An algorithm for solving the optical problem for stratified anisotropic media. *J. Exp. Theor. Phys.* 2001;92(4): 552–560. https://doi.org/10.1134/1.1371338
[Original Russian Text: Palto S.P. An algorithm for solving the optical problem for stratified anisotropic media. *Zhurnal Eksperimental'noi i Teorieticheskoi Fiziki* (*ZhETF*). 2001;119(4):638–648 (in Russ.). https://jetp.ras.ru/cgi-bin/dn/r_119_0638.pdf]
13. Nasedkina Yu.F., Sementsov D.I. Polarization distribution in a Gaussian beam reflected from a resonant medium. *Tech. Phys. Lett*. 2006;32(4):323–326. https://doi.org/10.1134/S1063785006040158
[Original Russian Text: Nasedkina Yu.F., Sementsov D.I. Polarization distribution in a Gaussian beam reflected from a resonant medium. *Pis'ma v Zhurnal Tekhnicheskoĭ Fiziki* (*Pis'ma v ZhTF*). 2006;32(8):1–9 (in Russ.).]
14. Vinogradova M.B., Rudenko O.V., Suhorukov A.P. *Teoriya voln. Lineinye i nelineinye volny.* (*Theory of Waves. Linear and Nonlinear Waves*). Moscow: Lenand; 2019, 448 p. (In Russ.).
15. Gladyshev I.V. *Program for Calculating the Magnitude of the Equatorial Kerr Effect of a Multilayer Thin-Film Structure:* Certificate of state registration of computer program. 2025617039 RF. Publ. (registered) 21.03.2025 (in Russ.).
16. Gladyshev I.V. Reflection of light from multilayer structures, including both coherent and incoherent layers. In: *Optical Technologies, Materials and Systems* ("*Optotech 2024*"): Conference Proceedings. Moscow. 2024. P. 520–525 (in Russ.).

СПИСОК ЛИТЕРАТУРЫ

1. Рыльков В.В., Емельянов А.В., Николаев С.Н., Никируй К.Э., Ситников А.В., Фадеев Е.А., Демин В.А., Грановский А.Б. Транспортные свойства магнитных наногранулированных композитов с диспергированными ионами в изолирующей матрице. *Журнал экспериментальной и теоретической физики* (*ЖЭТФ*). 2020;158(1): 164–183. https://doi.org/10.31857/S0044451020070159
2. Martyshov M.N., Emelyanov A.V., Demin V.A., et al. Multifilamentary Character of Anticorrelated Capacitive and Resistive Switching in Memristive Structures Based on $(Co{-}Fe{-}B)_x(LiNbO_3)_{100-x}$ Nanocomposite. *Phys. Rev. Appl*. 2020;14:034016. https://doi.org/10.1103/PhysRevApplied.14.034016
3. Gao C.N., Yang Y.X., Xiong Y.Q., et al. Low critical current density for spin-transfer torque in Fe-MgO granular film at room temperature. *J. Phys. D*. 2014;47(4):045003. https://doi.org/10.1088/0022-3727/47/4/045003
4. Ганьшина Е.А., Гаршин В.В., Перова Н.Н., Припеченков И.М., Юрасов А.Н., Яшин М.М., Рыльков В.В., Грановский А.Б. Магнитооптическая Керр-спектроскопия нанокомпозитов. *Журнал экспериментальной и теоретической физики* (*ЖЭТФ*). 2023;164(4):662–672.
5. Павлов В.В., Усачев П.А., Нефедов С.Г., Стогний А.И., Новицкий Н.Н., Писарев Р.В. Оптические и магнитооптические свойства многослойных наноразмерных пленок $[Co/TiO_2]_n$. *Физика твердого тела* (*ФТТ*). 2018;60(11): 2203–2212. https://doi.org/10.21883/FTT.2018.11.46664.26NN
6. Gladyshev I.V., Ganshina E.A., Simdyanova M.A., et al. Optical and Magneto-Optical Properties of Multilayer Magnetic Structures Based on Permalloy. *Bull. Russ. Acad. Sci. Phys.* 2024;88(1):S76–S79. https://doi.org/10.1134/S1062873824708821
7. Яшин М.М., Рябухин В.Е., Юрасов А.Н. Магнитооптический экваториальный эффект Керра в нанокомпозитах $Co_x(CoO)_{1-x}$. *Russian Technological Journal*. 2025;13(1):115–121. https://doi.org/10.32362/2500-316X-2025-13-1-115-121
8. Макаров Д.Г., Данилов В.В., Коваленко В.Ф. Многослойные структуры с управляемым магнитным полем пропусканием света. *Журнал технической физики* (*ЖТФ*). 2004;74(5):77–82.
9. Маевский В.М. Теория магнетооптических эффектов в многослойных системах с произвольной ориентацией намагниченности. *Физика металлов и металловедение*. 1985;50(2):213–219.
10. Berreman D.W. Optics in Stratified and Anisotropic Media: 4x4-Matrix Formulation. *J. Opt. Soc. Am*. 1972;62(4):502–510. https://doi.org/10.1364/JOSA.62.000502
11. Moler C., Van Loan C. Nineteen Dubious Ways to Compute the Exponential of a Matrix, Twenty-Five Years Later. *SIAM Review*. 2003;45(1):3–49. https://doi.org/10.1137/s00361445024180
12. Палто С.П. Алгоритм решения оптической задачи для слоистых анизотропных сред. *Журнал экспериментальной и теоретической физики* (*ЖЭТФ*). 2001;119(4):638–648. https://jetp.ras.ru/cgi-bin/dn/r_119_0638.pdf
13. Наседкина Ю.Ф., Семенцов Д.И. Распределение поляризации в гауссовом пучке, отраженном от резонансной среды. *Письма в Журнал технической физики*. 2006;32(8):1–9.

14. Виноградова М.Б., Руденко О.В., Сухоруков А.П. *Теория волн. Линейные и нелинейные волны.* М.: Ленанд; 2019, 448 с.
15. Гладышев И.В. *Программа расчета величины экваториального эффекта Керра многослойной тонкопленочной структуры:* Свидетельство о государственной регистрации программы для ЭВМ. 2025617039 РФ. Заявка № 2025615502; заявл. 13.03.2025: опубл. (зарег.) 21.03.2025.
16. Гладышев И.В. Отражение света от многослойных структур, включающих как когерентные, так и некогерентные слои. В сб.: *Оптические технологии, материалы и системы* («*Оптотех 2024*»): сборник докладов конференции. Москва. 2024. С. 520–525. https://www.elibrary.ru/otpmij

About the Authors

Igor V. Gladyshev, Cand. Sci. (Phys.–Math.), Associate Professor, Department of Nanoelectronics, Institute for Advanced Technologies and Industrial Programming, MIREA – Russian Technological University (78, Vernadskogo pr., Moscow, 119454 Russia). E-mail: i_gladyshev@mirea.ru. ResearcherID N-1535-2016, Scopus Author ID 6701612553, RSCI SPIN-code 6735-1887, https://orcid.org/0000-0002-7627-4978

Alexey N. Yurasov, Dr. Sci. (Phys.–Math.), Professor, Department of Nanoelectronics, Institute for Advanced Technologies and Industrial Programming, MIREA – Russian Technological University (78, Vernadskogo pr., Moscow, 119454 Russia). E-mail: alexey_yurasov@mail.ru. ResearcherID M-3113-2016, Scopus Author ID 6602974416, RSCI SPIN-code 4259-8885, https://orcid.org/0000-0002-9104-3529

Maxim M. Yashin, Cand. Sci. (Phys.–Math.), Associate Professor, Department of Nanoelectronics, Institute for Advanced Technologies and Industrial Programming, MIREA – Russian Technological University (78, Vernadskogo pr., Moscow, 119454 Russia). E-mail: ihkamax@mail.ru. ResearcherID G-6809-2017, Scopus Author ID 57210607470, RSCI SPIN-code 2438-6135, https://orcid.org/0000-0001-8022-9355

Об авторах

Гладышев Игорь Васильевич, к.ф.-м.н., доцент, кафедра наноэлектроники, Институт перспективных технологий и индустриального программирования, ФГБОУ ВО «МИРЭА – Российский технологический университет» (119454, Россия, Москва, пр-т Вернадского, д. 78). E-mail: i_gladyshev@mirea.ru. ResearcherID N-1535-2016, Scopus Author ID 6701612553, SPIN-код РИНЦ 6735-1887, https://orcid.org/0000-0002-7627-4978

Юрасов Алексей Николаевич, д.ф.-м.н., профессор, кафедра наноэлектроники, Институт перспективных технологий и индустриального программирования, ФГБОУ ВО «МИРЭА – Российский технологический университет» (119454, Россия, Москва, пр-т Вернадского, д. 78). E-mail: alexey_yurasov@mail.ru. ResearcherID M-3113-2016, Scopus Author ID 6602974416, SPIN-код РИНЦ 4259-8885, https://orcid.org/0000-0002-9104-3529

Яшин Максим Михайлович, к.ф.-м.н., доцент, кафедра наноэлектроники, Институт перспективных технологий и индустриального программирования, ФГБОУ ВО «МИРЭА – Российский технологический университет» (119454, Россия, Москва, пр-т Вернадского, д. 78). E-mail: ihkamax@mail.ru. ResearcherID G-6809-2017, Scopus Author ID 57210607470, SPIN-код РИНЦ 2438-6135, https://orcid.org/0000-0001-8022-9355

Translated from Russian into English by L. Bychkova
Edited for English language and spelling by Thomas A. Beavitt

Micro- and nanoelectronics. Condensed matter physics

Микро- и наноэлектроника. Физика конденсированного состояния

UDC 004.832.32
https://doi.org/10.32362/2500-316X-2026-14-3-83-105
EDN QIOHGI

REVIEW ARTICLE

Physically unclonable functions in analog integrated circuits

Evgenii Ph. Pevtsov [@], Tatyana A. Demenkova, Mikhail I. Maleto, Alexander S. Sigov, Yuri A. Korotaev [@], Nikita D. Evgenev

MIREA – Russian Technological University, Moscow, 119454 Russia
[@] Corresponding authors, e-mail: korotaevyua@yandex.ru, pevtsov@mirea.ru

• **Submitted:** 16.09.2025 • **Revised:** 20.10.2025 • **Accepted:** 27.03.2026

Abstract

Objectives. The paper provides a comprehensive overview of analog and passive physical unclonable functions (PUFs), analyzing their vulnerabilities to machine-learning (ML) attacks, and assessing their practical deployment in modern integrated circuits and Internet of Things (IoT) devices.

Methods. Quantitative metrics were used to compare PUF implementations and their formal properties, such as computability, uniqueness, implementability, difficulty of cloning, and protection against unauthorized access.

Results. Analog PUFs were shown to belong to the class of “strong” PUFs. However, special measures are required to counteract environmental and ageing effects. Examples are cited to demonstrate their near-ideal uniqueness (inter-Hamming distance ≈ 50%), high stability (intra-Hamming distance < 1%), and excellent energy performance (from units to tens of femtojoules per bit). While characterized by high stability, passive PUFs are classified as “weak” PUFs. A consideration of ML-based modeling attacks confirmed that convolutional neural networks and multilayer perceptrons outperform classical approaches. By limiting the amount of data available to an attacker, protocol-level protection prevents the PUF architecture from being modified.

Conclusions. Analog and passive PUFs expand the range of tools available for hardware authentication and anti-counterfeiting, particularly in low-power, resource-constrained IoT nodes. The most promising directions include architectures with on-chip self-calibration and minimal area/power overhead, as well as passive schemes for one-time identification and tamper evidence. However, open challenges remain in terms of standardizing readout and digitization procedures, increasing robustness to environmental variation and diverse attacks, and integrating error correction and post-processing on the chip. The practical adoption and selection of architectures requires conservative threat modeling and defense-in-depth strategies that account for current attack capabilities and likely future advances in ML.

Keywords: physically unclonable function, analog PUFs, passive PUFs, ML attacks, hardware security, device authentication, Internet of Things

For citation: Pevtsov E.Ph., Demenkova T.A., Maleto M.I., Sigov A.S., Korotaev Yu.A., Evgenev N.D. Physically unclonable functions in analog integrated circuits. *Russian Technological Journal.* 2026;14(3):83–105. https://doi.org/10.32362/2500-316X-2026-14-3-83-105, https://www.elibrary.ru/QIOHGI

Financial disclosure: The authors have no financial or proprietary interest in any material or method mentioned.

The authors declare no conflicts of interest.

ОБЗОРНАЯ СТАТЬЯ

Физически неклонируемые функции в аналоговых интегральных схемах

Е.Ф. Певцов [@], Т.А. Деменкова, М.И. Малето, А.С. Сигов, Ю.А. Коротаев [@], Н.Д. Евгеньев

МИРЭА – Российский технологический университет, Москва, 119454 Россия
[@] Авторы для переписки, e-mail: korotaevyua@yandex.ru, pevtsov@mirea.ru

• **Поступила:** 16.09.2025 • **Доработана:** 20.10.2025 • **Принята к опубликованию:** 27.03.2026

Резюме

Цели. Целью работы является комплексный обзор аналоговых и пассивных физически неклонируемых функций (ФНФ), анализ уязвимостей к атакам на основе машинного обучения и разбор практических сценариев применения в современных интегральных схемах и устройствах интернета вещей.

Методы. Использованы методы количественной оценки различий реализаций ФНФ и признаков их формального описания, включая вычислимость, уникальность, реализуемость, сложность создания клонов, защиту от несанкционированного доступа.

Результаты. Показано, что аналоговые ФНФ относятся к классу «сильных» ФНФ, но требуют специальных мер для подавления влияния факторов внешней среды и старения. Приведены примеры, демонстрирующие близкую к идеальной уникальность (inter-HD[1] ≈ 50%) при высокой стабильности (intra-HD[2] < 1%) и рекордные энергетические показатели (единицы – десятки фДж/бит). Пассивные ФНФ характеризуются высокой стабильностью, но относятся к «слабым» ФНФ. Рассмотрены атаки на основе машинного обучения, показано, что конволюционные нейронные сети и многослойные перцептроны превосходят классические подходы. Средства защиты на уровне протокола, ограничивающие объем доступной злоумышленнику информации, позволяют избежать модификации архитектуры ФНФ.

Выводы. Аналоговые и пассивные ФНФ расширяют спектр средств аппаратной аутентификации и защиты от подделок, особенно для маломощных и ресурсно-ограниченных устройств интернета вещей. Наиболее перспективны архитектуры с внутренней калибровкой и малыми накладными расходами по площади/потреблению, а также пассивные решения для задач однократной идентификации и контроля вмешательства. Остаются открытыми задачи стандартизации процедур чтения/оцифровки, повышения устойчивости к изменениям внешней среды и различным атакам, а также совмещения с коррекцией ошибок и постобработкой на кристалле. Для выбора архитектур ФНФ необходимо тщательное моделирование угроз и применение стратегий глубокой защиты с учетом будущих достижений машинного обучения.

Ключевые слова: физически неклонируемая функция, аналоговые ФНФ, пассивные ФНФ, ML-атаки, аппаратная безопасность, аутентификация устройств, интернет вещей

Для цитирования: Певцов Е.Ф., Деменкова Т.А., Малето М.И., Сигов А.С., Коротаев Ю.А., Евгеньев Н.Д. Физически неклонируемые функции в аналоговых интегральных схемах. *Russian Technological Journal*. 2026;14(3):83–105. https://doi.org/10.32362/2500-316X-2026-14-3-83-105, https://www.elibrary.ru/QIOHGI

Прозрачность финансовой деятельности: Авторы не имеют финансовой заинтересованности в представленных материалах или методах.

Авторы заявляют об отсутствии конфликта интересов.

[1] Inter-Hamming distance – внешнее расстояние Хэмминга.
[2] Intra-Hamming distance – внутреннее расстояние Хэмминга.

INTRODUCTION

Physically unclonable functions (PUFs) provide a hardware basis for trust, offering authentication and protection against counterfeiting, as well as enabling secure key derivation. The first part of the study [1] considers digital PUFs. The second part focuses on analog and passive PUFs, as well as their role in countering contemporary machine learning (ML) attacks and practical implementation scenarios.

Analog PUFs use continuous technological variations in the parameters of active and passive elements as a source of entropy. Unlike digital hardware security primitives, which are formed using discrete logic, analog PUFs rely on subtle variations in threshold voltages, currents, capacitances, and resistances. Once the circuit is switched on, these variations result in reproducible steady-state voltage and current levels that are unique to each chip. Here, digitization is performed by a comparator or analog-to-digital converter, while response stability is ensured by circuit techniques for drift and noise suppression. Due to their potentially providing higher entropy density and a larger set of challenge–response pairs (CRPs), many analog implementations are classified as "strong" PUFs. However, the characteristics of transistors and passive structures are sensitive to external factors, necessitating special measures to correct possible errors.

Passive PUFs include resistive imprints of power networks, Via PUFs[3] based on probabilistic contact formation, and Coating PUFs[4] which are provided with a specific pattern of random coating. While such passive PUFs are characterized by minimal overhead in terms of area, high stability, and—in some cases—ideal stability, this comes at the cost of a limited number of CRPs.

Due to the rapid development of ML having radically changed the understanding of PUF security, contemporary attacks are more frequently focused on extracting patterns from CRPs rather than traditional cryptanalysis. Concurrently, the research community is advancing increasingly sophisticated ML approaches to PUF modeling and countermeasures at the architecture and protocol levels, thereby refining the fundamental boundaries of resilience. The present work therefore sets out a comprehensive overview of the current state of the art, including the theoretical foundations, ML-based attack classes, and practical countermeasures, as well as their implications for PUF design and deployment.

ANALOG PUFS

Individual devices can be identified by measuring the original parameters of an electrical or electronic quantity. Sources of entropy include variations in transistor threshold voltages (TV-) and integrated circuit identification (ICID-) PUF, current arbiters, diode structures, quasi-adiabatic logic based (QUAL-) PUF, and adiabatic static random-access memory (SRAM).

The simplest TV-PUF variant, which is used to identify integrated circuits (ICs), involves analyzing the change in threshold voltages of integrated transistors resulting from inevitable technological variations during manufacturing. Here the challenge consists in the number or location of the transistor component, while the response is the corresponding threshold voltage value.

The approach to applying PUF proposed in [2] involves the assignation of a unique identification tag to each instance of a conventional IC without requiring special processing or programming after manufacture. Here several transistors of the same design are combined into an addressable matrix in an ICID-PUF device (Fig. 1).

In the described ICID-PUF, an addressable transistor controls a resistive load. Due to manufacturing variations, the threshold voltages of these transistors fluctuate, resulting in an unpredictable current flowing through the load. The voltage across the load is then measured and converted into a sequence of bits using an auto-zero comparator. This method has been verified via experimentation on 55 microchips manufactured using 0.35-μm complementary metal–oxide–semiconductor (CMOS)[5] technology. At the highest environmental fluctuations, an intra-Hamming distance (intra-HD) μ_{intra} = 1.3% is obtained, while the inter-Hamming distance (inter-HD) value of μ_{intra} is very close to 50%. With an input clock frequency of 1 GHz, the PUF design featuring 64-bit keys on transistors consumes 0.18 μW/bit while achieving uniqueness and uniformity indices of 50%. Reproducibility of this PUF variant has been demonstrated to be independent of IC ageing processes, as well as 45 nm-, 65 nm-, and 90 nm process nodes.

In [3], a cascade comprising three stages of 20 CMOS inverters and diode-connected transistors is proposed. This forms a voltage divider whose output voltage depends on variations in the threshold voltage. This PUF demonstrates an

[3] Via PUF is a technology based on the use of microscopic via holes in metallic layers of semiconductors.

[4] Coating PUF is a technology that utilizes a protective coating for its operation.

[5] The complementary metal–oxide–semiconductor (CMOS) structure is a collection of semiconductor technologies used for the fabrication of integrated circuits and the related circuitry in microcircuits.

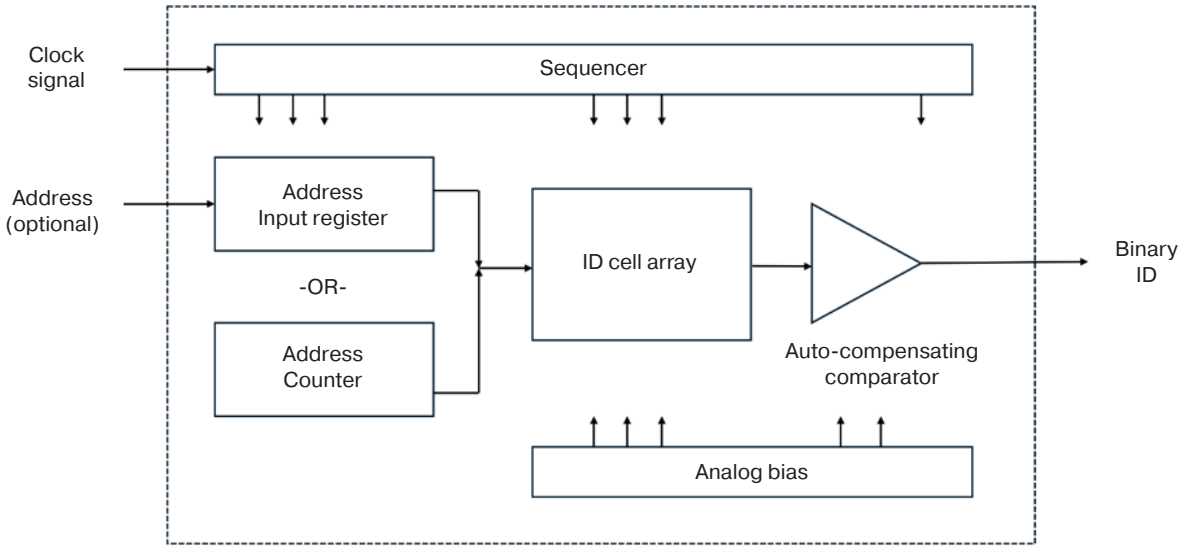


Fig. 1. Block diagram of a device for IC identification based on PUF [2]. ID is identifier

inter-HD of approximately 50.65% and an intra-HD of approximately 6.96%. Due to operating below the threshold voltage, the circuit demonstrates low power consumption (no more than 0.43 pJ/bit); however, it is sensitive to comparator noise. Reliability of recognition is enhanced by averaging 15 samples.

It should be noted that these characteristics may change in the event of increased noise due to changes in the library element parameters (characterization by PVT corners: process, voltage, and temperature variations) or the ageing of the active device, which can affect the reliability of PUF responses.

As described in [4], the difference between nominally identical currents is accurately measured when transistors are turned on. This is achieved using a specialized hardware element of an artificial neural winner-takes-all network, which acts as an arbiter. The obtained values are intra-HD ≈ 1.57%, inter-HD ≈ 49.8% and 97.7% reliability in the temperature range from −20°C to +120°C, with a ±300 mV variation in supply voltage. The circuit's energy consumption is 5.67 pJ/bit. An improved version of this circuit, which is implemented in a 130 nm technology node, uses a pair of cascode current mirrors with double gain-boosting[6] amplification to reliably maintain the operating point at around 50 nA, thereby increasing the output resistance. Consequently, bit instability does not exceed 1.56% across a supply voltage range of 0.6–2 V and a temperature range of 0–75°C. Intra-HD does not exceed 0.49% on average, with virtually perfect uniqueness (inter-HD ≈ 50%). Together with an energy consumption of 5.36 fJ/bit and an area of 72 μm^2/bit, these characteristics render the cell highly attractive for integration into Internet of Things (IoT) modules. Experiments on 21 128-bit chips confirm the statistical stability of the solution, as well as demonstrating one of the best aggregate quality factor indicators (figure of merit is 17 relative units) while maintaining ease of integration into ICs and potential for further scaling.

Another version of an analog diode PUF is implemented in [5], where the device signature is formed from diodes present in the output ports of the IC. Measures are provided to increase the uniqueness of the signatures, as well as compensating for temperature differences and losses in the supply conductors.

The study [6] examines the use of PUFs based on diode structures in ICs. The authors demonstrate that changes in the conductivity of oxide-based Schottky diodes as a result of the technology process provide an effective source of random numbers for using PUFs without the need for switching operations. It is demonstrated that the naturally occurring electron accumulation region in an oxide semiconductor film can be partially eliminated using a mild oxygen plasma treatment. This leads to a significant change in nonlinearity, thus providing an exotic source of entropy. In general, soft plasma-treated Schottky diodes demonstrate near-perfect uniformity, uniqueness, and an ideal entropy value, thus eliminating the need for additional equipment, as well as reducing space requirements and energy costs. These results are promising for the development of hardware-embedded PUFs that enable the implementation of energy-efficient cryptographic equipment.

A QUAL-PUF is formed from the composite capacitor and transistor components of an adiabatic logic circuit (an energy-efficient system that converts

[6] Gain-boosting is an analog circuit technique in which auxiliary gain in the feedback loop increases the effective gain of the cascade node, thereby increasing its output resistance.

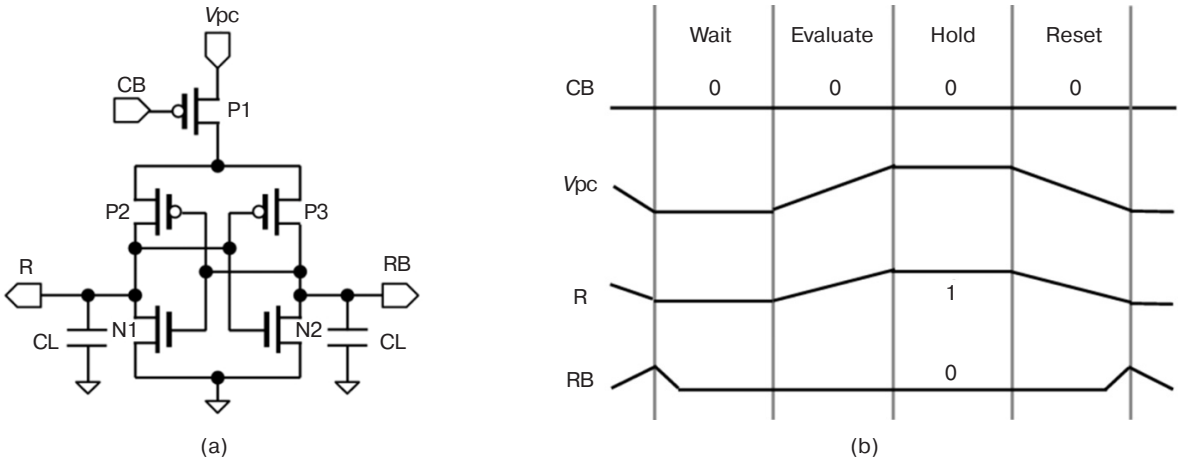


Fig. 2. QUAL-PUF circuits [7]:
(a) QUAL-PUF; (b) timing diagram.
*V*pc (power clock) is supply clock signal; CB is challenge bit;
P1 is control transistor; P2 and P3 are *p*-MOS[7] transistors of the bistable element;
R (response) is main response bit; RB (response bit) is complementary response bit output,
N1 and N2 are *n*-MOS transistors of the bistable element;
and CL is equivalent load capacitance of the output node

the charge accumulated in the load capacitor after operations into a signal) due to technological changes during the manufacturing process. Circuits that allow for such complete recycling are usually complex and occupy a large area. In [7], however, a quasi-adiabatic circuit is used which only recovers most of the capacitor charge. As shown in Fig. 2, this PUF implementation involves applying an increasing voltage to two theoretically identical transistor elements in the circuit.

Mismatches in transistor parameters are caused by differences that occur during production. This results in one transistor having greater conductivity and charging the load capacitor faster. This creates a stable response bit for each elementary cell of the circuit, which is similar to the effect of MOS transistor mismatch in traditional PUFs. Since adiabatic logic operates within specific charge/discharge cycles, a distinctive feature of this implementation is the evaluation of each FIFO (First-In, First-Out) cell at only one of four identical time intervals. To account for this, each PUF bit unit consists of four cells operating with a time shift of a quarter of a clock cycle. The implementation of a 4-bit adiabatic PUF module composed of four QUAL-PUFs is shown in Fig. 3.

Each elementary cell is controlled by a challenge bit (CB), which initiates the process, along with four clock pulses ($V\varphi_1$–$V\varphi_4$), which are trapezoidal power-clock signals of quasi-adiabatic logic shifted by 90° in phase. These phases correspond to the wait, evaluate, hold, and reset states. Each local block (QUAL-PUF1–QUAL-PUF4) generates a pair of output signals comprised of the main response bit (R1–R4) and its complementary output (RB_1–RB_4). Connections between the blocks are made via pass keys, which are controlled by the corresponding phase signals $V\varphi_i$ to ensure sequential cycles of charging and restoration of load energy.

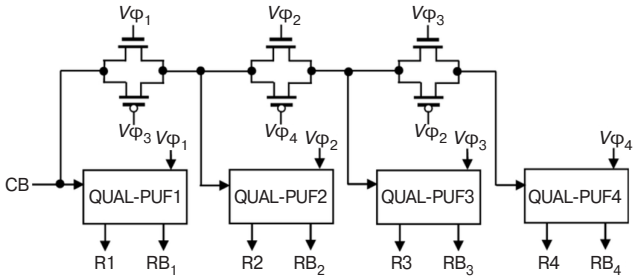


Fig. 3. 4-bit adiabatic PUF module composed of four QUAL-PUFs [7]

[7] Metal–oxide–semiconductor.

As shown in Fig. 3, if the first cell operates in the hold phase, the next cell operates in the reset phase, while the remaining two cells operate in the wait and evaluate phases, respectively. By sampling all four outputs simultaneously, four bits of response can be obtained. Sampling at different phases of the clock signal forms different combinations of bits, thus enabling the four PUF modules to generate up to 16 bits of response. Since each module contains four repeating random bits, this structure significantly complicates the PUF modeling process for a potential attacker.

Study [7] provides an example of implementing a 4-bit low-power chip with memory based on a 6T adiabatic memory cell manufactured using a standard 0.18 μm CMOS process with a supply voltage of 1.8 V, whose module dimensions measure 58.7 × 5.7 μm. The simulation results show an inter-HD value of 47.58%, a reliability of 95.10%, and a dissipated energy of 29.73 fJ/bit/cycle.

These results are confirmed in later implementations of a similar circuit [8], which demonstrate that such a PUF, when simulated, provides an average reliability of 98.51% under temperature and supply voltage fluctuations. The circuit offers a uniqueness of 49.75% and a power consumption of 15.92 fJ/bit/cycle. These estimates are consistent with measurements taken on manufactured samples, which confirm the PUF's required functionality with a reliability of 96.92% at room temperature.

An example of a circuit based on adiabatic SRAM using 0.18 μm CMOS technology is given in [9]. As well as consuming less energy than conventional memory-based PUFs, this circuit offers good uniqueness and reliability. When simulated using simulation program with integrated circuit emphasis (SPICE)[8], the proposed circuit consumes 13.88 fJ/bit/cycle, and the uniqueness and reliability values when the PUF circuit is connected in a 4-bit cascode are 50.07% and 99.51%, respectively.

PASSIVE PUFS

Analog PUFs include passive devices that use statistical variations in passive elements or structures to create a unique "fingerprint" for the device. Passive PUFs typically do not require a special stimulus signal due to their random parameters being embedded in the structure from the outset and thus capable of being read directly.

Further on, three examples of passive PUFs are considered: (1) based on variations in conductor resistance; (2) based on via holes (Via PUF); and (3) based on using a protective coating (Coating PUF).

[8] Simulation Program with Integrated Circuit Emphasis (SPICE) is an open-source simulator for general-purpose electronic circuits.

Differences in the manufacture of component connections can result in variations in power distribution maps in ICs, which can also serve as indicators of the uniqueness of a particular instance. In this case, a PUF is implemented by adding extra components to the device to enable the connection of each branch of the power distribution network to the ground bus, thus bypassing the existing components. The device signature is formed by measuring the voltage drops (or resistance values) across these sections of the circuit. Fluctuations in the thickness, width, and grain size of the metal during production cause slight variations in the resistance of the busbar segments from crystal to crystal. This can be achieved by applying test currents through specific sections of the power network and measuring the resulting voltage drops, which depend on the circuit's total resistance. By combining the results of several such measurements, a unique response vector is created that characterizes a given instance of the microcircuit. The challenge specifies the number or location of the circuit sections, while the response specifies the corresponding voltage drop or resistance value. An example of this process is presented in [10], which provides the results of measuring changes in equivalent resistances in the power distribution systems of 24 identical microcircuits manufactured using 65 nm technology.

A resistive PUF is highly stable to external conditions due to being based on passive metal structures. Variations in metal resistance depend linearly on temperature and weakly on supply voltage, making compensation for external influences simpler than with transistor effects. In particular, transistor (active) PUFs typically require calibrations or correction algorithms to account for changes in PVT parameters, whereas a passive metal network provides more reproducible results without complex adjustments. Since the only necessary additional circuits are those required for readout and digitizing the response, the use of a distributed power network already presents in each chip results in minimal overhead in terms of area. The distributed nature of the metal grid creates statistical variation across different areas to increase uniqueness; as such, the probability of two chips producing the same resistive fingerprint by chance is negligible. Another advantage of this PUF is its hardware stability, which is achieved by designing the metallization in such a way as to ensure it does not degrade over time (electromigration is eliminated by selecting the dimensions of the conductors). Since it is practically impossible to copy the resistances in all the power nodes of a clone without replicating the entire technological process of the origin, the counterfeiting of such devices is highly complex. This approach has two main limitations: (1) the relatively small

amount of data generated (usually a unique key/ID is obtained, although varying the measurement points can produce several bits); (2) the need for precise analog measurements of small resistance differences. Nevertheless, experimental samples of resistive PUF have proven viable; for instance, in 65 nm CMOS, all 36 tested crystals could be reliably distinguished based on power network variations.

In [10], which considers analog PUFs, it is proposed that voltage drops and equivalent resistances be recorded in microcircuit power supply circuits due to these electrical parameters being affected by random factors in manufacturing technology. Experiments carried out on chips manufactured using 65 nm CMOS technology demonstrate that, when measuring equivalent resistances, the quantitative characteristics μ_{intra} and μ_{inter} are approximately 0.04 and 1.5 Ohm, respectively.

The idea behind Via PUF, which is based on variations in the formation of via holes, is to deliberately violate the design standards for via holes in IC layout by using sizes that are slightly smaller than the permissible minimum. The exact size of each via gives it about a 50% chance of successfully forming a connection between the layers by being filled with metal or of remaining open [11]. Such events occur randomly due to uncontrollable fluctuations in the manufacturing process. Consequently, once the chip is manufactured, many of the embedded connections in it turn out to be either conductive ("1") or open ("0"), forming a pattern unique to the crystal. Via PUF is read by measuring the resistance of the embedded contacts: high resistance indicates an absence of a metal connection ("0"), while low resistance indicates a formed contact ("1"). An important advantage of this approach is its exceptional reliability: either the contact is formed or it is not, and the metal connection is unaffected by changes in temperature or supply voltage. Experiments have shown that the bit error rate for PUFs based on the formation of vias is practically zero, thus obviating the need for error correction using redundant codes. Additional processing, such as two-stage XOR conversion, is employed to eliminate offsets and improve bit uniformity. The main advantages of Via PUFs include high fingerprint uniqueness (with inter-codeword Hamming distances of around 50%) and resistance to ageing. Such random contacts can be distributed throughout the crystal among conventional vias, which makes them difficult to detect when an attacker reverse engineers the chip. Since standard CMOS layers and materials are used with the sole addition of "embedded" contacts of a specific size, the implementation does not require non-standard technological processes. However, the number of bits generated by Via PUF is fixed by the circuit (usually acting as a unique identifier rather than a reusable CRP); for this reason, the size of the vias must be calibrated for a specific technical process to ensure a ~50% probability of filling and any borderline cases rejected to avoid unstable bits.

A study of methods to increase the reliability of a system on a chip based on ARM Cortex-M4 microprocessors (ARM, United Kingdom) conducted by Lee[9] justifies the chosen method of forming a PUF. It notes that there is a medium-sized area, conventionally referred to as the PUF zone, in which the probability of contact forming is 50% if the size of the via is smaller than that specified in the design (Fig. 4a).

A microscopic image of a silicon Via PUF in cross section is shown in Fig. 4b. Contact vias of varying quality are clearly visible, with the following definitions: (1) the contact is open if it does not provide a connection to the silicon substrate; (2) the contact is shorted if it provides an electrical connection between the layers. In order to form the PUF after manufacturing is complete, it is necessary to exclude unreliable contacts by measuring the resistance of the through or contact connection.

For example, in one process node, resistance greater than 1 MOhm is identified as an open circuit, while resistance less than 50 kOhm is identified as a short circuit. All vias between these two values are cut off, these cut-off values being selected separately for each process. Once classified as a short circuit or open circuit, the connection remains unchanged regardless of PVT variations to ensure a zero-bit error rate. This important feature of Via PUF confirms the basic reliability of the technology.

As mentioned previously, achieving complete randomness is essential for PUFs, where ideal randomness is defined as an HD of 0.5, or 50%. To increase reliability, this work used a two-stage selection based on the XOR criterion (Fig. 5). A total of 405 test chips were manufactured with 16 different via sizes, each having 7680 bits, thus providing a total of 122800 bits of input data for the PUF. The first XOR stage reduces this to 7680 bits, which then pass through the second XOR stage to generate the final result of 640 bits, achieving a value of μ_{inter} = 0.4972 at a standard deviation σ_{inter} = 0.0205. Figure 5 shows the selection of Via PUF chips based on the results of the United States National Institute of Standards and Technology's[10] standard randomness tests SP 800-22 and SP 800-90B.

A similar PUF using a binary response to uniqueness (learning resilient and reliable digital PUF) based on variations in interconnection lithography is also used in [12].

[9] Lee T.K. *Via PUF technology as a root of trust in IoT supply chain*. Global Semiconductor Alliance; 2024. https://www.gsaglobal.org/forums/via-puf-technology-as-a-root-of-trust-in-iot-supply-chain. Accessed June 16, 2025.

[10] The National Institute of Standards and Technology, NIST. https://www.nist.gov/. Accessed June 16, 2025.

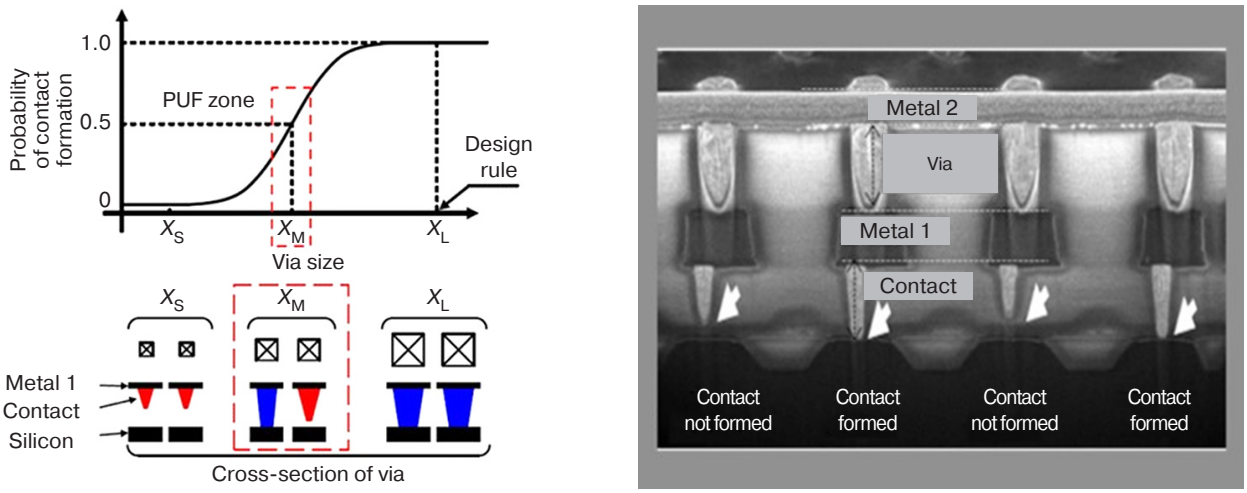


Fig. 4. Implementation of a PUF based on variations in via formation:
(a) probability of contact formation depending on the size of the via;
(b) cross-section[11] of a microscopic image of a Via PUF.
X_S is the size of the via at which no contact is formed; X_L is the size of the via determined by the design rules;
X_M is the intermediate size of the via at which contact formation is probabilistic

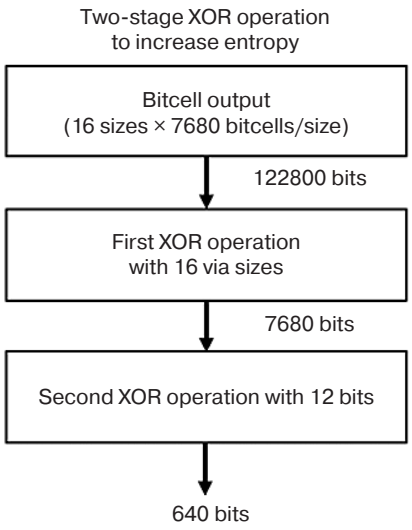
The protective Coating PUF uses an external, random, dielectric coating that is applied over the crystal in order to form a unique fingerprint. The classic implementation, as described in references [13, 14], involves placing a grid of metal conductors (e.g., a comb-like electrode structure) on top of the IC and filling the space between them with an optically opaque polymer containing randomly distributed dielectric nanoparticles. Due to the chaotic arrangements, sizes, and dielectric properties of these particles, the electrical capacitance between each pair of conductors becomes a random variable. In other words, nominally identical capacitors formed by the upper conductors acquire a range of capacitance values that is unique to each microcircuit instance. By reading a set of such capacitors (for example, by measuring leakage currents or charge/discharge time constants), it a set of random bits can be obtained depending on local variations in the dielectric permeability of the coating. These bits constitute a unique device identifier that is physically unclonable due to the unique distribution of particles in the coating layer.

As shown in [15], when an opaque, chemically inert dielectric layer is applied to the upper layer of IC metallization, the electrical capacitance values measured for each chip are random and unique. The implementation scheme and operation principle of the Coating PUF are shown in Fig. 6.

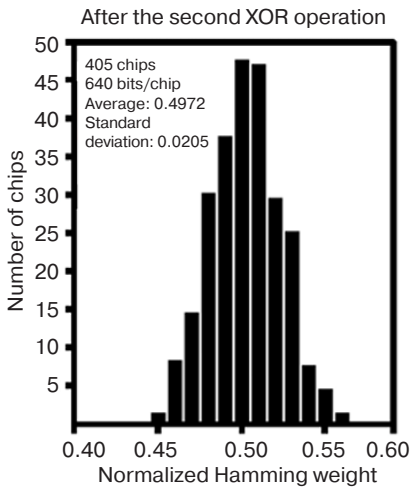
The results of digitized measurement carried out on 36 manufactured chips, in which each chip is used to test 31 capacitive sensors, show a high degree of randomness ($\mu_{inter} \approx 50\%$) and low noise ($\mu_{intra} < 5\%$).

A protective Coating PUF offers two key advantages. Firstly, it enhances the hardware security of the crystal by preventing direct optical examination and reading of the internal circuits. The opaque top layer acts as a protective mask. As such, any attempt to remove or damage this layer inevitably changes the distribution of particles and capacitance, destroying the device's original "fingerprint." Thus, Coating PUF not only provides a unique key, but also serves as a kind of tamper sensor. If tampered with, the original identifier is lost, thus revealing the hack. Secondly, due to the large-scale random formation process involving the mixing of millions of particles, the probability of two such PUFs coinciding is extremely low. Reproduction would require copying processes at the atomic level, which is practically impossible. However, a key disadvantage in this approach is the need for an additional manufacturing process, namely the application and curing of a special coating, which increases the cost. Additional analog circuits or high-precision external measurements required for reading the imprint may be affected by external conditions (e.g., temperature affecting the dielectric). Devices with Coating PUF have demonstrated their effectiveness in authentication systems, particularly in the implementation of radio frequency identification tags, where a random epoxy coating is used as a source of a 128-bit key that is destroyed when physical access is attempted.

[11] Lee T.K. *Via PUF technology as a root of trust in IoT supply chain*. Global Semiconductor Alliance; 2024. https://www.gsaglobal.org/forums/via-puf-technology-as-a-root-of-trust-in-iot-supply-chain. Accessed June 16, 2025.



(a)



(b)

NIST SP 800-22 test for Via PUF

Test type	Average *p*-value	Test result
Frequency	0.4327	+
Block frequency (*M* = 128)	0.4734	+
Cumulative sums	0.4259	+
Runs	0.4956	+
Long run	0.4506	+
Rank	0.4569	+
FFT	0.5090	+
Approximate entropy (*m* = 6)	0.4549	+
Sequential test	0.4920	+

(c)

Fig. 5. Two-stage Via PUF selection algorithm based on the XOR criterion and NIST test results[12].
FFT is fast Fourier transform. *M* is the length of the block in the frequency test within the block,
m is the length of the bit pattern in the approximate entropy test. A block is a fragment of a fixed-length bit sequence into which the NIST test divides the entire sequence being tested

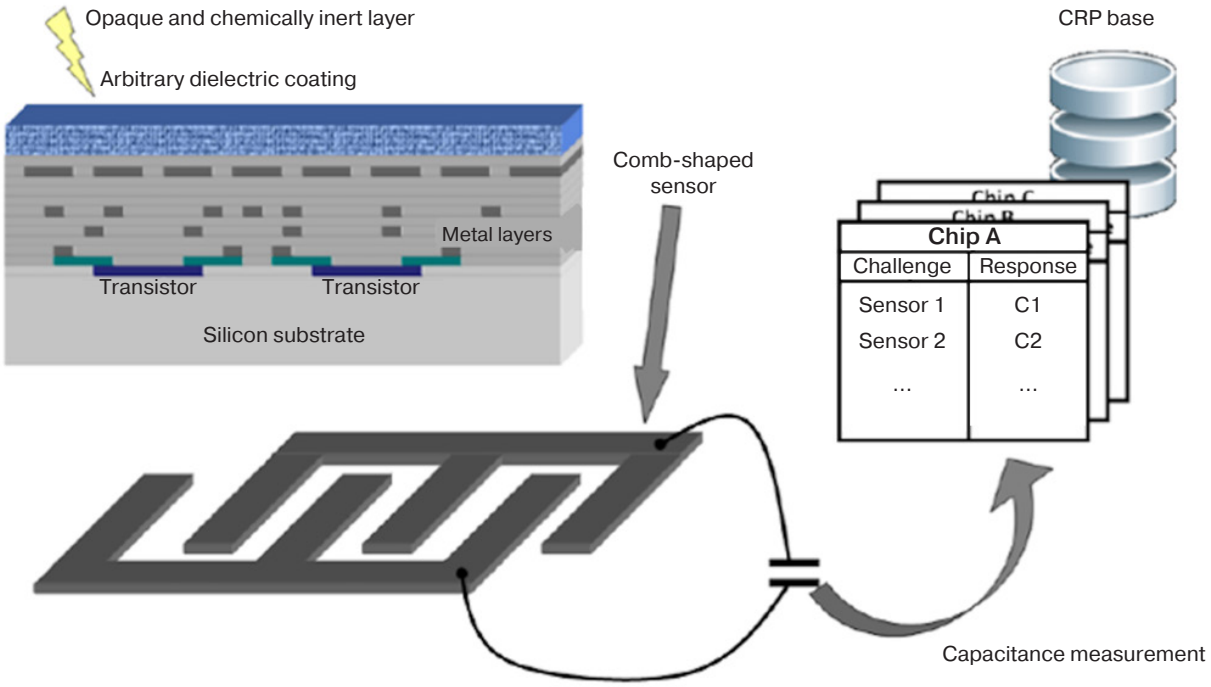


Fig. 6. Example of passive PUF implementation based on an inert dielectric layer [14]

[12] Lee T.K. *Via PUF technology as a root of trust in IoT supply chain*. Global Semiconductor Alliance; 2024. https://www.gsaglobal.org/forums/via-puf-technology-as-a-root-of-trust-in-iot-supply-chain. Accessed June 16, 2025.

Table. Characteristics of analog and passive PUFs

PUF type / reference	Year of publication	PUF characteristics					
		Inter-HD	Intra-HD	Platform	Sensitivity to external conditions		Estimated implementation complexity
					Temperature	Voltage	
Subthreshold [3]	2019	50.65%	~7%	SPICE model TSMC 65 nm CMOS	From −20°C to 85°C	0.75–0.9 V	High
Current Mirror[13] [4]	2023	49.84%	1.57%	SPICE model TSMC 65 nm CMOS	From −20°C to 120°C	±300 mV	High
Adiabatic SRAM[14] [7]	2020	47.58%	4.9%	ASIC 180 nm CMOS	From −40°C to 100°C	1.8 V	Medium
6T Adiabatic[15] [8]	2024	49.75%	1.49%	ASIC 180 nm CMOS	From −40°C to 100°C	0.9–1.8 V	Medium
Adiabatic Logic[16] [9]	2024	50.07%	0.49%	ASIC 180 nm CMOS	From −50°C to 100°C	–	Medium
Via[17] [11]	2020	49.99%	~0%	ASIC 130 nm CMOS	From −55°C to 125°C	1.65 V	Low

The table below summarizes data from publications presenting original results on the implementation of PUFs based on analog circuits and circuits with variations in vias. The key metrics selected are inter-HD, i.e., the distance between two PUF responses from different PUF instances using the same call, and intra-HD, which is the distance between two PUF responses obtained from the same PUF instance and using the same call. These are referred to as "uniqueness" and "reliability" in a number of studies, respectively.

The table also shows the voltage/temperature variations at which the characteristics are measured, as well as the extent to which they change (the change in intra-HD is indicated in parentheses) if such data is provided. The estimated complexity of implementation (high, medium, or low) indicates the relative hardware costs and technical complexity involved in implementing a particular type of PUF (e.g., balancing paths, selecting element parameters, or change technical processes).

VIOLATION OF PUF SECURITY BASED ON ML METHODS

Recent advancements in ML techniques have significantly altered the perception of PUF security. The vulnerability of PUFs to ML-based attacks has led to extensive research on attack methodologies and defense measures [3, 16–19]. The paradigm shift in the approach to analyzing PUF security due to such attacks tends away from traditional cryptanalytic approaches and towards data-driven modeling techniques that exploit patterns inherent in PUF CRPs.

As well as developing more sophisticated attack strategies that use adversarial ML techniques [17, 20], researchers have proposed new PUF architectures and protocols designed to counter these attacks [3, 18, 19]. The following section presents an analysis of the current state of ML attacks and defenses in PUF systems, which is based on the latest advances in both attack methods and countermeasures. As well as examining the theoretical foundations of PUF security, it analyzes various categories of ML-based attack, evaluates proposed defense measures, and discusses the implications for future PUF development and deployment strategies.

PUF architectures and security properties

Based on their behavior in CRPs, PUFs can be categorized into two distinct types: strong and weak [21]. The possibility of using strong PUFs to generate a large

[13] PUF based on the use of an array of current mirrors.
[14] A type of SRAM that employs adiabatic energy recovery methods.
[15] A six-transistor PUF that operates on adiabatic logic principles.
[16] PUF based on adiabatic logic.
[17] PUF derived from variations in the formation of vias.

number of CRPs makes them suitable for authentication protocols that use multiple CRPs without exhausting the available challenge space. Examples include arbiter-type PUFs, ring oscillator PUFs, and various composite architectures [17–19]. In contrast, weak PUFs, which have limited challenge space, are generally employed for key generation, where one or more responses are extracted and processed using error correction procedures [22].

The Arbiter physical unclonable function (A-PUF) is one of the most widely studied strong PUF architectures. This architecture uses the difference in delays between two nominally identical signal transmission paths to generate a response [16, 18]. The challenge bits control the switching elements that determine the path configuration in such a way that the arbiter circuit can then evaluate which path is faster to produce a binary response. Despite its conceptual simplicity, this type of PUF has proven vulnerable to ML attacks due to its linear additive delay model [17, 19]. To improve security against ML attacks, more complex architectures have been developed that attempt to fundamentally change the mathematical relationships between inputs and outputs as a means of preventing effective ML modeling [18, 19, 22].

The most important aspect of PUF implementation is the processing of noise inherent in PUF responses. This noise is caused by environmental fluctuations, aging, and measurement errors [16, 22]. Reliable key generation and authentication relies on error correction mechanisms, which are typically implemented using fuzzy extractors or auxiliary data algorithms. Here, the most common approach is to store auxiliary data for enabling the correction of noisy PUF responses without disclosing the actual response values. However, recent research has shown that this auxiliary data can leak significant information about the PUF responses, enabling ML attacks even when the actual response values are unavailable [22]. This represents a fundamental vulnerability in PUF-based systems that use standard error correction approaches, particularly those using linear block codes, such as repetition codes.

Attacks on PUFs using ML

The vulnerability of PUFs to ML-based attacks has been demonstrated through the application of traditional ML techniques to A-PUFs and their variations [3, 17]. These attacks are successful due to the underlying mathematical relationships that underpin PUF designs. In the case of A-PUFs, the delay difference can be modeled as a linear function of challenge bits and delay parameters. This makes A-PUFs vulnerable to linear classification techniques [16, 17]. Despite increased complexity introduced by XOR operations or other nonlinear transformations, the fundamental structure is often still analyzable with appropriate ML algorithms. Typically, classic A-PUFs can be accurately modeled with over 95% accuracy using fewer than 10000 CRPs along with conventional logistic regression methods. Due to the linear nature of their delay model, these PUFs are particularly vulnerable to mathematical analysis and ML modeling.

Deep learning (DL) models can automatically detect relevant features and nonlinear relationships in CRPs, thus eliminating the need for manual feature selection. DL-based attacks demonstrate superior performance on all tested architectures compared to classical ML approaches. Convolutional neural networks and multilayer perceptrons also outperformed classical ML approaches, particularly against complex physical unclonable function (PUF) architectures designed to defend against traditional attacks [3, 17]. Convolutional neural networks achieve modeling accuracy above 90% for various types of PUFs while requiring fewer training samples than traditional methods. This suggests that the complexity of modern ML algorithms exceeds the defensive capabilities of current PUF designs.

Transfer learning methods have also been investigated, whereby models trained on a single PUF instance or architecture are adapted for use in attacks on different PUFs [21]. This approach reduces the amount of training data required for successful attacks at the same time as demonstrating the generalizability of PUF models trained on similar architectures.

Evolutionary strategies, particularly the covariance matrix adaptation evolution strategy, have been shown to be effective in ML attacks against more complex PUF designs [17, 21]. These methods enable the modeling of PUF behavior by evolving populations of candidate models and selecting those that best match the observed CRPs. The flexibility of evolutionary approaches makes them particularly dangerous for PUFs with complex internal structures.

One particularly challenging approach involves using Siamese neural networks to model a PUF using auxiliary data [22]. This technique exploits the redundancy inherent in error correction codes to extract trainable features and labels without direct access to PUF responses. By applying XOR relationships in linear block codes, attackers can train models to predict PUF behavior using only publicly available auxiliary data and tasks.

Advanced attack strategies, which have moved beyond simple CRP modeling, now utilize additional sources of information. Reliability-based attacks exploit the fact that PUF responses near decision boundaries are more sensitive to noise and environmental

fluctuations [16, 17]. By analyzing the stability of responses over multiple measurements, attackers can gain insights into the internal structure and parameters of PUF circuits.

Another significant threat vector is side-channel attacks, in which physical information such as power consumption, electromagnetic emissions, or synchronization variations is exploited to strengthen attacks on ML models [21]. Such attacks can be highly effective when combined with traditional ML approaches due to the provision of additional constraints and information that can improve model accuracy.

The combination of multiple attack vectors can create especially powerful threats. For example, by combining partial response data with side-channel measurements, the number of CRPs required for successful ML attacks is substantially decreased. Such multimodal approaches underline the importance of considering all potential sources of information leakage when analyzing the security of PUF models.

Quantum computing poses a potential threat in the future that could radically alter the security level of PUF. While current quantum algorithms may not be directly applicable to PUF modeling, advanced computing capabilities could facilitate new attack strategies or render currently impractical attacks feasible.

Defense methods against ML attacks

Several architectural improvements have been proposed for enhancing the resistance of PUFs against attacks implemented using ML. One such fundamental approach is the cyclic redundancy check PUF (CRC-PUF) design, which breaks the direct correlation between challenges and responses using cryptographic transformations [18]. By applying CRC operations with randomly chosen polynomials, the CRC-PUF ensures that the likelihood of recovering the transformed challenge is cryptographically small, thus circumventing traditional ML attacks.

Another PUF architecture employs a two-round challenge processing mechanism in which subsequent challenges are modified using intermediate responses [17]. By concealing the direct connection between input challenges and final responses, this approach significantly complicates the creation of predictive models by ML algorithms. Here, the randomization introduced by the intermediate processing stages increases the effective challenge space to reduce the correlation between different CRPs.

Innovations at the hardware level have also shown promise in countering ML attacks. Subthreshold PUF with a voltage divider operates in weak inversion regions, in which large threshold voltage fluctuations increase randomness [3]. The cascading connection of several stages combined with careful bias control provide this design with strong statistical properties that maintain its resistance to various ML algorithms. These properties include support vector machines, logistic regression, and artificial neural networks.

Rather than modifying the underlying architecture of the PUF, protocol-level defenses aim to limit the information available to potential attackers. Challenge restriction strategies limit the number of CRPs that can be observed, thereby preventing attackers from accumulating sufficient training data for effective ML attacks [16, 21]. However, due to inherently restricting the capabilities of strong PUFs, this approach may not be feasible for applications requiring numerous authentication operations. To minimize information leakage during the configuration and operation phases of PUFs, improved logging protocols have been developed [21, 23]. These protocols prevent unauthorized access to training data by carefully controlling the distribution of CRPs and implementing secure computation methods. Multilateral computation and homomorphic encryption can ensure that PUFs operate without revealing confidential information to potential attackers.

Advanced methods (Fig. 7) offer a new approach to protecting PUFs by deliberately introducing errors into the CRP process [20]. These methods can significantly reduce the accuracy of ML models by poisoning the training data available to potential attackers while maintaining correct operation for legitimate users who understand the poisoning strategy. This involves periodically providing incorrect responses to challenges, thereby creating a dataset from which standard ML algorithms cannot effectively learn.

The security implications of using auxiliary data in PUF systems have led to the development of specialized error correction approaches [16, 22]. It has been demonstrated that traditional concatenated coding schemes, particularly those employing repetition codes as inner codes, are highly vulnerable to ML attacks involving auxiliary data analysis. The redundancy of these codes gives attackers enough information to create effective models without needing to access the actual PUF responses. Codes with higher frequencies and more complex structures are more resistant to auxiliary data attacks, whereas simple codes such as repetition codes should be avoided in critical applications [22]. The analysis of various code families confirms that vulnerability to ML attacks is directly affected by the number and complexity of XOR relationships in the code structure.

Safe error correction approaches include syndrome construction methods, systematic coding with low data leakage, and specialized polar codes that minimize

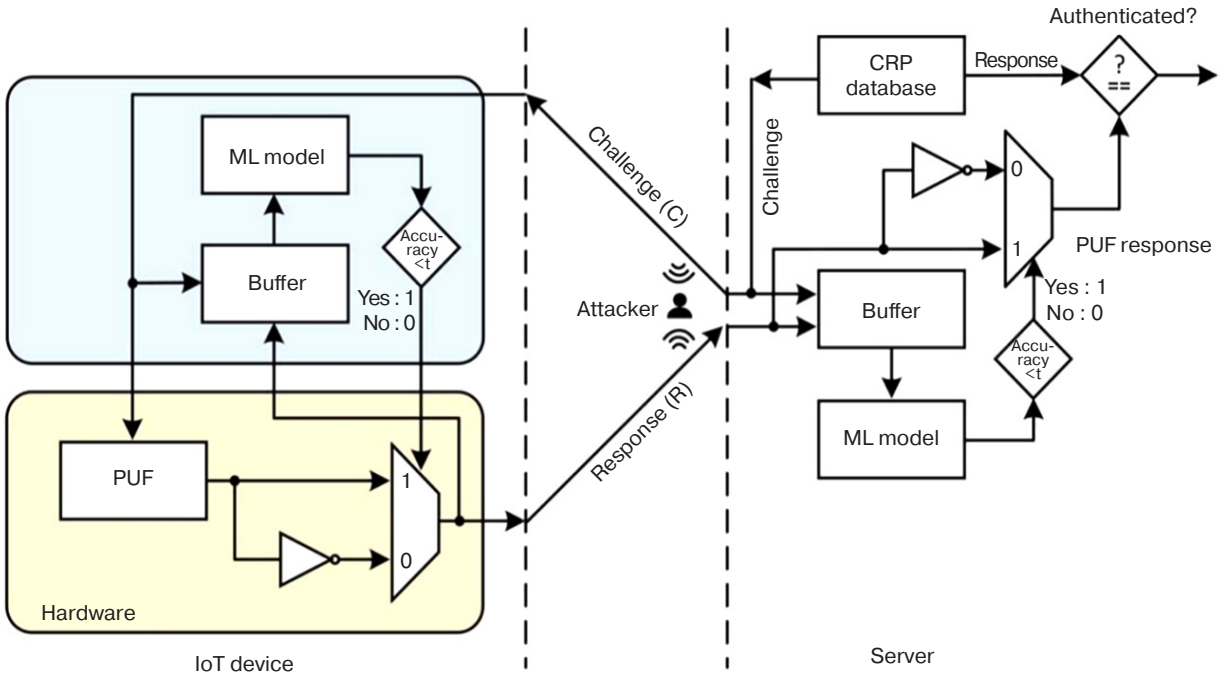


Fig. 7. Block diagram of the error injection method

data leakage while preserving the ability to correct errors [22]. These methods aim to reduce the correlation between auxiliary data and PUF responses, thereby making it more difficult for attackers to extract useful training data.

The practical effectiveness of both attacks and defenses can be understood through research into physical implementation [3, 18, 19]. Significant differences in the success rates of ML attacks and the required amounts of training data have been revealed by empirical studies of various PUF architectures. The field-programmable gate array implementation of CRC-PUF is particularly noteworthy for its ability to withstand ML attacks while still maintaining efficient area and power consumption. The achievement during development of normalized inter-HD and intra-HD values of 0.5065 and 0.0696, respectively, indicates favorable statistical properties for security applications.

Area and power analyses confirm the possibility of implementing ML-resistant PUFs at a reasonable cost as compared to classical architectures. For example, the requirement of 1032 equivalent gates in CRC-PUF as compared to 646 for the base Arbiter PUF is a modest increase, but one that results in a significant improvement in security. Similarly, while the ML resistant PUF architecture achieves ML resistance, its area significantly exceeds the practical limits for embedded applications.

More complex architectures demonstrate varying degrees of PUF resistance. Arbiter-type functions using multiple XOR cascades require an exponentially increasing number of CRPs to mount a successful attack, yet remain vulnerable to advanced techniques when sufficient data is available [17, 19, 20]. Conversely, reducing the frequency of successful attacks to below 60% even when using large training data sets represents a significant improvement over classical design.

Implementing PUF using an array of subthreshold voltage dividers in 65 nm CMOS technology achieves promising results with a power consumption of only 0.43 pJ/bit [3]. When exposed to various ML attacks, including logistic regression, artificial neural networks, and support vector machines with nonlinear radial basis function kernels, the prediction accuracy remains at around 60%, demonstrating practical resistance to ML.

Research directions for methods of protection against ML attacks

Developing ML-resistant PUFs requires a balance between security requirements and practical considerations such as size, power consumption, reliability, and performance [3, 18, 19]. High-security designs may require additional hardware, more advanced error correction mechanisms, or reduced performance, all of which may be unacceptable for cost-sensitive applications.

Analysis of ML attacks on PUFs reveals fundamental limitations in the security provided by current implementations. Many practical PUF implementations

do not achieve the maximum theoretical level of entropy, making them vulnerable to statistical, pattern-based ML attacks [16, 17, 22]. The mathematical models that underpin PUF designs often contain learnable patterns that can be exploited by sophisticated ML algorithms. Therefore, achieving true resilience against ML attack may require fundamentally distinct approaches to PUF design rather than incremental enhancements to existing systems.

The scalability of ML attacks poses a significant challenge for the implementation of PUFs in the future. As ML techniques continue to advance and become more accessible, the likelihood of successful attacks decreases. This trend indicates that PUF security should not rely solely on computational complexity, but rather focus on fundamental information-theoretic principles.

Although the challenge restriction approach is theoretically sound, it severely limits the potential of strong PUFs and thus may not be practical for applications requiring frequent authentication or key generation [16, 21]. By making strong PUFs as ineffective as weak PUFs, this approach negates many of the initial motivations for their development.

The error correction requirements present another significant challenge. Study [22] reveals that auxiliary data is vulnerable, suggesting that seemingly secure error correction approaches may introduce vulnerabilities. When selecting error correction codes, it is essential to consider traditional factors such as correction capability and implementation complexity, as well as security against ML attacks through the analysis of auxiliary data.

Analysis of PUF-based authentication and key exchange protocols reveals varying degrees of vulnerability to ML attacks [21, 23]. Protocols based on direct CRP exchange are particularly vulnerable to simulation attacks when there is a large number of CRPs. The security of these protocols primarily relies on PUF resilience against ML attacks rather than on protocol-level protections.

Recent advancements in protocol design aim to minimize the information available to intruders while maintaining functional requirements [23]. Methods such as secure multi-party computing, homomorphic encryption and zero-knowledge proofs enable the use of PUF without disclosing CRP to potential attackers. However, these approaches often incur significant computational costs, which may be impractical for devices with limited resources. More complex protocols use challenge obfuscation, response masking, and temporary security mechanisms to restrict intruders' capabilities [20, 21]. Nevertheless, many of these approaches have proven inadequate against determined opponents with access to modern ML techniques. The main challenge lies in balancing security with practical constraints such as communication costs, computational requirements, and error tolerance.

Integrating PUFs into complex systems creates additional opportunities for attacks through side-channel analysis to introduce malfunctions and vulnerabilities at the system level [21]. As PUFs become more widely adopted in mission-critical applications, the incentive for sophisticated attacks increases, necessitating more robust security analyses and protection mechanisms. The rapid development of ML methods poses ongoing challenges for PUF security. Emerging techniques such as meta-learning, transfer learning, and multitask learning have the potential to create attack strategies that can overwhelm existing defenses [20, 23].

In order to facilitate fair comparisons between different attack designs and techniques, it is essential to develop standardized methodologies for evaluating the security of PUFs [17, 21]. Future PUF research should focus on developing architectures with provably secure properties rather than simply relying on empirical resistance to current attack methods [16, 18]. Due to the use of different datasets, attack parameters, and success criteria in existing assessments, it is difficult to make an objective comparison of the relative security levels of different systems. Information-theoretic approaches capable of ensuring security even against unlimited computational resources provide a sounder basis for long-term security.

The dynamics of the evolution of trusted design systems suggest that in order to achieve long-term security, it may be necessary to transition from approaches based on incrementally improving current architectures to a fundamentally different approach that leverages information-theoretic security principles [24, 25]. The findings of this study emphasize the significance of careful threat modeling, conservative security assumptions, and robust protection strategies for professionals implementing PUF-based systems [26–29].

PUF IMPLEMENTATION EXAMPLES

Depending on the level of protection required by the embedded chip in the device (weak or strong), PUF modules are used in a variety of applications to safeguard devices. PUF responses can be used directly for authentication in a manner similar to biometric verification. As demonstrated in the first part of this publication series [1], the cutoff value used to determine positive authentication relies on intra-HD and inter-HD histograms. Typically, when the histograms overlap, the cutoff value is set by balancing the likelihood of

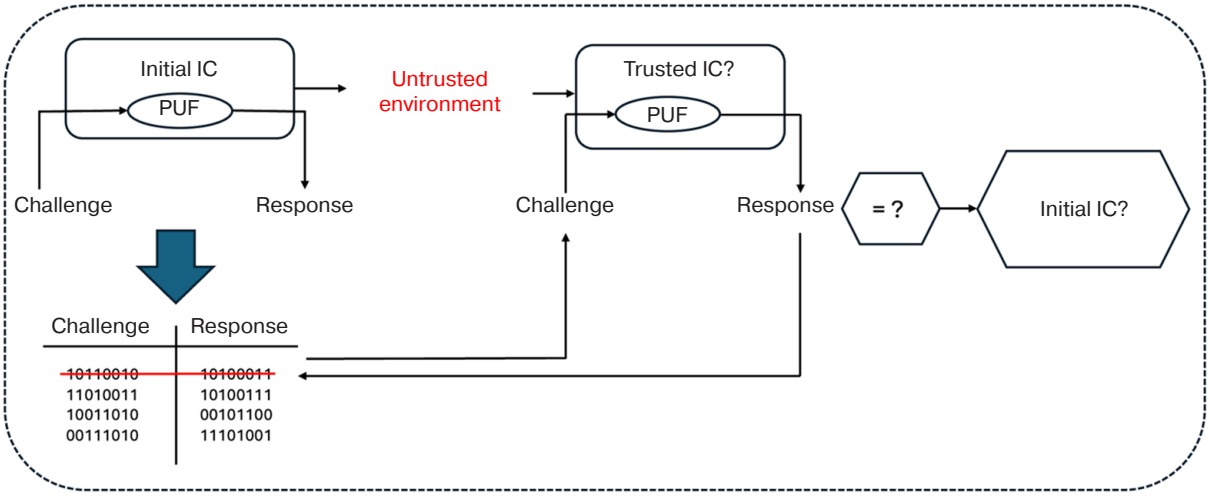


Fig. 8. PUF application in authentication protocol [40]

false positives (false acceptance rate, FAR) and false negatives (false rejection rate, FRR) using an approach similar to those employed in metrology. The optimal choice, which minimizes the sum of FARs and FRRs, is achieved by setting a threshold at the intersection of the two histograms. However, other trade-offs may be necessary for specific applications. Here it should also be noted that unique identification is only possible with a high degree of certainty if the response contains sufficient entropy relative to the sample size.

PUFs are used for the following:

1) the mutual authentication of IoT devices [30, 31];
2) the identification of message flows involving authentication requests from mobile unmanned aerial vehicles equipped with a PUF and ground stations [32, 33];
3) the communication between devices, sensors, and a health monitoring system in which the server is also equipped with an appropriate PUF, and a secure database is used to store the collected variants of CRPs [34];
4) the protection and safety of vehicles [35, 36];
5) the security and confidentiality of data transmission in networks [37–39].

Device identification

At registration, the CRP from each PUF is stored in the database along with the identification data of the physical system in which it is built. As outlined in [1], during the identification process, the verifier compares the random CRP pair with the PUF responses for the presented system stored in the database. If the observed response is close enough to the response in the database, authentication succeeds; otherwise, it fails. To prevent repeated attacks, each CRP should only be used once for each PUF instance and removed from the database after identification (Figs. 8 and [40]).

Generating encryption keys

Since the PUF is based on randomness due to technological tolerances, no programming is required when generating or storing secret keys based on the PUF. Due to this randomness being fixed in the microscopic physical details of the chip, the key remains unchanged and can be reproduced several times. This eliminates the need for non-volatile key memory, providing additional protection against third-party channel attacks.

However, since PUF responses are typically noisy, an intermediate processing step is necessary to extract the cryptographic key. This problem, which is known in information theory as extracting an encryption key from a noisy signal, is typically solved using a two-stage algorithm. The PUF request during the first generation stage involves the algorithmic generation of a secret key and some additional auxiliary data. This key and the auxiliary data are then stored in a secure, device-independent database. During the reproduction stage, the auxiliary data is provided to the algorithm, which uses it to extract the same key from the PUF created during the generation stage. Such algorithms can be designed in such a way that the key remains top secret even if the auxiliary data are transmitted openly. To ensure the reliability of this method of key generation, special troubleshooting methods are employed [41, 42], such as the generation (Gen) and playback (Rep) algorithms. These algorithms ensure the extraction of stable, reproducible information from PUF responses by comparing two messages composed of noisy, encrypted, random data with unclassified auxiliary data

attached for identification purposes. Practical examples of these algorithms are provided in [43–45].

The use of PUF allows for the implementation of hardware-based cryptography as a special method. In this method, the digital cipher key is not stored in memory, while the secret element consists in the unique behavior of the PUF instance in the embedded device. This significantly hinders attempts by attackers to use non-volatile memory as a means of obtaining useful information. Since the PUF can also be used to detect unauthorized access to the key store, device-entangled cryptography is closely related to proven physical security (see, for example, [46]).

Intellectual property protection

Protecting the intellectual property of chips is a crucial issue for semiconductor companies due to numerous security threats that can result in financial losses. These threats include forgery, cloning, reverse engineering, and reliance on substandard components. Studies in this area include works on the hardware protection of SRAM PUF blocks on programmable logic ICs and methods of preventing the copying of IP addresses to protect against unauthorized access to firmware, based on PUF and neural network models [47–49].

Random number generation

Silicon PUFs are used as a source for generating random numbers, which are necessary for cryptographic systems. Exemplary studies in this domain include [50, 51], wherein the authors employ PUF responses to generate the primary data for random number generation.

Payment protection

In [52, 53], PUF responses are used in authentication bit strings, encryption keys, and electronic cash token generation (PUF-Cash). The aim is to develop an application architecture that can be used in electronic payment schemes. It also ensures the anonymity of user identities for organizations such as banks and sellers. It is proposed in [54] that private keys, secure communication and data authentication be provided by equipping credit/debit cards with built-in PUF chips.

Protection of memory and software integrity, ensuring secure communication

Currently, a number of companies and research centers specialize in developing special measures to increase the power of equipment and software. A typical example is the technical documentation of an association [54], which was created to develop, define and promote open, vendor-independent, global industry standards. These standards support a hardware trust framework for interchangeable trusted computing platforms. The specification of this framework refers to large-scale security threats arising from geopolitical and data sovereignty issues, which threaten to slow down the adoption and growth of the IoT industry. This in turn stimulates the need to create reliable supply chain ecosystems in Asia, Europe, and the Americas. Here, the importance of the keys used for digital signing and verification in ensuring the security of the entire system, along with cryptographic functions that provide a secure process for loading operating systems, is emphasized. The main and only effective means of ensuring safe operation is recognized as the use of microcircuits equipped with a PUF, in particular, a new type developed by eMemory Technology Inc.[18] The NeoPUF technology developed by this company exploits differences in the tunnel effect of the oxide layer to achieve high PUF performance (inter-HD = 50%; intra-HD ~ 0%).[19]

An example of using PUF to protect the confidentiality and integrity of instructions and data in memory from physical and software attacks is given in [38, 55].

Software licensing is required to protect against unauthorized modifications and usage on unauthorized platforms. The idea behind using the PUF is that critical operations such as starting or restarting the system are performed using the generated keys, while the software interacts with the PUF [56]. A software licensing mechanism based on PUF is proposed in [57]. The user's computer is equipped with a PUF based on the SRAM scheme to provide it with a unique identification. When a user needs to purchase the necessary software, the company establishes a connection to the user's computer to obtain the PUF output and make it available as a license in the software. During installation by the client, an authentication process occurs between the software and the personal computer. The built-in license is then compared with the PUF output to ensure that the software instance is only executed on a specific device.

Recently, a new patented PUF approach called equipotential timing has been proposed by Granite

[18] eMemory Technology Inc. https://www.ememory.com.tw/. Accessed July 19, 2025.

[19] PUFsecurity. NeoPUF® – A Reliable and Non-Traceable Quantum Tunneling PUF. https://www.pufsecurity.com/document/neopuf-a-reliable-and-non-traceable-quantum-tunneling-puf/. Accessed July 19, 2025.

Mountain Technologies.[20] This approach uses gigabyte PUF (Giga-PUF) to design features that cannot be physically disabled. The equipotential timing approach provides stable, synthesized PUF implementations that can be implemented as soft IP blocks and integrated into any design at a low cost. Such Giga-PUFs can be quickly and easily scaled to any circuit, even across technology nodes and silicon fabricators. This gives all companies access to PUF exponential solutions that provide true trust in the hardware, thus enabling them to protect their products.

Another application of the PUF modules included in the IC is to enable secure communication for the authentication of IoT devices in the key exchange protocol [58–60].

CONCLUSIONS

As confirmed in this review, analog and passive PUFs represent an important class of hardware security primitive that complement latency- and memory-based solutions. While analog circuits made from transistors and diodes offer high entropy and low response power consumption, careful stabilization and calibration are required to counteract the effects of PVT factors and ageing.

Passive approaches, such as resistive fingerprints of the power grid Via PUF and Coating PUF, are attractive due to their minimal overhead, high stability and counterfeiting complexity.

For practical application, the following are recommended:

- use of internal compensation and auto-zero mechanisms, as well as rejection/masking of unstable elements;
- single standardized digitization chain (sensor—amplifier/comparator—code post-processing);
- integration with light error correction techniques or phase filtration where justified;
- assessment of resistance to simulation attacks and side channels, taking into account analog specifics (temperature, power supply, and noise injection).

The development of ML has revealed the vulnerability of many classic PUFs, particularly in linear delay models. The insufficient design complexity (e.g., XOR and cascades) results in leaks of auxiliary or side data to enable attacks without direct access to the responses.

Defense mechanisms demonstrate varying degrees of success, while architectural innovations such as the CRC-PUF offer increased resistance to ML attacks. However, the fundamental issue remains that most contemporary PUF designs are based on mathematical models that can be trained using complex algorithms.

When selecting suitable PUF architectures, error correction codes and operational protocols, it is important to consider not only current attack methods, but also potential future advances in ML techniques. As this area evolves, it will be essential to integrate PUF security research with broader developments in ML, cryptography, and hardware security in order to develop reliable solutions. Here, the ultimate goal is to develop truly unclonable functions that retain their protective properties even in the face of highly sophisticated computing attacks.

PUFs can be used as important building blocks in authentication systems, particularly in hardware tokens with limited resources and IoT systems (see the services of eMemory Technology Inc. in particular).

Promising future research directions include the following:

1) standardizing and developing algorithms and analysis tools, and increasing energy efficiency;
2) improving security and protection against attacks by analyzing side channels and ML attacks;
3) reducing environmental impact;
4) applying new special algorithms, particularly controlled and reconfigurable PUFs;
5) developing PUFs based on promising physical effects, particularly quantum PUFs utilizing nuclear magnetic moments, resonant tunnel diodes, plasmon and lanthanide luminescence, Josephson transitions, changes in electron flow in nanocells (Aaron–Bohm effect), and quantum entanglement.

[20] Granite Mountain Technologies. Physical Unclonable Functions. https://gmt-semi.com/solutions/puf. Accessed July 19, 2025.

ACKNOWLEDGMENTS

This work was supported by the Ministry of Science and Higher Education of the Russian Federation (State task for universities No. FSFZ-2026-0003) and using the equipment of the Center for Collective Use of RTU MIREA (agreement dated September 01, 2021, No. 075-15-2021-689, unique identification number 2296.61321X0010).

Authors' contributions

E.Ph. Pevtsov—study conceptualization, review outline and structure, and manuscript writing.
T.A. Demenkova—study conceptualization, review outline and structure, and synthesis of the results.
M.I. Maleto—literature analysis and systematization, synthesis of the results.
A.S. Sigov—scientific consulting, scientific editing, and final approval of the manuscript.
Yu.A. Korotaev—literature analysis and systematization, manuscript writing, and synthesis of the results.
N.D. Evgenev—literature analysis and systematization, manuscript writing, and synthesis of the results.
All authors have read and approved the published version of the manuscript.

REFERENCES

1. Pevtsov E.Ph., Demenkova T.A., Korotaev Yu.A., Sigov A.S. Physically unclonable functions in digital integrated circuits. *Russian Technological Journal.* 2026;14(2):80–102. https://doi.org/10.32362/2500-316X-2026-14-2-80-102
2. Lofstrom K., Daasch R., Taylor D. IC identification circuit using device mismatch. In: *Proceedings of the IEEE International Solid-State Circuits Conference* (*ISSCC 2000*). February 7–9, 2000. San Francisco, CA, USA. Piscataway, NJ: IEEE; 2000. P. 372–373. https://doi.org/10.1109/ISSCC.2000.839821
3. Venkatesh A., Sanyal A. A machine learning resistant strong PUF using subthreshold voltage divider array in 65nm CMOS. In: *Proceedings of the 2019 IEEE International Symposium on Circuits and Systems* (*ISCAS 2019*). May 26–29, 2019. Sapporo, Japan. Piscataway, NJ: IEEE; 2019. P. 1–5. https://doi.org/10.1109/ISCAS.2019.8702525

4. Mitchell-Moreno J.H., Espinosa Flores-Verdad G. A low bit instability CMOS PUF based on current mirrors and WTA cells. *J. Electron. Test.* 2023;39:611–620. https://doi.org/10.1007/s10836-023-06085-4
5. Jadhav V.D., Kalloor R., Poola L., Prabhakar T.V. Diode-PUF for intelligent electronic devices. In: *Proceedings of the 16th International Conference on Communication Systems & Networks* (*COMSNETS 2024*). January 2–6, 2024. Bengaluru, India. Piscataway, NJ: IEEE; 2024. P. 330–332. https://doi.org/10.1109/COMSNETS59351.2024.10427169
6. Kim N., Jeon S.-B., Jang B. Hardware-intrinsic physical unclonable functions by harnessing nonlinear conductance variation in oxide semiconductor-based diode. *Nanomaterials* (*Basel*). 2023;13(4):675. https://doi.org/10.3390/nano13040675
7. Takahashi Y., Koyasu H., Kumar S.D., et al. Quasi-adiabatic SRAM based silicon physical unclonable function. *SN Comput. Sci.* 2020;1:237. https://doi.org/10.1007/s42979-020-00253-5
8. Liu J., Takahashi Y. Design of low-power 6T adiabatic PUF circuit. In: *Proceedings of the 2024 IEEE Asia Pacific Conference on Circuits and Systems* (*APCCAS 2024*). October 27–30, 2024. Taipei, Taiwan. Piscataway, NJ: IEEE; 2024. P. 599–603. https://doi.org/10.1109/APCCAS62602.2024.10808318
9. Nagata S., Takahashi Y. A design of PUF circuit using adiabatic logic. In: *Proceedings of the 2024 IEEE Asia Pacific Conference on Circuits and Systems* (*APCCAS 2024*). October 27–30, 2024. Taipei, Taiwan. Piscataway, NJ: IEEE; 2024. P. 595–598. https://doi.org/10.1109/APCCAS62602.2024.10808900
10. Helinski R., Acharyya D., Plusquellic J. A physical unclonable function defined using power distribution system equivalent resistance variations. In: *Proceedings of the 46th ACM/IEEE Design Automation Conference* (*DAC 2009*). July 26–31, 2009. San Francisco, CA, USA. New York: ACM; 2009. P. 676–681. https://doi.org/10.1145/1629911.1630089
11. Jeon D., Baek J.H., Kim Y.-D., Lee J., Kim D.K., Choi B.-D. A physical unclonable function with bit error rate $<2.3 \times 10^{-8}$ based on contact formation probability without error correction code. *IEEE J. Solid-State Circuits*. 2020;55(3):805–816. https://doi.org/10.1109/JSSC.2019.2951415
12. Csaba G., Ju X., Chen Q., Porod W., Schmidhuber J., Schlichtmann U., Lugli P., Rührmair U. On-chip electric waves: an analog circuit approach to physical uncloneable functions [preprint]. *IACR Cryptology ePrint Archive*. 2009;2009/246.
13. Tuyls P., Schrijen G-J., Škorić B., van Geloven J., Verhaegh N., Wolters R. Read-proof hardware from protective coatings. In: Goubin L., Matsui M. (Eds.). *Cryptographic Hardware and Embedded Systems*. *CHES 2006,* Yokohama, Japan, October 10–13, 2006. Book Series: Lecture Notes in Computer Science. Berlin: Springer; 2006. V. 4249. P. 369–383. https://doi.org/10.1007/11894063_29
14. Skoric B., Maubach S., Kevenaar T., Tuyls P. Information-theoretic analysis of coating PUFs [preprint]. *IACR Cryptology ePrint Archive*. 2006;2006/101.
15. Aysu A., Farhady Ghalaty N., Franklin Z., Yali M., Schaumont P. Digital fingerprints for low-cost platforms using MEMS sensors. In: *Proceedings of the Workshop on Embedded Systems Security* (*WESS ’13*). September 29, 2013. Montreal, QC, Canada. New York: ACM; 2013. Article 2. P. 1–6. https://doi.org/10.1145/2527317.2527319
16. Yu M.D., M’Raihi D., Sowell R., Devadas S. Lightweight and secure PUF key storage using limits of machine learning. In: Preneel B., Takagi T. (Eds.). *Cryptographic Hardware and Embedded Systems. CHES 2011.* Book Series: Lecture Notes in Computer Science. Berlin Heidelberg: Springer; 2011. V. 6917. P. 358–373. https://doi.org/10.1007/978-3-642-23951-9_24
17. Saadvikaa N., Saketi K.J., Gopishetti A., et al. PUF modeling attacks using deep learning and machine learning algorithms. *Eng. Proceedings.* 2023;56(1):187. https://doi.org/10.3390/ASEC2023-15948
18. Dubrova E., Näslund O., Degen B., et al. CRC-PUF: A machine learning attack resistant lightweight PUF construction. In: *2019 IEEE European symposium on security and privacy workshops* (*EuroS&PW*). IEEE; 2019. P. 264–271. https://doi.org/10.1109/EuroSPW.2019.00036
19. Tripathy S., Rai V.K., Mathew J. MARPUF: physical unclonable function with improved machine learning attack resistance. *IET Circuits, Devices & Systems*. 2021;15(5):465–474. https://doi.org/10.1049/cds2.12042
20. Ebrahimabadi M., Lalouani W., Younis M., et al. Countering PUF modeling attacks through adversarial machine learning. In: *2021 IEEE Computer Society Annual Symposium on VLSI* (*ISVLSI*). IEEE. 2021. P. 356–361. https://doi.org/10.1109/ISVLSI51109.2021.00071
21. Khalfaoui S., Leneutre J., Villard A., et al. Security analysis of machine learning-based PUF enrollment protocols: A review. *Sensors*. 2021;21(24):8415. https://doi.org/10.3390/s21248415

22. Strieder E., Frisch C., Pehl M. Machine learning of physical unclonable functions using helper data: Revealing a pitfall in the fuzzy commitment scheme. *IACR Transactions on Cryptographic Hardware and Embedded Systems*. 2021;2:1–36. https://doi.org/10.46586/tches.v2021.i2.1-36
23. Ali-Pour A., Afghah F., Hely D., et al. Secure PUF-based authentication and key exchange protocol using machine learning. In: *2022 IEEE Computer Society Annual Symposium on VLSI* (*ISVLSI*). IEEE. 2022. P. 386–389. https://doi.org/10.1109/ISVLSI54635.2022.00086
24. Yadav A., Kumar S., Singh J. A review of physical unclonable functions (PUFs) and its applications in IoT environment. In: Hu Y.C., Tiwari S., Trivedi M.C., Mishra K.K. (Eds.). *Ambient Communications and Computer Systems.* Book Series: Lecture Notes in Networks and Systems. Singapore: Springer; 2022. V. 356. P. 1–3. https://doi.org/10.1007/978-981-16-7952-0_1
25. Gao Y., Al-Sarawi S.F., Abbott D. Physical unclonable functions. *Nat. Electron*. 2020;3(2):81–91. https://doi.org/10.1038/s41928-020-0372-5
26. Wisiol N., Mühl C., Pirnay N., et al. Splitting the interpose PUF: A novel modeling attack strategy. *IACR Transactions on Cryptographic Hardware and Embedded Systems.* 2020;3:97–120. https://doi.org/10.13154/tches.v2020.i3.97-120
27. Arapinis M., Delavar M., Doosti M., et al. Quantum physical unclonable functions: Possibilities and impossibilities. *Quantum*. 2021;5:475. https://doi.org/10.22331/q-2021-06-15-475
28. Kayaci N., Ozdemir R., Kalay M., et al. Organic light-emitting physically unclonable functions. *Adv. Funct. Mater.* 2022;32(14):2108675. https://doi.org/10.1002/adfm.202108675
29. Awano H., Iizuka T., Ikeda M. PUFNet: A deep neural network based modeling attack for physically unclonable function. In: *2019 IEEE International Symposium on Circuits and Systems* (*ISCAS*). IEEE. 2019. P. 1–4. https://doi.org/10.1109/ISCAS.2019.8702431
30. Idriss T.A., Idriss H.A., Bayoumi M.A. A lightweight PUF-based authentication protocol using secret pattern recognition for constrained IoT devices. *IEEE Access*. 2021;9:80546–80558. https://doi.org/10.1109/ACCESS.2021.3084903
31. Shah A., Pandya H., Soni M., Karimov A., Maaliw R.R., Keshta I. PUF-based lightweight authentication protocol for IoT devices. In: Balas V.E., Semwal V.B., Khandare A. (Eds.). *Intelligent Computing and Networking. IC-ICN 2023*. Book Series: Lecture Notes in Networks and Systems. Singapore: Springer; 2023. V. 699. P. 401–412. https://doi.org/10.1007/978-981-99-3177-4_29
32. Alladi T., Deo M., Chamola V., Sikdar B., Chao H.C. SecAuthUAV: a novel authentication scheme for UAV-ground station and UAV-UAV communication. *IEEE Trans. Veh. Technol.* 2020;69(12):15068–15077. https://doi.org/10.1109/TVT.2020.3033060
33. Bansal G., Sikdar B. S-MAPS: scalable mutual authentication protocol for dynamic UAV swarms. *IEEE Trans. Veh. Technol.* 2021;70(11):12088–12100. https://doi.org/10.1109/TVT.2021.3116163
34. Yanambaka V.P., Mohanty S.P., Kougianos E., Puthal D. PMsec: physical unclonable function-based robust and lightweight authentication in the Internet of Medical Things. *IEEE Trans. Consum. Electron.* 2019;65(3):388–397. https://doi.org/10.1109/TCE.2019.2926192
35. Jiang Q., Zhang X., Zhang N., et al. Three-factor authentication protocol using physical unclonable function for IoV. *Comput. Commun.* 2021;173:45–55. https://doi.org/10.1016/j.comcom.2021.03.022
36. Mershad K., Cheikhrouhou O., Ismail L. Proof of accumulated trust: a new consensus protocol for the security of the IoV. *Veh. Commun.* 2021;32:100392. https://doi.org/10.1016/j.vehcom.2021.100392
37. Kaveh M., Aghapour S., Martín D., Mosavi M.R. A secure lightweight signcryption scheme for smart grid communications using reliable physically unclonable function. In: *2020 IEEE International Conference on Environment and Electrical Engineering & 2020 IEEE Industrial and Commercial Power Systems Europe* (*EEEIC/I&CPS Europe*). June 9–12, 2020. Madrid, Spain. Piscataway. NJ: IEEE; 2020. P. 1–6. https://doi.org/10.1109/EEEIC/ICPSEurope49358.2020.9160596
38. Cao Y.-N., Wang Y., Ding Y., Zheng H., Guan Z., Wang H. A PUF-based lightweight authenticated metering data collection scheme with privacy protection in smart grid. In: *2021 IEEE International Conference on Parallel & Distributed Processing with Applications, Big Data & Cloud Computing, Sustainable Computing & Communications, Social Computing & Networking* (*ISPA/BDCloud/SocialCom/SustainCom*). August 30 – September 3, 2021. New York, USA. Piscataway, NJ: IEEE; 2021. P. 876–883. https://doi.org/10.1109/ISPA-BDCloud-SocialCom-SustainCom52081.2021.00124
39. Maqsooq B., Qadri S., Shamshad S., Ayub M.F., Mahmood K., Kumar N. An identity-based authentication protocol for smart grid environment using physical unclonable function. *IEEE Trans. Smart Grid*. 2021;12(5):4426–4434. https://doi.org/10.1109/TSG.2021.3072244
40. Zerrouki F., Ouchani S., Bouarfa H. PUF-based mutual authentication and session key establishment protocol for IoT devices. *J. Ambient Intell. Human. Comput.* 2023;14:12575–12593. https://doi.org/10.1007/s12652-022-04321-x

41. Müelich S., Bossert M. *A New Error Correction Scheme for Physical unclonable Functions. arXiv*. arXiv:1611.01960 [cs.CR]. 2016. https://doi.org/10.48550/arXiv.1611.01960
42. Shamsoshoara A., Korenda A.R., Afghah F., Zeadally S. *A Survey on Hardware-Based Security Mechanisms for Internet of Things. arXiv*. arXiv:1907.12525 [cs.CR]. 2019. https://doi.org/10.48550/arXiv.1907.12525
43. Maes R., Verbauwhede I. Physically Unclonable Functions: A Study on the State of the Art and Future Research Directions. In: Sadeghi A.-R., Naccache D. *Towards Hardware-Intrinsic Security*: *Foundations and Practice*. Book series: Information Security and Cryptography. Berlin: Springer; 2010. P. 3–37. https://doi.org/10.1007/978-3-642-14452-3_1
44. Dodis Y., Ostrovsky R., Reyzin L., Smith A. Fuzzy extractors: how to generate strong keys from biometrics and other noisy data. *SIAM J. Comput.* 2008;38(1):97–139. https://doi.org/10.1137/060651380
45. Muthammal R., Sindhuja N. VLSI architecture of turbo codes for dedicated short-range communication. *Int. J. Eng. Res. Online.* 2015;3(5):412–416. Available from URL: https://www.researchgate.net/publication/321669464_VLSI_Architecture_of_Turbo_Codes-IP_Secure_With_PUF_for_DSRC_systems. Accessed July 19, 2025.
46. Wong C.-W., Wu M. Counterfeit detection using paper PUF and mobile cameras. In: *Proceedings of the IEEE International Workshop on Information Forensics and Security* (*WIFS 2015*). November 16–19, 2015. Rome, Italy. Piscataway, NJ: IEEE; 2015. P. 1–6. https://doi.org/10.1109/WIFS.2015.7368579
47. Zheng J., Potkonjak M. A digital PUF-based IP protection architecture for network embedded systems. In: *Proceedings of the Tenth ACM/IEEE Symposium on Architectures for Networking and Communications Systems* (*ANCS'14*). 2014. P. 255–256. https://doi.org/10.1145/2658260.2661776
48. Zhang J., Lin Y., Lyu Y., Qu G. A PUF-FSM binding scheme for FPGA IP protection and pay-per-device licensing. *IEEE Trans. Inf. Forensics Secur.* 2015;10(6):1137–1150. https://doi.org/10.1109/TIFS.2015.2400413
49. Guo Q., Gong Y., Hu Y., Li X.-W. PUF-based pay-per-device scheme for IP protection of CNN model. In: *2018 IEEE Asian Test Symposium* (*ATS 2018*). December 10–13, 2018. Hefei, China. Piscataway, NJ: IEEE; 2018. P. 115–120. https://doi.org/10.1109/ATS.2018.00032
50. Kalanadhabhatta S., Kumar D., Anumandla K.K., Reddy A., Acharyya A. PUF-based secure chaotic random number generator design methodology. *IEEE Trans. Very Large Scale Integr.* (*VLSI*) *Syst.* 2020;28(9):1994–2004. https://doi.org/10.1109/TVLSI.2020.2979269
51. Kaya T. A true random number generator based on a Chua and RO-PUF: design, implementation and statistical analysis. *Analog Integr. Circ. Sig. Process.* 2020;102:577–588. https://doi.org/10.1007/s10470-019-01474-2
52. Calhoun J., Minwalla C., Helmich C., Saqib F., Che W., Plusquellic J. Physical Unclonable Function (PUF)-based e-cash transaction protocol (PUF-Cash). *Cryptography*. 2019;3(3):18. https://doi.org/10.3390/CRYPTOGRAPHY3030018
53. Zhang Y., Qin Y., Feng D., Yang B., Wang W. An efficient Trustzone-based in-application isolation schema for mobile authenticators. In: Lin X., Ghorbani A., Ren K., Zhu S., Zhang A. (Eds.). *Security and Privacy in Communication Networks. SecureComm 2017.* Book Series: Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering. Cham: Springer; 2018. V. 238. P. 585–605. https://doi.org/10.1007/978-3-319-78813-5_30
54. Kish L.B., Entesari K., Granqvist C.G., Kwan C. Unconditionally secure credit/debit card chip scheme and physical unclonable function. *Fluctuation Noise Lett.* 2017;16(1):1750002. https://doi.org/10.1142/S021947751750002X
55. Suh G.E., O'Donnell C., Devadas S. Aegis: a single-chip secure processor. *IEEE Des. Test Comput.* 2007;24(6):570–580. https://doi.org/10.1109/MDT.2007.179
56. Suresh V., Manimegalai R. SPIC-SRAM PUF integrated chip based software licensing model. In: Thampi S., Madria S., Wang G., Rawat D., Alcaraz Calero J. (Eds.). *Security in Computing and Communications. SSCC 2018. Communications in Computer and Information Science*. Springer; 2018. V. 969. P. 377–388. https://doi.org/10.1007/978-981-13-5826-5_29
57. Kohnhäuser F., Schaller A., Katzenbeisser S. PUF-based software protection for low-end embedded devices. In: Conti M., Schunter M., Askoxylakis I. (Eds.). *Trust and Trustworthy Computing. Trust 2015.* Book Series: Lecture Notes in Computer Science. Cham: Springer; 2015. V. 9229. P. 3–21. https://doi.org/10.1007/978-3-319-22846-4_1
58. Zheng Y., Liu W., Gu C., Chang C-H. *PUF-based Mutual Authentication and Key-Exchange Protocol for Peer-to-Peer IoT Applications* [preprint]. *TechRxiv*; 2021.
59. Mahmood K., Shamshad S., Rana M., et al. PUF-enabled lightweight key-exchange and mutual authentication protocol for multi-server-based D2D communication. *J. Inf. Secur. Appl.* 2021;61:102900. https://doi.org/10.1016/j.jisa.2021.102900

60. Bathalapalli V.K.V.V., Mohanty S.P., Pan C., Kougianos E. QPUF: quantum physical unclonable functions for security-by-design of industrial Internet-of-Things. In: *2023 IEEE International Symposium on Smart Electronic Systems* (*iSES 2023*). December 18–20, 2023. Hyderabad, India. Piscataway, NJ: IEEE; 2023. P. 296–301. https://doi.org/10.1109/iSES58672.2023.00067

About the Authors

Evgenii Ph. Pevtsov, Cand. Sci. (Eng.), Director of Center for the Design of Integrated Circuits, Nanoelectronics Devices and Microsystems, MIREA – Russian Technological University (78, Vernadskogo pr., Moscow, 119454 Russia). E-mail: pevtsov@mirea.ru. Scopus Author ID 6602652601, ResearcherID M-2709-2016, RSCI SPIN-code 1410-2483, http://orcid.org/0000-0001-6264-1231

Tatyana A. Demenkova, Cand. Sci. (Eng.), Associate Professor, Computer Technology Department, Institute of Information Technologies, MIREA – Russian Technological University (78, Vernadskogo pr., Moscow, 119454 Russia). E-mail: demenkova@mirea.ru. Scopus Author ID 57192958412, ResearcherID AAB-3937-2020, RSCI SPIN-code 3424-7489, http://orcid.org/0000-0003-3519-6683

Mikhail I. Maleto, Cand. Sci. (Eng.), Center for the Design of Integrated Circuits, Nanoelectronics Devices and Microsystems, MIREA – Russian Technological University (78, Vernadskogo pr., Moscow, 119454 Russia). E-mail: maleto@mirea.ru. RSCI SPIN-code 2958-3989, http://orcid.org/0009-0006-3603-6322

Alexander S. Sigov, Academician at the Russian Academy of Sciences, Dr. Sci. (Phys.–Math.), Professor, President, MIREA – Russian Technological University (78, Vernadskogo pr., Moscow, 119454 Russia). E-mail: sigov@mirea.ru. Scopus Author ID 35557510600, ResearcherID L-4103-2017, RSCI SPIN-code 2869-5663, https://www.researchgate.net/profile/A_Sigov

Yuri A. Korotaev, Postgraduate Student, Department of Nanoelectronics, Institute for Advanced Technologies and Industrial Programming, MIREA – Russian Technological University (78, Vernadskogo pr., Moscow, 119454 Russia). E-mail: korotaevyua@yandex.ru. RSCI SPIN-code 7428-6831, https://orcid.org/0009-0000-3976-7872

Nikita D. Evgenev, Student, Computer Technology Department, Institute of Information Technologies, MIREA – Russian Technological University (78, Vernadskogo pr., Moscow, 119454 Russia). E-mail: nikita.evgenev.10@gmail.com. RSCI SPIN-code 1034-0447, https://orcid.org/0009-0006-9073-8798

Об авторах

Певцов Евгений Филиппович, к.т.н., директор структурного подразделения «Центр проектирования интегральных схем, устройств наноэлектроники и микросистем», ФГБОУ ВО «МИРЭА – Российский технологический университет» (119454, Россия, Москва, пр-т Вернадского, д. 78). E-mail: pevtsov@mirea.ru. Scopus Author ID 6602652601, ResearcherID M-2709-2016, SPIN-код РИНЦ 1410-2483, https://orcid.org/0000-0001-6264-1231

Деменкова Татьяна Александровна, к.т.н., доцент, кафедра вычислительной техники, Институт информационных технологий, ФГБОУ ВО «МИРЭА – Российский технологический университет» (119454, Россия, Москва, пр-т Вернадского, д. 78). E-mail: demenkova@mirea.ru. Scopus Author ID 57192958412, ResearcherID AAB-3937-2020, SPIN-код РИНЦ 3424-7489, https://orcid.org/0000-0003-3519-6683

Малето Михаил Иванович, к.т.н., ведущий инженер структурного подразделения «Центр проектирования интегральных схем, устройств наноэлектроники и микросистем», ФГБОУ ВО «МИРЭА – Российский технологический университет» (119454, Россия, Москва, пр-т Вернадского, д. 78). E-mail: maleto@yandex.ru. SPIN-код РИНЦ 2958-3989, http://orcid.org/0009-0006-3603-6322

Сигов Александр Сергеевич, академик Российской академии наук, д.ф.-м.н., профессор, президент ФГБОУ ВО «МИРЭА – Российский технологический университет» (119454, Россия, Москва, пр-т Вернадского, д. 78). E-mail: sigov@mirea.ru. Scopus Author ID 35557510600, ResearcherID L-4103-2017, SPIN-код РИНЦ, 2869-5663, www.researchgate.net/profile/A_Sigov

Коротаев Юрий Александрович, аспирант, кафедра наноэлектроники, Институт перспективных технологий и индустриального программирования, ФГБОУ ВО «МИРЭА – Российский технологический университет» (119454, Россия, Москва, пр-т Вернадского, д. 78). E-mail: korotaevyua@yandex.ru. SPIN-код РИНЦ 7428-6831, https://orcid.org/0009-0000-3976-7872

Евгеньев Никита Давидович, студент, кафедра вычислительной техники, Институт информационных технологий, ФГБОУ ВО «МИРЭА – Российский технологический университет» (119454, Россия, Москва, пр-т Вернадского, д. 78). E-mail: nikita.evgenev.10@gmail.com. SPIN-код РИНЦ 1034-0447, https://orcid.org/0009-0006-9073-8798

Translated from Russian into English by K. Nazarov
Edited for English language and spelling by Thomas A. Beavitt

Analytical instrument engineering and technology

Аналитическое приборостроение и технологии

UDC 537.553.3; 681.7
https://doi.org/10.32362/2500-316X-2026-14-3-106-114
EDN RKATXE

RESEARCH ARTICLE

Optimal electrode design for microminiature electronic optics

Pavel S. Kuznetsov [1, @], Anton O. Sinelnikov [2]

[1] *State Scientific Research Institute of Instrument Engineering, Moscow, 129226, Russia*
[2] *Peoples' Friendship University of Russia (RUDN University), Moscow, 117198 Russia*
[@] *Corresponding author, e-mail: ps_kuznetsov@mail.ru*

• **Submitted:** 12.09.2025 • **Revised:** 15.11.2025 • **Accepted:** 27.03.2026

Abstract

Objectives. The work set out to systematically analyze and optimize the overall design and technological characteristics of microminiature electron-optical systems for achieving maximum performance indicators. The study paid special attention to establishing relationships between the geometric parameters of the system and its functional characteristics.

Methods. The research is based on comprehensive mathematical modeling of electron dynamics in a complex five-electrode scheme that accurately reproduces the actual design of a compact electron-beam microcolumn. This approach was used to establish the quantitative dependencies of resolution and electron beam intensity critical system performance indicators on fundamental geometric parameters: interelectrode distances, diaphragm aperture configurations, and output angular size. The main efforts focused on determining the optimal parameter values while ensuring minimal focal spot size and simultaneously maximizing beam energy.

Results. The computer modeling revealed the determining influence of each component of the five-element electron-optical structure on the formation of qualitative electron flow characteristics. A pronounced minimum in electron beam diameter was established at a specific combination of geometric and electrical system parameters. The thus-obtained optimum was used to develop a new methodology for designing and calibrating compact electron-beam devices that ensures maximum resolution and high sensitivity with minimal power consumption. Detailed analysis demonstrated that the optimal electrode configuration reduces spherical aberration by 25% compared to traditional solutions.

Conclusions. The developed design approach for microcolumn electron-optical systems significantly enhances performance while expanding the functional capabilities of electron microscopes and related analytical instruments. The practical significance of the work is confirmed by the possibility of creating devices with record resolution indicators in compact sizes. An important achievement is the establishment of quantitative optimization criteria for enabling targeted improvement of electron-beam system characteristics.

Keywords: electron-beam microsystem, electrostatic optics, microcolumn, microlens, Schottky emitter, optimal design, mathematical modeling, multi-beam lithography

For citation: Kuznetsov P.S., Sinelnikov A.O. Optimal electrode design for microminiature electronic optics. *Russian Technological Journal.* 2026;14(3):106–114. https://doi.org/10.32362/2500-316X-2026-14-3-106-114, https://www.elibrary.ru/RKATXE

Financial disclosure: The authors have no financial or proprietary interest in any material or method mentioned.

The authors declare no conflicts of interest.

НАУЧНАЯ СТАТЬЯ

Оптимальная конструкция электродов для микроминиатюрной электронной оптики

П.С. Кузнецов [1, @], А.О. Синельников [2]

[1] *АО «Государственный научно-исследовательский институт приборостроения», Москва, 129226 Россия*
[2] *Российский университет дружбы народов, Москва, 117198 Россия*
[@] *Автор для переписки, e-mail: ps_kuznetsov@mail.ru*

• **Поступила:** 12.09.2025 • **Доработана:** 15.11.2025 • **Принята к опубликованию:** 27.03.2026

Резюме

Цели. Цель работы – систематический анализ и оптимизация основных конструкционно-технологических характеристик микроминиатюрных электронно-оптических систем для достижения максимальных показателей их работоспособности. В ходе исследования особое внимание уделялось установлению взаимосвязей между геометрическими параметрами системы и ее функциональными характеристиками.

Методы. В основе исследования лежит комплексное математическое моделирование динамики электронов в сложной пятиэлектродной схеме, достоверно воспроизводящей реальную конструкцию компактной электронно-лучевой микроколонны. Данный подход позволил установить количественные зависимости критических показателей производительности системы – разрешающей способности и интенсивности электронного пучка – от фундаментальных геометрических параметров: межэлектродных расстояний, конфигурации апертур диафрагм и углового размера выходного отверстия. Основные усилия были сосредоточены на определении оптимальных значений указанных параметров, обеспечивающих минимальный размер фокального пятна при одновременной максимизации энергии пучка.

Результаты. Проведенное компьютерное моделирование выявило определяющее влияние каждого компонента пятиэлементной электронно-оптической структуры на формирование качественных характеристик электронного потока. Установлено наличие выраженного минимума диаметра электронного пучка при определенной комбинации геометрических и электрических параметров системы. Обнаруженный оптимум позволил разработать новую методику проектирования и калибровки компактных электронно-лучевых приборов, обеспечивающую достижение максимального разрешения и высокой чувствительности при минимальном энергопотреблении. Детальный анализ продемонстрировал, что оптимальная конфигурация электродов позволяет снизить сферическую аберрацию на 25% по сравнению с традиционными решениями.

Выводы. Разработанный подход к проектированию электронно-оптических систем микроколонн существенно повышает производительность и расширяет функциональные возможности электронных микроскопов и родственных аналитических приборов. Практическая значимость работы подтверждается возможностью создания устройств с рекордными показателями разрешения при компактных размерах. Важным достижением является установление количественных критериев оптимизации, позволяющих целенаправленно улучшать характеристики электронно-лучевых систем.

Ключевые слова: электронно-лучевая микросистема, электростатическая оптика, микроколонна, микролинза, эмиттер Шоттки, оптимальное проектирование, математическое моделирование, многолучевая литография

Для цитирования: Кузнецов П.С., Синельников А.О. Оптимальная конструкция электродов для микроминиатюрной электронной оптики. *Russian Technological Journal*. 2026;14(3):106–114. https://doi.org/10.32362/2500-316X-2026-14-3-106-114, https://www.elibrary.ru/RKATXE

Прозрачность финансовой деятельности: Авторы не имеют финансовой заинтересованности в представленных материалах или методах.

Авторы заявляют об отсутствии конфликта интересов.

INTRODUCTION

The introduction of microtechnologies into electronic optics has led to a qualitative leap in the characteristics and utility of electron beam devices and equipment. The creation of electronic sources, lenses, and deflectors having dimensions reduced by an order of magnitude or more compared to conventional ones [1–4] has given impetus to the development of microcolumns for miniature electron microscopes [5, 6] and multi-beam lithography systems [7–9] offering new technical and economic capabilities. The transition to smaller scales has expanded the prospects for the application of electron beam systems in areas related to the storage, processing, and display of information [10–16], where they have been supplanted by semiconductor devices and devices based on liquid crystals, as well as other materials having pronounced sensory and electroluminescent properties.

For design, technological, and electronic-optical reasons, microcolumns are made completely electrostatic. The first such column, having a total length of about 3.5 mm, consisted of a thermoelectric electron source with a two-electrode cathode lens, an aperture diaphragm, an octupole deflector, and a three-electrode focusing lens [17, 18]. The lenses were comprised of stacks of silicon crystals separated by insulating spacers made of Pyrex glass, with membrane windows for the electrodes. Traditional electron beam lithography and reactive ion etching microelectronics technologies were used to form holes in the electrodes, while multilayer anodic welding was used to connect the silicon components. The microcolumn [17], which demonstrated very high performance in low-voltage mode (resolution of about 10 nm at an accelerating voltage of 1 kV and a beam current of ~1 nA), served as the prototype for all subsequent modifications, representing a kind of reference standard. Ongoing research aimed at expanding microcolumn functionality and simplifying their manufacture concerns the structure (composition and mutual arrangement) and operating parameters of functional elements. One noteworthy experimental variant consists in an integrated electron-optical system (microlens) having dimensions of 1 × 1 × 0.05 cm, which performs electron extraction, focusing, and beam deflection [19–21]. The positioning of the cathode at 1–2 mm away from the microlens reduces the requirements for its installation accuracy and makes it possible to use any type of source: cold field, thermionic, or thermofield (Schottky emitter). Another promising modification is a dual-lens microcolumn [22, 23], which, while maintaining optimal aperture conditions in the objective lens, ensuring minimum aberrations and sufficient resolution for sub-100-nm technologies (no worse than 20–50 nm per 1 kV), significantly increases the solid angle of reception and, accordingly, the beam current (up to 50 nA at a source angular brightness of 100 μA/cm²), albeit at the expense of a slight increase in the size of the system (~7 mm).

PROBLEM STATEMENT

The task of optimal design consists in determining the optimal parameters of the geometry and mode of the microlens to ensure the minimum diameter of the probe (focused beam in the object plane) for given source characteristics, focusing distance (working distance), accelerating voltage (determining the final energy of electrons), as well as physical and technological limitations. Unlike [2, 15], the optimization problem was solved with fewer variable geometric parameters (the diameters of the holes and the interelectrode distances were selected to be the same), but with a larger number of electrodes and corresponding variable potentials.

Regardless of the optimization criterion (maximum resolution, i.e., the smallest possible size of the probe formed, the minimum size of the probe at a given current, or the minimum of individual aberrations), the optimal parameters are determined for a narrow range of electron energies, working distances, and source characteristics. For other conditions, the obtained parameters are nonoptimal. Therefore, lenses that are optimized to a greater extent by the selection of potentials rather than by geometry, which cannot be changed, appear to be more flexible in terms of utility. This, as well as the focus on simplifying the technology, which also includes the

adjustment function, determines the formulation of the optimization problem with a minimum number of variable geometric parameters. To evaluate the electron-optical advantages and select the preferred option, the maximum calculated resolutions of optimized lenses consisting of three, four, and five identical silicon membrane electrodes were compared with each other and with those obtained under the same conditions by a three-electrode lens optimized for all geometric parameters [2].

MATHEMATICAL MODEL OF AN ELECTRONIC LENS

Studies [24–32] present the results of various numerical experiments aimed at structural and parametric optimization to evaluate the limiting focusing properties of micro-miniature diaphragm systems in a given range of geometric parameters. In this case, optimization is understood as a search—under conditions of sequential increase in the number of electrodes, hole diameters, distances between electrodes, and potentials applied to them—for the linear increase, accelerating voltage, and working segment that provide minimum axial aberration coefficients when varied within specified limits. In each of the cases presented, a direct optimization method is used with original algorithms that enable tracking and correction of the axial distribution of the potential. The electrostatic field is typically defined analytically as a superposition of the fields of individual diaphragms.

The practical optimization methods used, which provide the initial data for designing high-quality optics, are therefore entirely adequate for the task at hand. Moreover, the same programs can be used to obtain an approximate solution to the optimization problem, which is presented as a linear combination of a sufficiently large number of functions of the appropriate type [9]. This allows us to estimate how close we can get to the limiting values of aberration coefficients under given conditions. A similar approach was tested in the problem of optimal synthesis of magnetic focusing fields when compared with the solution obtained by the optimal control method [10].

With optimal design of the electrostatic microminiature optics of electron beams shown in Fig. 1, the following data becomes available for processing. The variable parameters will be the voltage U_i, interelectrode distances l_i, diaphragm diameters d_i, and angular aperture α_0. The initial data for optimization are the working length (l_0 = 1 mm) and source parameters (source radius r_0 = 2 nm and angular brightness 10 μA/cr, voltage variation range ΔU_0 = 0.2 V, distance from the source to the first electrode l_{src} = 50–250 μm).

The objective of optimization is to achieve a minimum beam diameter equal to:

$$d = \sqrt{(Md_0)^2 + d_{chr}^2 + d_{sf}^2 + d_{dif}^2} = \min, \qquad (1)$$

where M is the linear magnification of the lens; Md_0 is the diameter of the probe in the cross-sectional plane; d_{chr} is the diameter of the chromatic aberration spot; d_{sf} is the diameter of the spherical aberration spot; d_{dif} is the diameter of the electron beam cross-section formed as a result of diffraction.

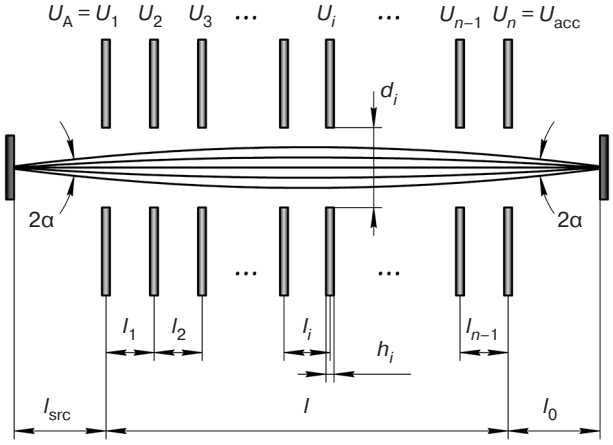


Fig. 1. Initial model of an electronic lens. U_A is the voltage at the starting point A of the lens, U_{acc} is the accelerating voltage

Given the existing limitations on electric field strength ($E \le 10^4$ V/mm) and geometric parameters (l_i = 100–500 μm, d_i = 20–200 μm, hole diameter h_i = 1–3 μm) the following technical characteristics are expected: accelerating voltage of about 1–3 kV; beam current greater than 1 nA; probe diameter about 10 nm; column height about 5 mm.

According to the selected mathematical model, we obtain the following characteristics:

- minimizable beam radius:

$$r = \sqrt{r_G^2 + \left(r_{dif}\right)^2 + \left(r_{sf}\right)^2 + \left(r_{chr}\right)^2}, \qquad (2)$$

where $r_G = Mr_0$ is the radius of the Gaussian image, $r_{dif} = \dfrac{0.75M}{\alpha_0\sqrt{U_0}}$ is the radius of the diffraction blur, $r_{sf} = MC_{sf}\alpha_0^3$ is the radius of the spherical aberration discs, $r_{chr} = MC_{chr}\alpha_0\dfrac{\Delta U_0}{U_0}$ is the radius of the chromatic aberration discs;

- axial distribution of potential:

$$E(z) = \frac{2}{\pi}\sum_{i=1}^{n}\tilde{E}_i\left[\operatorname{arctg}\left(\frac{2(z-z_i)}{h_i}\right) + \frac{\dfrac{2(z-z_i)}{h_i}}{1+\left(\dfrac{2(z-z_i)}{h_i}\right)^2}\right], \qquad (3)$$

where $\tilde{E}_i = \frac{E_i - E_{i-1}}{2}$, $E_i = \frac{U_i - U_{i+1}}{l_i}$, $i = \overline{1, n-1}$,

$E_0 = E_n = 0$; z is the axial position of the source, z_i is the axial coordinate of the ith electrode.

- spherical and chromatic aberration coefficients:

$$C_{\text{sf}} = \frac{1}{32}\int_{z_0}^{z_i}\sqrt{\frac{U}{U_0}}r^4 \times$$

$$\times\left\{\left(\frac{U'}{U}\right)^2\left[\frac{r'}{r}+\frac{5}{6}\cdot\frac{U'}{U}\right]^2+\frac{3}{2}\left(\frac{U''}{U}+\frac{U'}{U}\cdot\frac{r'}{r}-\left(\frac{U'}{U}\right)^2\right)^2+\right.$$

$$+\left(\frac{U'}{U}\right)^2\left(\frac{r'}{r}+\frac{5}{6}\cdot\frac{U'}{U}\right)^2+ \quad (4)$$

$$\left.+\frac{1}{36}\left(\frac{U'}{U}\right)^4+\left[\left(\frac{U''}{U}+\frac{U'}{U}\cdot\frac{r'}{r}\right)-\frac{5}{4}\left(\frac{U'}{U}\right)^2\right]^2\right\}dz,$$

$$C_{\text{chr}} = \frac{3}{8}\int_{z_0}^{z_i}\sqrt{\frac{U_0}{U}}r^2\left(\frac{U'}{U}\right)^2 dz. \quad (5)$$

RESULTS AND DISCUSSION

The results of modeling and optimization of the five-electrode lens are shown in Figs. 2–4. Figure 2 shows the dependence of the minimum radius of the electron lens on the angular aperture, which has a characteristic minimum at $\alpha_0 = 4$ mrad. It can be seen that at first the lens radius decreases rapidly with increasing aperture to reach its minimum value, after which it begins to increase gradually with further growth of the angular aperture.

As can be seen from the graphs shown in Fig. 3, the maximum focus and corresponding beam aperture decrease almost linearly with increasing length of the five-electrode lens under consideration.

The dependencies of the minimum radius of the electron lens on the aperture diameter (Fig. 4) at values of α_0, equal to 4 and 6 mrad, is distinctly nonlinear and is graphically described by a U-shaped curve with a local minimum. This means that as the aperture diameter increases, the minimum radius value first decreases smoothly, passing through the point of extremum (minimum value). After passing the minimum point, a further increase in the aperture diameter causes a gradual increase in the radius.

Thus, there is an optimal aperture diameter at which the minimum focus radius of the electron beam is achieved. At the minimum point, the electron lens has the following parameters:

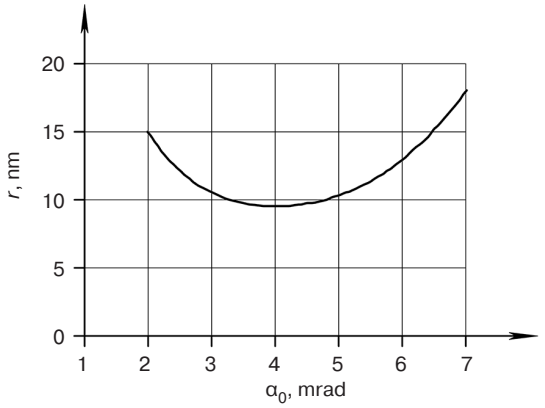


Fig. 2. Dependence of the minimum radius on the angular aperture

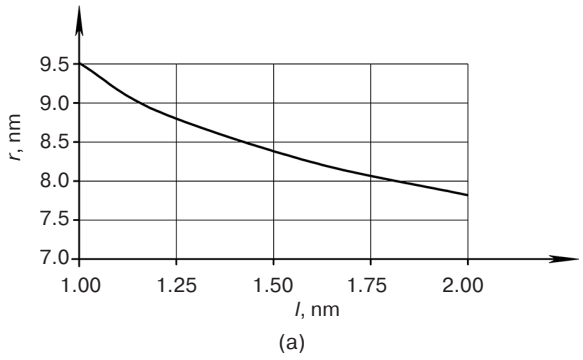


Fig. 3. Maximum focus (a) and corresponding optimal beam aperture (b) depending on the length of the five-electrode lens, $d_i = 1/4$

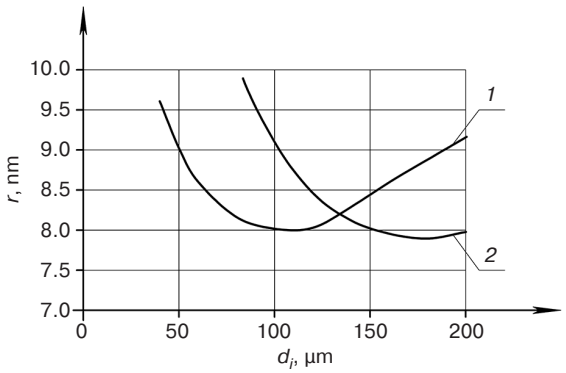


Fig. 4. Dependence of the minimum radius of the electron lens on the diameter of the diaphragm: (*1*) $\alpha_0 = 4$ mrad, (*2*) $\alpha_0 = 6$ mrad; $l_0 = 0.25$, $l_{\text{src}} = 1$, $U_0 = 1$ kV

1. Electro-optical characteristics: C_{chr} = 1.65 mm, C_{sf} = 51.42 mm, M = 1.34, r_G = 2.64 nm, r_{chr} = 1.76 nm, r_{sf} = 4.33 nm, r_{dif} = 7.83 nm.
2. Mode parameters: U_1 = 0.62 kV, U_2 = 1.17 kV, U_3 = 3.5 kV, U_A = U_{acc} = 1 kV, l_{src} = 250 μm, l_i = 250 μm, d_i = 50 μm, l_0 = 1 mm.

Based on the beam current value I = 1 nA and the angular aperture α_0 = 6 mrad, we obtain the following values for the minimum beam diameter: d_{min} = 26 nm (at ΔU_0 = 0.2 V), d_{min} = 54 nm (at ΔU_0 = 2 V).

Based on the results of the study, it can be concluded that the most appropriate configuration is a five-electrode electronic lens having evenly spaced electrodes, which reduces aberrations by approximately 10% by increasing the number of electrodes and varying the inter-electrode distances while maintaining the overall length. Maximum efficiency is achieved with an increased lens length, which is associated with an increase in the number of potential focusing areas and improved interaction of voltage fields with electrons. In addition, the optimal ratio between lens length, beam diameter, and aperture diameter was determined to minimize energy losses and scattering effects.

CONCLUSIONS

The methods of numerical, analytical modeling and software-support tools for optimizing computational experiments developed over the course of the research were aimed at finding optimal configurations and operating modes of low-voltage electrostatic optics for electron beam microcolumns in accordance with the formulated criteria. A detailed study of the electron-optical properties of microminiature diaphragm lenses implemented on the basis of microtechnologies was carried out.

The obtained estimates of the minimum probe size formed by an optimized five-electrode lens under different source parameters and aperture conditions showed that thermoemissive cathodes can be used to produce optimized five-electrode lenses having a length of 2.5 mm, which can form a probe having a diameter of 4 to 10 nm at distances of ~1 mm depending on the angular aperture (and, accordingly, the current in the probe). Such distances are sufficient for manipulating the object and placing special miniaturized secondary radiation detectors.

Authors' contribution

Both authors contributed equally to the conceptualization, methodology, investigation, and writing of this work. All authors have read and agreed to the published version of the manuscript.

REFERENCES

1. Zhao X., Fan B., Ma Z., Zhong S., Chen J., Zhang T., Su H. Optical-digital joint design of multi-order diffractive lenses for lightweight high-resolution computational imaging. *Opt. Lasers Eng.* 2024;180:108308. https://doi.org/10.1016/j.optlaseng.2024.108308
2. Thomson M.G.R., Chang T.H.P. Lens and deflector design for microcolumns. *J. Vac. Sci. Technol. B*. 1995;13(6):2445–2449. https://doi.org/10.1116/1.588018
3. Perng D.C., Crewe D.A., Feinerman A.D. Micromachined thermionic emitters. *J. Micromech. Microeng.* 1992;2(1):25–30. https://doi.org/10.1088/0960-1317/2/1/006
4. Lee J.-W., Park I.-Y., Ogawa T. Design and optimization of a conical electrostatic objective lens of a low-voltage scanning electron microscope for surface imaging and analysis in ultra-high-vacuum environment. *Ultramicroscopy*. 2024;257:113908. https://doi.org/10.1016/j.ultramic.2023.113908
5. Laszczyk K., Krysztof M. Electron beam source for the miniaturized electron microscope on-chip. *Vacuum*. 2021;189:110236. https://doi.org/10.1016/j.vacuum.2021.110236
6. Krysztof M., Białas M., Szyszka P., Grzebyk T., Górecka-Drzazga A. Fabrication and characterization of a miniaturized octupole deflection system for the MEMS electron microscope. *Ultramicroscopy*. 2021;225:113288. https://doi.org/10.1016/j.ultramic.2021.113288
7. Hofmann U., Ünal N., Klikovits J. From ghost to state-of-the-art process corrections – PEC enabled e-beam nanofabrication. *Micro and Nano Engineering* (*MNE*). 2024;25:100286. https://doi.org/10.1016/j.mne.2024.100286
8. Karkantonis T., Penchev P., Nasrollahi V., Le H., See T.L., Bruneel D., Ramos-de-Campos J.A., Dimov S. Laser micro-machining of freeform surfaces: Accuracy, repeatability and reproducibility achievable with multi-axis processing strategies. *Precis. Eng*. 2022;78:233–247. https://doi.org/10.1016/j.precisioneng.2022.08.009
9. Parker N.W., Brodie A.D., McCoy J.H. A high throughput NGL electron-beam direct-write lithography system. In: *Proceeding SPIE Emerging Lithographic Technologies*. V. 3997. 2000. https://doi.org/10.1117/12.390042
10. Itoh S., Tanaka M., Tonegawa T. Development of field emission displays. *J. Vac. Sci. Technol. B*. 2004;22(3):1362–1366. https://doi.org/10.1116/1.1691409
11. Villarroya M., Barniol N., Martin C., Pérez-Murano F., Esteve J., Bruchhaus L., Jede R., Bourhis E., Gierak J. Fabrication of nanogaps for MEMS prototyping using focused ion beam as a lithographic tool and reactive ion etching pattern transfer. *Microelectron. Eng*. 2007;84(5-8):1215–1218. https://doi.org/10.1016/j.mee.2007.01.074

12. Appleton B.R., Tongay S., Lemaitre M., Gila B., Fridmann J., Mazarov P., Sanabia J.E., Bauerdick S., Bruchhaus L., Mimura R., Jede R. Materials modifications using a multi-ion beam processing and lithography system. *Nuclear Instruments and Methods in Physics Research Section B: Beam Interactions with Materials and Atoms*. 2012;272:153–157. https://doi.org/10.1016/j.nimb.2011.01.054
13. Pease R.F., Ioakeimidi K., Aldana R., et al. Photoelectronic analog-to-digital conversion using miniature electron optics: Basic design considerations. *J. Vac. Sci. Technol. B*. 2003;21(6):2826–2829. https://doi.org/10.1116/1.1621664
14. Srivastava K., Le-The H., Lozeman J.J.A., van den Berg A., van der Stam W., Odijk M. Prospects of nano-lithographic tools for the fabrication of surface-enhanced Raman spectroscopy (SERS) substrates. *Micro and Nano Engineering* (*MNE*). 2024;23:100267. https://doi.org/10.1016/j.mne.2024.100267
15. Kuznetsov P.S. Issues and prospects for the development of mechatronics and microsystem technology. *Nano- i mikrosistemnaja tehnika = Nano- and Microsystem Technique.* 2024;26(4):159–169 (in Russ.). https://doi.org/10.17587/nmst.25.159-169
16. Kuznetsov P.S. Microelectromechanical systems for improved gyroscope design. *Russian Technological Journal*. 2025;13(3):103–121. https://doi.org/10.32362/2500-316X-2025-13-3-103-121
17. Chang T.H.P., Thomson M.G.R., Yu M.L., et al. Electron beam technology – SEM to microcolumn. *Microelectron. Eng*. 1996;32(1-4):113–130. https://doi.org/10.1016/0167-9317(95)00366-5
18. Kim H.S., Ahn S., Kim D.W., Oh T.-S., Ahn S.J. Efficient electron beam condensing for low-energy microcolumn lithography. *Microelectronics J*. 2008;39(1):94–98. https://doi.org/10.1016/j.mejo.2007.09.030
19. Hu J., Yue L., Ma Y., Liu F., Kang Y. Aberration calculation of microlens array using differential algebraic method. *Ultramicroscopy*. 2025;269:114085. https://doi.org/10.1016/j.ultramic.2024.114085
20. Zlatkin A., Garcia N. Low-energy (300eV) versatile scanning electron microscope with 30 nm resolution. *Microelectron. Eng*. 1999;45(1):39–46. https://doi.org/10.1016/S0167-9317(98)00260-3
21. Ximen J. Chapter One – The electron optical imaging system and its aberrations. In: Hawkes P.W., Hÿtch M. (Eds.). *Advances in Imaging and Electron Physics*. Elsevier; 2023. V. 226. P. 1–88. https://doi.org/10.1016/bs.aiep.2023.03.003
22. Weigand H., Gautsch S., Strohmaier W., Blideran M., Staufer U., de Rooij N.F., Kern D.P. Microcolumn with variable axis lens for large scan fields and pixel numbers. *Microelectron. Eng*. 2011;88(8):2431–2434. https://doi.org/10.1016/j.mee.2010.12.120
23. Mankos M., Lee K.Y., Muray L., et al. Optimization of microcolumn electron optics for high current application. *J. Vac. Sci. Technol. B.* 2000;18(6):3057–3060. https://doi.org/10.1116/1.1321756
24. Feng Y., Li W., Chen Y., Kang X., Li J., Tang K., Zhao Z., Liu X., Zhou K., You Y., Li M., Li P., Xu Z., Zhao T., Mao R. Design and performance of a low-emittance electron gun for electron beam probe. *Nuclear Inst. Methods in Physics Research, A: Accelerators, Spectrometers, Detectors and Associated Equipment*. 2024;1066:169604. https://doi.org/10.1016/j.nima.2024.169604
25. Fox J., Lee H.K., Alajo A.B., Avachat A. Simulation study of electron beam optics for a distributed X-ray source toward stationary CT architecture. *Nuclear Inst. Methods in Physics Research, A: Accelerators, Spectrometers, Detectors and Associated Equipment*. 2024;1062:169149. https://doi.org/10.1016/j.nima.2024.169149
26. Sinelnikov A.O., Smetanin I.A., Basov B.A., Smetanin E.A., Bykanova U.F. Effect of Ignition Voltage Amplitude on Breakdown Delay Time of Ring Laser Discharge Gap. In: *2025 7th International Youth Conference on Radio Electronics, Electrical and Power Engineering* (*REEPE*). 2025. https://doi.org/10.1109/REEPE63962.2025.10971095
27. Kazmiruk V.V., Kurganov I.G., Podkopaev A.A., et al. Optimization of the Electron Optical System of a Scanning Electron Microscope for Measuring the Size of Micro- and Nanoobjects. *J. Surf. Investig.* 2019;13(4):594–599. https://doi.org/10.1134/S1027451019040074
[Original Russian Text: Kazmiiruk V.V., Kurganov I.G., Podkopaev A.A., Savitskaya T.N. Optimization of the Electron Optical System of a Scanning Electron Microscope for Measuring the Size of Micro- and Nanoobjects. *Poverkhnost'. Rentgenovskie, sinkhrotronnye i neitronnye issledovaniya.* 2019;7:36–41 (in Russ.). https://doi.org/10.1134/S0207352819070072]
28. Bakulin E.M., Kurushin G.V., Stroganov K.A. Results of simulation of the microaccelerometer design based on integrated technology of micromechanics and acoustoelectronics. *Ehlektronnaya tekhnika. Seriya 3: Mikroehlektronika = J. Electronic Engineering. Series 3*: *Microelectronics*. 2018;2(170):43–49 (in Russ.). https://elibrary.ru/xtxzvz
29. Trapashko G.A. Synthesis of optoelectronic system for control of microelectronic structures. *Nauka i tekhnika = Science and Technique*. 2014;1:38–43 (in Russ.). https://elibrary.ru/rylmup
30. Belsky M.D., Suvorinov A.V., Filipchuk T.S., Shakhbazov S.Yu. Calculation of technological tolerances in electrostatic lenses for electron beam micro columns. *Izvestiya vysshikh uchebnykh zavedenii. Ehlektronika = Proceedings of Universities. Electronics*. 2006;1:76–83 (in Russ.). https://elibrary.ru/hsywrb
31. Kazmiiruk V.V., Kurganov I.G., Savitskaya T.N. Miniature forming lens for high-voltage electron-beam lithography system. *J. Surf. Investig.* 2020;14:1366–1370. https://doi.org/10.1134/S1027451020060336
[Original Russian Text: Kazmiiruk V.V., Kurganov I.G., Savitskaya T.N. Miniature forming lens for high-voltage electron-beam lithography system. *Poverkhnost'. Rentgenovskie, sinkhrotronnye i neitronnye issledovaniya* 2020;12:89–94 (in Russ.). https://doi.org/10.31857/S1028096020120158]
32. Belsky M.D., Lvov B.G., Rybalko V.V. Modeling the focusing microlenses of an auger analyzer equipped with a CMA. *Bull. Russ. Acad. Sci. Phys.* 2013;77(8):959–962. https://doi.org/10.3103/S1062873813080066
[Original Russian Text: Belsky M.D., Lvov B.G., Rybalko V.V. Modeling the focusing microlenses of an auger analyzer equipped with a CMA. *Izvestiya Rossiiskoi akademii nauk. Seriya fizicheskaya.* 2013;77(10):1059–1062 (in Russ.). https://doi.org/10.7868/S0367676513080061]

СПИСОК ЛИТЕРАТУРЫ

1. Zhao X., Fan B., Ma Z., Zhong S., Chen J., Zhang T., Su H. Optical-digital joint design of multi-order diffractive lenses for lightweight high-resolution computational imaging. *Opt. Lasers Eng.* 2024;180:108308. https://doi.org/10.1016/j.optlaseng.2024.108308
2. Thomson M.G.R., Chang T.H.P. Lens and deflector design for microcolumns. *J. Vac. Sci. Technol. B*. 1995;13(6):2445–2449. https://doi.org/10.1116/1.588018
3. Perng D.C., Crewe D.A., Feinerman A.D. Micromachined thermionic emitters. *J. Micromech. Microeng.* 1992;2(1):25–30. https://doi.org/10.1088/0960-1317/2/1/006
4. Lee J.-W., Park I.-Y., Ogawa T. Design and optimization of a conical electrostatic objective lens of a low-voltage scanning electron microscope for surface imaging and analysis in ultra-high-vacuum environment. *Ultramicroscopy*. 2024;257:113908. https://doi.org/10.1016/j.ultramic.2023.113908
5. Laszczyk K., Krysztof M. Electron beam source for the miniaturized electron microscope on-chip. *Vacuum*. 2021;189:110236. https://doi.org/10.1016/j.vacuum.2021.110236
6. Krysztof M., Białas M., Szyszka P., Grzebyk T., Górecka-Drzazga A. Fabrication and characterization of a miniaturized octupole deflection system for the MEMS electron microscope. *Ultramicroscopy*. 2021;225:113288. https://doi.org/10.1016/j.ultramic.2021.113288
7. Hofmann U., Ünal N., Klikovits J. From ghost to state-of-the-art process corrections – PEC enabled e-beam nanofabrication. *Micro and Nano Engineering* (*MNE*). 2024;25:100286. https://doi.org/10.1016/j.mne.2024.100286
8. Karkantonis T., Penchev P., Nasrollahi V., Le H., See T.L., Bruneel D., Ramos-de-Campos J.A., Dimov S. Laser micro-machining of freeform surfaces: Accuracy, repeatability and reproducibility achievable with multi-axis processing strategies. *Precis. Eng*. 2022;78:233–247. https://doi.org/10.1016/j.precisioneng.2022.08.009
9. Parker N.W., Brodie A.D., McCoy J.H. A high throughput NGL electron-beam direct-write lithography system. In: *Proceeding SPIE Emerging Lithographic Technologies*. V. 3997. 2000. https://doi.org/10.1117/12.390042
10. Itoh S., Tanaka M., Tonegawa T. Development of field emission displays. *J. Vac. Sci. Technol. B*. 2004;22(3):1362–1366. https://doi.org/10.1116/1.1691409
11. Villarroya M., Barniol N., Martin C., Pérez-Murano F., Esteve J., Bruchhaus L., Jede R., Bourhis E., Gierak J. Fabrication of nanogaps for MEMS prototyping using focused ion beam as a lithographic tool and reactive ion etching pattern transfer. *Microelectron. Eng*. 2007;84(5-8):1215–1218. https://doi.org/10.1016/j.mee.2007.01.074
12. Appleton B.R., Tongay S., Lemaitre M., Gila B., Fridmann J., Mazarov P., Sanabia J.E., Bauerdick S., Bruchhaus L., Mimura R., Jede R. Materials modifications using a multi-ion beam processing and lithography system. *Nuclear Instruments and Methods in Physics Research Section B: Beam Interactions with Materials and Atoms*. 2012;272:153–157. https://doi.org/10.1016/j.nimb.2011.01.054
13. Pease R.F., Ioakeimidi K., Aldana R., et al. Photoelectronic analog-to-digital conversion using miniature electron optics: Basic design considerations. *J. Vac. Sci. Technol. B*. 2003;21(6):2826–2829. https://doi.org/10.1116/1.1621664
14. Srivastava K., Le-The H., Lozeman J.J.A., van den Berg A., van der Stam W., Odijk M. Prospects of nano-lithographic tools for the fabrication of surface-enhanced Raman spectroscopy (SERS) substrates. *Micro and Nano Engineering* (*MNE*). 2024;23:100267. https://doi.org/10.1016/j.mne.2024.100267
15. Кузнецов П.С. Вопросы и перспективы развития мехатроники и микросистемной техники. *Нано- и микросистемная техника*. 2024;25(4):159–169. https://doi.org/10.17587/nmst.25.159-169
16. Кузнецов П.С. Микроэлектромеханические системы: путь к совершенствованию гироскопов. *Russian Technological Journal*. 2025;13(3):103–121. https://doi.org/10.32362/2500-316X-2025-13-3-103-121
17. Chang T.H.P., Thomson M.G.R., Yu M.L., et al. Electron beam technology – SEM to microcolumn. *Microelectron. Eng*. 1996;32(1-4):113–130. https://doi.org/10.1016/0167-9317(95)00366-5
18. Kim H.S., Ahn S., Kim D.W., Oh T.-S., Ahn S.J. Efficient electron beam condensing for low-energy microcolumn lithography. *Microelectronics J*. 2008;39(1):94–98. https://doi.org/10.1016/j.mejo.2007.09.030
19. Hu J., Yue L., Ma Y., Liu F., Kang Y. Aberration calculation of microlens array using differential algebraic method. *Ultramicroscopy*. 2025;269:114085. https://doi.org/10.1016/j.ultramic.2024.114085
20. Zlatkin A., Garcia N. Low-energy (300eV) versatile scanning electron microscope with 30 nm resolution. *Microelectron. Eng*. 1999;45(1):39–46. https://doi.org/10.1016/S0167-9317(98)00260-3
21. Ximen J. Chapter One – The electron optical imaging system and its aberrations. In: Hawkes P.W., Hÿtch M. (Eds.). *Advances in Imaging and Electron Physics*. Elsevier; 2023. V. 226. P. 1–88. https://doi.org/10.1016/bs.aiep.2023.03.003
22. Weigand H., Gautsch S., Strohmaier W., Blideran M., Staufer U., de Rooij N.F., Kern D.P. Microcolumn with variable axis lens for large scan fields and pixel numbers. *Microelectron. Eng*. 2011;88(8):2431–2434. https://doi.org/10.1016/j.mee.2010.12.120
23. Mankos M., Lee K.Y., Muray L., et al. Optimization of microcolumn electron optics for high current application. *J. Vac. Sci. Technol. B.* 2000;18(6):3057–3060. https://doi.org/10.1116/1.1321756
24. Feng Y., Li W., Chen Y., Kang X., Li J., Tang K., Zhao Z., Liu X., Zhou K., You Y., Li M., Li P., Xu Z., Zhao T., Mao R. Design and performance of a low-emittance electron gun for electron beam probe. *Nuclear Inst. Methods in Physics Research, A: Accelerators, Spectrometers, Detectors and Associated Equipment*. 2024;1066:169604. https://doi.org/10.1016/j.nima.2024.169604
25. Fox J., Lee H.K., Alajo A.B., Avachat A. Simulation study of electron beam optics for a distributed X-ray source toward stationary CT architecture. *Nuclear Inst. Methods in Physics Research, A: Accelerators, Spectrometers, Detectors and Associated Equipment*. 2024;1062:169149. https://doi.org/10.1016/j.nima.2024.169149

26. Sinelnikov A.O., Smetanin I.A., Basov B.A., Smetanin E.A., Bykanova U.F. Effect of Ignition Voltage Amplitude on Breakdown Delay Time of Ring Laser Discharge Gap. In: *2025 7th International Youth Conference on Radio Electronics, Electrical and Power Engineering* (*REEPE*). 2025. https://doi.org/10.1109/REEPE63962.2025.10971095
27. Казьмирук В.В., Курганов И.Г., Подкопаев А.А., Савицкая Т.Н. Оптимизация электронно-оптической системы растрового электронного микроскопа для измерения размеров микро- и нанообъектов. *Поверхность. Рентгеновские, синхротронные и нейтронные исследования*. 2019;7:36–41. https://doi.org/10.1134/S0207352819070072
28. Бакулин Е.М., Курушин Г.В., Строганов К.А. Результаты моделирования конструкции микроакселерометра на основе интегрированной технологии микромеханики и акустоэлектроники. *Электронная техника. Серия 3: Микроэлектроника*. 2018;2(170):43–49. https://elibrary.ru/xtxzvz
29. Трапашко Г.А. Синтез оптоэлектронной системы установки контроля микроэлектронных структур. *Наука и техника*. 2014;1:38–43. https://elibrary.ru/rylmup
30. Бельский М.Д., Суворинов А.В., Филипчук Т.С., Шахбазов С.Ю. Расчет технологических допусков в электростатических линзах для электронно-лучевых микроколонн. *Известия высших учебных заведений. Электроника*. 2006;1:76–83. https://elibrary.ru/hsywrb
31. Казьмирук В.В., Курганов И.Г., Савицкая Т.Н. Расчет миниатюрной формирующей линзы высоковольтного электронного литографа. *Поверхность. Рентгеновские, синхротронные и нейтронные исследования*. 2020;12:89–94. https://doi.org/10.31857/S1028096020120158
32. Бельский М.Д., Львов Б.Г., Рыбалко В.В. Моделирование фокусирующей микролинзы для оже-анализатора с ЦЗА. *Известия Российской академии наук. Серия физическая*. 2013;77(8):1059–1062. https://doi.org/10.7868/S0367676513080061

About the Authors

Pavel S. Kuznetsov, Cand. Sci. (Eng.), Deputy Head of the Experimental Complex of Microelectronics and Micromechanical Systems, State Scientific Research Institute of Instrument Engineering (GosNIIP) (125, Mira pr., Moscow, 129226 Russia). E-mail: ps_kuznetsov@mail.ru. Scopus Author ID 58513707600, RSCI SPIN-code 6564-9540, https://orcid.org/0000-0001-5459-7883

Anton O. Sinelnikov, Cand. Sci. (Eng.), Associated Professor, Basic Department “Nanotechnology and Microsystem Technology,” RUDN University (6, Miklukho-Maklaya ul., Moscow, 117198 Russia). E-mail: mr.sinelnikov.a@mail.ru. Scopus Author ID 55382453500, ResearcherID AAC-2606-2022, RSCI SPIN-code 2442-7507, https://orcid.org/0000-0002-5579-3509

Об авторах

Кузнецов Павел Сергеевич, к.т.н., заместитель начальника экспериментального комплекса микроэлектроники и микромеханических систем, Акционерное общество «Государственный научно-исследовательский институт приборостроения» (АО «ГосНИИП») (129226, Россия, Москва, пр-т Мира, д. 125). E-mail: ps_kuznetsov@mail.ru. Scopus Author ID 58513707600, SPIN-код РИНЦ 6564-9540, https://orcid.org/0000-0001-5459-7883

Синельников Антон Олегович, к.т.н., доцент, кафедра «Нанотехнологии и микросистемная техника», ФГАОУ ВО «Российский университет дружбы народов имени Патриса Лумумбы) (РУДН) (117198, Россия, Москва, ул. Миклухо-Маклая, д. 6). E-mail: mr.sinelnikov.a@mail.ru. Scopus Author ID 55382453500, ResearcherID AAC-2606-2022, SPIN-код РИНЦ 2442-7507, https://orcid.org/0000-0002-5579-3509

Translated from Russian into English by L. Bychkova
Edited for English language and spelling by Thomas A. Beavitt

Mathematical modeling

Математическое моделирование

UDC 001.57; 004.942
https://doi.org/10.32362/2500-316X-2026-14-3-115-130
EDN MYWEJW

RESEARCH ARTICLE

Features of analytical modeling of nonlinear surface waves in gradient media

Sergey E. Savotchenko [1, @], Nadezhda O. Afanasyeva [2]

[1] *MIREA – Russian Technological University, Moscow, 119454 Russia*
[2] *Sergo Ordzhonikidze Russian State Geological Prospecting University, Moscow, 117997 Russia*
[@] *Corresponding author, e-mail: savotchenkose@mail.ru*

• **Submitted:** 13.10.2025 • **Revised:** 10.11.2025 • **Accepted:** 20.03.2026

Abstract

Objectives. An important role in modern physics, particularly in waveguide optics, is played by studies that involve the search for exact solutions to equations used in modeling to identify classes of exactly solvable models. This work set out to use analytical modeling methods to explore the properties of surface shear waves propagating without loss along the interface between a nonlinear and a graded-index nonmagnetic medium.

Methods. Methods of mathematical modeling, calculus, mathematical physics, differential equations, and the theory of special functions were used. Fundamental principles, methods, and physical models of nonlinear and waveguide optics were also applied.

Results. The properties of surface transverse waves propagating along the interface between a nonlinear and a graded-index medium are modeled. In order to model the nonlinearity of the medium to describe the nonlinear optical response of the medium to electric field perturbations, the linear dependence of permittivity on light intensity is chosen as a parameter. The graded-index medium is modeled using a spatial profile of permittivity as a function of distance from the interface for which an exact analytical solution to the stationary wave equation can be found. A mathematical formulation of the model is presented, consisting of a conjugation boundary value problem for a nonlinear equation with variable coefficients. Exact analytical solutions to this boundary value problem are found for the cases of focusing and defocusing nonlinearities to describe the spatial distributions of the electric field strength in the direction transverse to the interface. Analysis of the model revealed significant differences in the spatial distribution of the field intensity in surface waves propagating in the focusing and defocusing media. The effect of the values of model parameters used to characterize the optical properties of the contacting media on the spatial distribution of light intensity in surface waves was also studied in detail.

Conclusions. The obtained, which results supplement the existing theory of nonlinear and waveguide optics, can be applied in the design of new waveguide structures with user-defined properties. The obtained new solutions expand the class of exactly solvable models of planar waveguide structures with distributed inhomogeneous and nonlinear properties.

Keywords: mathematical modeling, mathematical model, boundary value problem, exact solution, surface wave, waveguide optics, nonlinear optics

For citation: Savotchenko S.E., Afanasyeva N.O. Features of analytical modeling of nonlinear surface waves in gradient media. *Russian Technological Journal.* 2026;14(3):115–130. https://doi.org/10.32362/2500-316X-2026-14-3-115-130, https://www.elibrary.ru/MYWEJW

Financial disclosure: The authors have no financial or proprietary interest in any material or method mentioned.

The authors declare no conflicts of interest.

НАУЧНАЯ СТАТЬЯ

Особенности аналитического моделирования нелинейных поверхностных волн в градиентных средах

С.Е. Савотченко [1, @], Н.О. Афанасьева [2]

[1] *МИРЭА – Российский технологический университет, Москва, 119454 Россия*
[2] *Российский государственный геологоразведочный университет им. Серго Орджоникидзе, Москва, 117997 Россия*
[@] *Автор для переписки, e-mail: savotchenkose@mail.ru*

• **Поступила:** 13.10.2025 • **Доработана:** 10.11.2025 • **Принята к опубликованию:** 20.03.2026

Резюме

Цели. Важную роль в современной физике и волноводной оптике играют и исследования, связанные с нахождением точных решений используемых при моделировании уравнений, позволяющие выявить классы точно решаемых моделей. Цель работы – изучение свойств поверхностных поперечных волн, распространяющихся вдоль границы раздела нелинейной и градиентной немагнитных сред без потерь.

Методы. В работе использованы методы математического моделирования, методы анализа и математической физики, дифференциальных уравнений и теории специальных функций. Использовались базовые принципы, методы и физические модели нелинейной и волноводной оптики.

Результаты. Проведено моделирование свойств поверхностных поперечных волн, распространяющихся вдоль границы раздела нелинейной и градиентной сред. В качестве модели нелинейности среды, описывающей нелинейно-оптический отклик среды на возмущения электрического поля, выбрана линейная форма зависимости диэлектрической проницаемости от интенсивности света. В качестве модели градиентной среды выбрана форма пространственного профиля диэлектрической проницаемости, описывающая ее изменение в зависимости от расстояния до границы раздела, для которой можно найти точное аналитическое решение стационарного волнового уравнения. Приведена математическая формулировка модели, которая представляет собой краевую задачу сопряжения для нелинейного уравнения с переменными коэффициентами. Найдены точные аналитические решения данной краевой задачи для случаев фокусирующей и дефокусирующей нелинейностей, которые описывают пространственные распределения напряженности электрического поля в поперечном границе раздела сред направлении. Анализ модели позволил выявить существенные различия пространственного распределения интенсивности поля в поверхностных волнах, распространяющихся в фокусирующих и дефокусирующих средах. Проведен детальный анализ влияния значений параметров модели, характеризующих оптические свойства контактирующих сред, на пространственное распределение интенсивности света в поверхностных волнах.

Выводы. Полученные результаты дополняют существующую теорию нелинейной и волноводной оптики и могут найти применение при проектировании новых волноводных структур с определяемыми пользователями свойствами. Полученные новые решения расширяют класс точно решаемых моделей планарных волноводных структур с распределенными неоднородными и нелинейными свойствами.

Ключевые слова: математическое моделирование, математическая модель, краевая задача, точное решение, поверхностная волна, волноводная оптика, нелинейная оптика

Для цитирования: Савотченко С.Е., Афанасьева Н.О. Особенности аналитического моделирования нелинейных поверхностных волн в градиентных средах. *Russian Technological Journal*. 2026;14(3):115–130. https://doi.org/10.32362/2500-316X-2026-14-3-115-130, https://www.elibrary.ru/MYWEJW

Прозрачность финансовой деятельности: Авторы не имеют финансовой заинтересованности в представленных материалах или методах.

Авторы заявляют об отсутствии конфликта интересов.

INTRODUCTION

Mathematical modeling occupies a central place in modern physics and waveguide optics. In particular [1], this approach can be used to identify, describe, and predict the behavior of a physical system under the influence of various parameters. Most mathematical models of physical processes are based on boundary value problems for partial differential equations, which under certain conditions can transform into ordinary differential equations. Modern computing systems can be used not only to find solutions to such problems using numerical methods but also to visualize them. However, an important role in applied mathematical research his also played by problems of finding exact solutions to formulated equations, which permit the identification of classes of exactly solvable models. The existence of an exact solution, as opposed to a numerical one, makes it possible not only to quickly and clearly analyze the effect of model parameters on the process under study, but also to obtain asymptotic expressions and other estimates in explicit analytical form [2].

The obtaining of exact solutions to boundary value problems arising in the modeling of surface wave propagation in planar waveguide structures is discussed in the work [3]. Although waveguide properties of interfaces between media with different optical characteristics have been examined in numerous works [4–6], many unexplored problems remain that involve the analytical description of surface wave properties within the framework of waveguide models that allow exact solutions. Such models are referred to as exactly solvable [7].

In nonlinear and waveguide optics, the fundamental model equation is the wave equation, which is derived from Maxwell's equations. For time-independent wave processes, the wave equation transforms into the Helmholtz equation [8]. The optical properties of the medium in which the wave process is being studied are determined by the refractive index or permittivity (for nonmagnetic media). Therefore, the choice of the form of dependence of these characteristics on the process function (e.g., light intensity) or spatial variables determines the possibility of an exact solution to the model equation. Such dependencies, which are themselves model equations, characterize the optical response of the medium to various perturbations.

The nonlinear response of a medium is described by models in which the permittivity depends on the light intensity, i.e., on the square of the amplitude of the electric (or magnetic) field strength [9]. Many of these models allow exact analytical solutions. The most common of these is the Kerr nonlinearity model, in which the permittivity depends linearly on the light intensity. In such a nonlinear model, depending on the geometry of the system, there are various classes of exact solutions for nonlinear equations, known as solitons [10–12], kinks, and cnoidal waves [13]. Furthermore [12], solitons can be not only optical, but also of a different physical nature, which emphasizes the wide application of similar mathematical models fora large class of phenomena that differ in physical nature.

The spatial inhomogeneity of the optical properties of a medium is characterized using models in which the permittivity depends on spatial coordinates. In waveguide optics, the media in which the refractive index smoothly depends on the spatial coordinate are referred to as graded-index media [14]. In this case, the equation contains variable (distributed) coefficients determined by the chosen inhomogeneity model. Exact analytical solutions are known for a wide range of models; these solutions are expressed through various special functions [15].

From a physical perspective, it is important to investigate the characteristics of wave propagation along interfaces between media having different optical properties. When modeling such processes, boundary value problems with conjugation conditions at the interface are formulated. A number of studies examined

various surface waves propagating along interfaces between nonlinear and inhomogeneous media [16]. Exact solutions to model conjugation boundary value problems describing various types of surface waves were obtained [17, 18].

Recently [19], a study was made of surface waves along the surface of an optically inhomogeneous crystal with permittivity ε modeled by a special type of profile generalizing a hyperbolic profile ($\varepsilon \sim 1/x$) [20] and an inverse square profile ($\varepsilon \sim 1/x^2$) [21]. Consideration was made of the contact of an inhomogeneous medium having such a profile with a homogeneous linear medium in which the field intensity decreases exponentially with distance from the interface between the media. An exact analytical solution to the posed conjugation boundary value problem was found, the possibility of whose expression through the Whittaker function has been demonstrated.

The present work examines the contact of an inhomogeneous medium with an optical nonlinear medium having the same spatial profile. Kerr nonlinearity was chosen as the model of the nonlinear response. It is confirmed that the boundary value problem has an exact solution with this choice of inhomogeneity and nonlinearity models. The resulting solution is used to identify the characteristics of the formation of the spatial profile of the electric field in the transverse direction of the interface between the media at various optical parameters of the model.

The main result of this work is the derivation of an exact solution for this configuration of media. This solution describes a new type of wave with a novel shape of its profile in the direction transverse to the interface between the media. Furthermore, the sensitivity of the properties of this wave to changes in the optical parameters of the media is revealed.

1. MATHEMATICAL FORMULATION OF THE MODEL

Consider a plane interface between two lossless nonmagnetic media having different optical properties. Let the interface plane be located in the $x = 0$ coordinate plane (the yOz plane in space), and let the x axis be perpendicular to the interface and aligned with the direction of surface wave propagation. If the media are assumed to be uniform along their interface, then the electric field distribution in the transverse wave will also be uniform along it and characterized by spatial differences only in the transverse direction. When modeling surface waves in such systems, the problem is reduced to a one-dimensional one in the direction perpendicular to the interface; its solution describes the spatial distribution of the y component of the electric field strength vector in this direction.

As is known [8], the fundamental equation of the theory of waveguide optics for the transverse distribution $u(x)$ of the electric field in a surface wave can be written in the form

$$u''(x) + \{k_0^2 \varepsilon(x, I) - \beta^2\} u(x) = 0, \quad (1)$$

where $\varepsilon(x, I)$ is the permittivity; $I = u^2$ is the intensity of the electric field (light); $k_0 = 2\pi/\lambda_0$ is the longitudinal wave number; λ_0 is the wavelength of the sent radiation; and β is the propagation constant, which is related to the effective refractive index n by the expression $\beta = nk_0$.

Equation (1) is, in general, a nonlinear equation with variable coefficients.

The optical properties of media and their spatial distribution are completely described by the permittivity $\varepsilon(x, I)$. Therefore, the differences in models of media are described by this function. In the case of a flat interface between media, such a function is represented in step form:

$$\varepsilon(x, I) = \begin{cases} \varepsilon_G(x), x > 0, \\ \varepsilon_N(I), x < 0, \end{cases} \quad (2)$$

where it is considered that the model describes the contact of a nonlinear medium with a graded-index medium and assumed that the nonlinear medium occupies a half-space $x < 0$, while the graded-index medium fills the half-space $x > 0$.

In (2), the function $\varepsilon_N(I)$ describes the dependence of the permittivity on the light intensity in the nonlinear medium. Its specific form is determined by the model of the medium's nonlinearity, which characterizes the nonlinear optical response of the medium to electric field perturbations arising from the redistribution of internal charges with varying intensity of the laser radiation exciting the surface wave. Note that there is a wide variety of real materials exhibiting Kerr nonlinearity. For example, these include crystals like $AgGaSe_xS_{2(1+x)}$, which have high nonlinearity, including Kerr nonlinearity, and in which the refractive index changes proportionally to the square of the light field strength, as well as GaP, InAs, InP, and InSb semiconductors.

The function $\varepsilon_G(I)$ actually characterizes the spatial inhomogeneity of the refractive index profile in a graded-index medium. Its specific form is determined by the graded-index medium model, in which the change in refractive index with distance from the surface depends on the spatial distribution of charges induced, for example, by implanted ions. In this case, the refractive index profile can be similar to the spatial distribution of the implanted ion concentration. This means that, if it is technologically possible to create a specific ion concentration profile by implantation, a corresponding

refractive index profile will be obtained to determine the function $\varepsilon_G(I)$.

In this paper, we consider the most common model of nonlinearity in the form of a linear dependence of permittivity on light intensity (i.e., a quadratic dependence on electric field). This model describes nonlinear media that exhibit the Kerr effect and are called Kerr media. In the case of Kerr nonlinearity, the function $\varepsilon_N(I)$ can be written as

$$\varepsilon_N(I) = \varepsilon_{0N}(I) + \alpha I, \quad (3)$$

where α is the Kerr nonlinearity coefficient, and $\varepsilon_{0N} = \varepsilon_N(I = 0)$ is the unperturbed value of the permittivity of the nonlinear medium.

In nonlinear optics, a positive value of the Kerr nonlinearity coefficient is characteristic of media in which a self-focusing effect of light beam propagation is exhibited (focusing media), while a negative value models media characterized by a defocusing effect (defocusing media) [9]. The choice of the nonlinearity model in form (3) is primarily due to the fact that Eq. (1) with coefficient (3) has exact analytical solutions, the specific form of which is determined by the sign of the Kerr nonlinearity coefficient and additional conditions (boundedness, periodicity, etc.).

To model the spatial inhomogeneity of the refractive index in the graded-index medium in this work, we use a function of the form

$$\varepsilon_G(x) = e_0 + \frac{e_1}{x+h} + \frac{e_2}{(x+h)^2}, \quad (4)$$

where e_0, e_1, e_2, and h are the parameters of the spatial profile of the permittivity.

The choice of the graded-index medium model in form (4) is primarily due to the fact that Eq. (1) with coefficient (4) has exact analytical solutions [19], whose specific form is determined by the profile parameters and expressed through special functions of mathematical physics. In addition, the shape of the permittivity profile of form (4) is a generalization of the monotonically decreasing profiles of the refractive index that were previously used in modeling surface waves propagating along the contact of graded-index media with a Kerr nonlinear medium. In particular, at $e_0 = e_2 = 0$, expression (4) takes the form of the profile $\varepsilon_G(x) = e_1/(x + h)$ [20], and at $e_0 = e_1 = 0$, Eq. (4) becomes the profile $\varepsilon_G(x) = e_2/(x + h)^2$ [21].

A refractive index profile that sharply monotonically (but not exponentially) decreases within a narrow region was mentioned in the development of special photonic crystal heterostructures [22, 23]. A profile of form (4) is a more general version of the hyperbolic profile mentioned in those works. Selecting the parameters of profile (4) allows for a more accurate approximation of experimentally obtained profiles of a similar class. In this regard, interest arises in theoretically studying the possibility of constructing a crystalline system as a composite of the above types of optical materials, which could lead to the discovery of new properties.

Since the coefficient ε in Eq. (1) is generally discontinuous (with a jump), its solution can be represented as

$$u(x) = \begin{cases} u_G(x), x > 0 \\ u_N(I), x < 0, \end{cases} \quad (5)$$

where the sought functions $u_G(x)$ and $u_N(x)$ describe the spatial transverse distributions of the electric field strength in the graded-index and nonlinear media, respectively; and are defined on the corresponding half-axes.

As a result, instead of Eq. (1), taking into account (2)–(5), two equations can be written on the half-axes:

$$u_G''(x) + \left(e_0 + \frac{e_1}{x+h} + \frac{e_2}{(x+h)^2} - n^2 \right) \times \times k_0^2 u_G(x) = 0, x > 0, \quad (6)$$

$$u_N''(x) + \{\varepsilon_{0N} + \alpha |u_N|^2 - n^2\} k_0^2 u_N(x) = 0, x < 0. \quad (7)$$

The requirement for continuity of the components of the electromagnetic field when moving from one of the media to the other leads to the need to use boundary conditions of conjugation at the interface between the media at $x = 0$:

$$u_G(+0) = u_G(-0), \quad (8)$$

$$u_G'(+0) = u_N'(-0). \quad (9)$$

From here on, we mean one-sided limits $\lim_{x \to \pm 0} f(x) = f(\pm 0)$.

Since the field must tend to zero at infinity, the following conditions at infinity naturally arise:

$$\lim_{x \to \pm\infty} u_{G,N}(x) = 0. \quad (10)$$

Thus, the mathematical formulation of the model is a boundary value problem of finding continuous and bounded solutions of Eqs. (6) and (7) on the corresponding half-axes that are related by conjugation conditions (8) and (9) and satisfy the conditions at infinity (10).

2. ANALYTICAL RESULTS

Equation (6) is an equation with variable coefficients, and its solution bounded on the positive half-axis can be written in various ways: through a hypergeometric function, a confluent Heun function, or also a Whittaker function. In our opinion, the preferred

form of representing the solution is through the Whittaker function.

A solution, bounded on the positive half-axis, to the Whittaker differential equation [24]

$$y'' + \left(\frac{\mu}{z} - \frac{1}{4} + \frac{1/4 - \nu^2}{z^2} \right) y = 0$$

is the Whittaker function $W_{\mu,\nu}(z)$. By changing the variables, Eq. (6) is reduced to the Whittaker equation, and then its solution bounded on the positive half-axis can be written as

$$u_G(x) = u_0 \frac{W_{\mu,\nu}\left(p(x+h)\right)}{W_{\mu,\nu}(ph)}, \quad (11)$$

where u_0 is the amplitude of the electric field strength at the interface between the media, and the subscripts and parameters of the Whittaker function are determined by the coefficients of Eq. (6) as

$$\mu = \frac{e_1 k_0}{2\sqrt{n^2 - e_0}}, \quad (12)$$

$$\nu = \sqrt{1 - 4k_0^2 e_2} / 2, \quad (13)$$

$$p = 2k_0\sqrt{n^2 - e_0}. \quad (14)$$

For such a solution to exist, the conditions $n^2 > e_0$ and $k_0^2 > 1/4e_2$ must be met. Due to the choice of the Whittaker function, solution (11) satisfies the condition at infinity (10), which for it has the form $u_G(x) \to 0$ as $x \to +\infty$.

Equation (7) is a nonlinear differential equation. It has several types of solutions, depending on the sign of the nonlinearity coefficient α. Therefore, further analysis of the model is carried out separately for the cases $\alpha > 0$ and $\alpha < 0$.

(1) The case of focusing nonlinearity ($\alpha > 0$).

At $\alpha > 0$, a solution, bounded on the negative half-axis, to Eq. (7) satisfies the condition at infinity (10), which for it has the form $u_N(x) \to 0$ as $x \to -\infty$, and is expressed through the hyperbolic cosine:

$$u_N(x) = \sqrt{\frac{2}{\alpha}} \cdot \frac{q}{k_0 \cosh\left(q(x - x_N)\right)}. \quad (15)$$

Here,

$$q^2 = k^2(n^2 - \varepsilon_{0N}), \quad (16)$$

and the quantity x_N is the position of the maximum (if any) of the intensity of the surface wave in the nonlinear medium, which is determined from the boundary conditions.

To determine the field amplitude at the interface and x_N, solutions (11) and (15) should be substituted into boundary conditions (8) and (9). Such transformations give

$$u_0 = \sqrt{\frac{2}{\alpha}} \cdot \frac{q}{k_0 \cosh(qx_N)}, \quad (17)$$

$$x_N = \frac{1}{q} \operatorname{artanh}\left(\frac{q_G}{q} \right), \quad (18)$$

where

$$q_G = p \frac{W'_{\mu,\nu}(ph)}{W_{\mu,\nu}(ph)}. \quad (19)$$

Taking into account expression (18), from relation (17), we can obtain the field intensity at the interface between the media in the form

$$I_0 = |u_0|^2 = \frac{2}{\alpha}(n^2 - \varepsilon_{0N} - q_G^2 / k_0^2). \quad (20)$$

Thus, an everywhere continuous, smooth, bounded, vanishing at infinity solution to the posed boundary value problem in the case of positive nonlinearity can be written in the form

$$u(x) = \sqrt{\frac{2}{\alpha}} \cdot \begin{cases} \dfrac{W_{\mu,\nu}\left(p(x+h)\right)}{W_{\mu,\nu}(ph)}, & x > 0, \\ \dfrac{q}{k_0 \cosh\left(q(x - x_N)\right)}, & x < 0. \end{cases} \quad (21)$$

Expression (21) describes a nonlinear surface wave propagating along the interface between the nonlinear focusing and the graded-index media using the selected models of nonlinearity and spatial distribution of the refractive index.

(2) The case of defocusing nonlinearity ($\alpha < 0$).

At $\alpha < 0$, a solution, bounded on the negative half-axis, to equation (7) satisfies the condition at infinity (10), which for it has the form $u_N(x) \to 0$ as $x \to -\infty$, and is expressed through the hyperbolic sine:

$$u_N(x) = -\sqrt{\frac{2}{|\alpha|}} \cdot \frac{q}{k_0 \sinh\left(q(x - x_N)\right)}. \quad (22)$$

To determine the field amplitude at the interface and x_N, solutions (11) and (22) should be substituted into boundary conditions (8) and (9). Such transformations give

$$u_0 = \sqrt{\frac{2}{\alpha}} \cdot \frac{q}{k_0 \sinh(qx_N)}, \quad (23)$$

$$x_N = \frac{1}{q} \operatorname{arcoth}\left(\frac{q_G}{q} \right). \quad (24)$$

Taking into account (24), from (23), we can obtain the field intensity at the interface between the media in the form

$$I_0 = \frac{2}{|\alpha|}(q_G^2 / k_0^2 + \varepsilon_{0N} - n^2). \quad (25)$$

Thus, an everywhere continuous, smooth, bounded, vanishing at infinity solution of the posed boundary value problem in the case of negative nonlinearity can be written in the form

$$u(x)=\sqrt{\frac{2}{|\alpha|}}\cdot\begin{cases}\dfrac{W_{\mu,\nu}\left(p(x+h)\right)}{W_{\mu,\nu}(ph)}, & x>0,\\[2ex] \dfrac{q}{k_0\sinh\left(q(x-x_N)\right)}, & x<0.\end{cases} \qquad (26)$$

Expression (26) describes a nonlinear surface wave propagating along the interface between the nonlinear defocusing and graded-index media using the selected models of nonlinearity and spatial distribution of the refractive index.

3. RESULTS OF MODELING THE PROPERTIES OF SURFACE WAVES

First of all, we note the differences in the spatial distribution of the field intensity I in the surface waves propagating in the focusing and defocusing media in contact with the graded-index medium in the considered heterogeneity model.

Figure 1 shows the characteristic transverse intensity profiles constructed using solutions (21) and (26) at the same values of the optical parameters. Note that we chose $|\alpha| = 1$, specifically, $\alpha = 1$ for the case of the focusing medium (line *1*) and $\alpha = -1$ for the case of the defocusing medium (line *2*). All other parameter values are the same, and they are chosen so that their values fall within the ranges of existence of both surface waves.

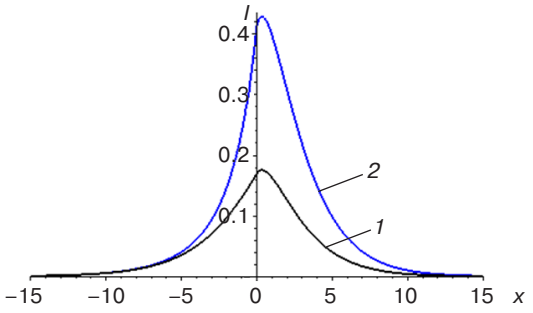


Fig. 1. Spatial distributions of the field intensity in the surface waves determined by expressions (21) (line *1*) and (26) (line *2*) at the following values of the system parameters (in conventional dimensionless units): $k_0 = 0.5$, $h = 0.5$, $e_0 = -0.1$, $e_1 = 0.8$, $e_2 = 0.2$, $\varepsilon_{0N} = 0.05$, $n = 0.44$, and $\alpha =$ (*1*) 1 and (*2*) −1

The surface waves are characterized by a single, clearly defined maximum in the spatial distribution of the field intensity, which decreases monotonically with distance from the maximum in both directions from the interface. In the selected range of parameter values, the maximum of the intensity for both types of surface waves is located in the graded-index medium. As will be shown below, in the case of defocusing nonlinearity, the maximum of the intensity is always located in the graded-index medium, while in the case of focusing nonlinearity, it can also be located in the nonlinear medium (not only in the graded-index medium), but at other parameter values. In other words, in the defocusing nonlinear medium, the field always decays strictly monotonically at any parameter values. In the focusing medium, the decay can be nonmonotonic, with an intensity spike near the interface, in which case the decay in the graded-index medium becomes monotonic.

The intensity peak height in the surface wave in the defocusing medium (Fig. 1, line *2*) significantly exceeds the intensity peak height in the surface wave in the focusing medium (Fig. 1, line *1*). For example, at the parameter values selected in Fig. 1, this excess is more than twofold. This results in a higher light intensity in the surface wave in the defocusing medium in comparison with the intensity in the surface wave in the focusing medium at the same distance from the interface.

Below, we will present the results of modeling the properties of the surface waves due to changes in the optical characteristics of the media. For the cases of focusing and defocusing media, this analysis is performed separately.

Note that varying the parameters of the models of nonlinearity (3) and graded-index media (4) means that different media correspond to different parameter values. Varying the effective refractive index n (which is equivalent to varying the propagation constant β) means changing the angle of incidence of the beam exciting the surface wave in a given medium. Varying the longitudinal wavenumber k_0 means changing the wavelength of the radiation exciting the surface wave (for example, the wavelength of a laser).

(1) The case of focusing nonlinearity ($\alpha > 0$).

Modeling using analytical solution (21) revealed that, depending on the effective refractive index, the surface wave propagating along the interface with the focusing nonlinear medium can be characterized by an intensity maximum located in both the nonlinear and the graded-index media. Specifically, at relatively low effective refractive index values, the intensity maximum is located in the graded-index medium. As the effective refractive index increases, the maximum shifts closer to the interface and then transitions into the nonlinear medium (Fig. 2), while its height also increases. The field penetration depth into the nonlinear medium increases; however, in the graded-index medium, it initially increases and then begins to decrease with increasing effective refractive index.

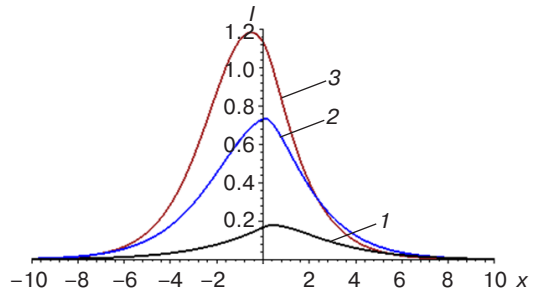


Fig. 2. Spatial distributions of the field intensity in the surface waves in the focusing medium (expression (21)) at the following values of the system parameters (in conventional dimensionless units): $\alpha = 1$, $k_0 = 0.5$, $h = 0.5$, $e_0 = -0.1$, $e_1 = 1$, $e_2 = 0.1$, $\varepsilon_{0N} = 0.05$; and n = (*1*) 0.5, (*2*) 0.65, and (*3*) 0.8

It can be seen that there are two distinct characteristic field intensity distributions in the surface wave propagating along the interface with the focusing nonlinear medium at two different values of the effective refractive index. One type corresponds to the maximum intensity in the graded-index medium (Fig. 2, line *1*), while the other corresponds to the maximum intensity in the nonlinear medium (Fig. 2, line *3*).

Next, we will analyze the change in the spatial distribution of the surface wave intensity at these two characteristic values of the effective refractive index when varying the parameters of the models of the media.

Figure 3 illustrates the effect of the parameter h of graded-index profile (4) on the spatial distribution of the surface wave intensity. With increasing h, the position of the intensity maximum shifts from the graded-index medium to the nonlinear medium. However, the shape of the field profiles depends on the effective refractive index. In particular, at small values of n (Fig. 3a), the field intensity is lower than that at its large values (Fig. 3b).

Figure 4 presents the effect of the parameter e_2 of graded-index profile (4) on the spatial distribution of the surface wave intensity. At small values of the effective refractive index (Fig. 4a), with increasing e_2, the position of the intensity maximum located in the graded-index medium remains unchanged, and its height decreases. At large values of the effective refractive index (Fig. 4b), with increasing e_2, the position of the intensity maximum located in the nonlinear medium shifts to the graded-index medium, and its height also decreases.

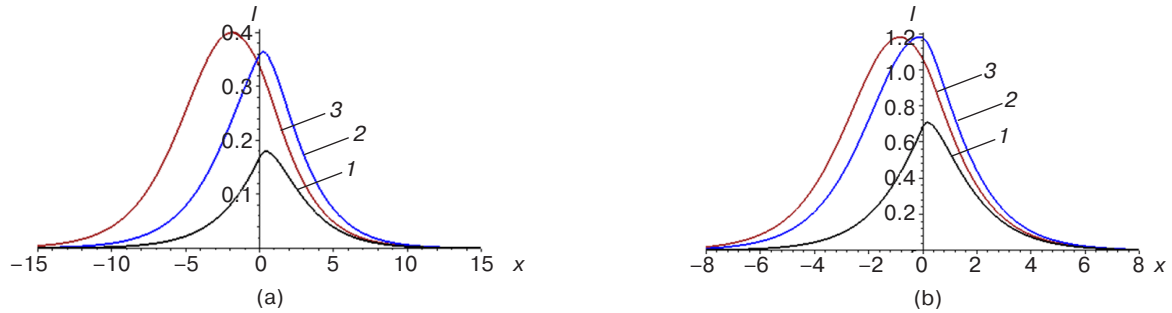


Fig. 3. Spatial distributions of the field intensity in the surface waves in the focusing medium (expression (21)) at the following values of the system parameters (in conventional dimensionless units): $\alpha = 1$, $k_0 = 0.5$, $e_0 = -0.1$, $e_1 = 1$, $e_2 = 0.2$, $\varepsilon_{0N} = 0.05$, (a) $n = 0.5$ and h = (*1*) 0.5, (*2*) 0.7, and (*3*) 1.5; and (b) $n = 0.8$ and h = (*1*) 0.2, (*2*) 0.4, and (*3*) 0.6

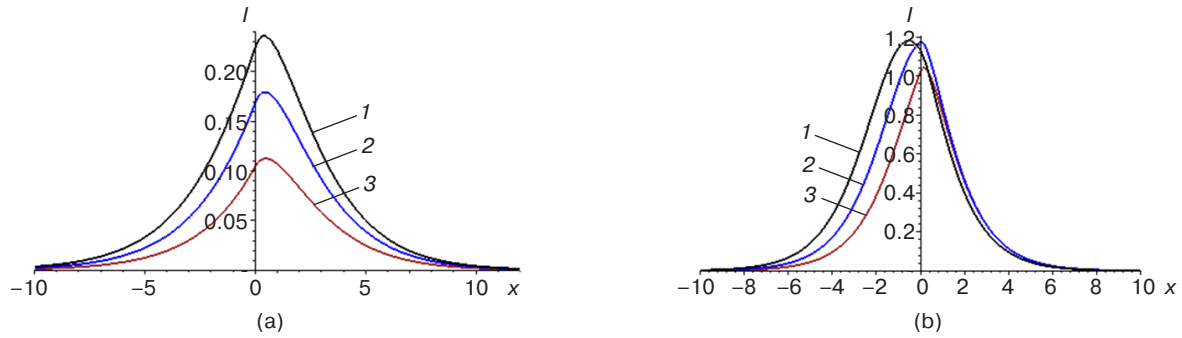


Fig. 4. Spatial distributions of the field intensity in the surface waves in the focusing medium (expression (21)) at the following values of the system parameters (in conventional dimensionless units): $\alpha = 1$, $k_0 = 0.5$, $e_0 = -0.1$, $e_1 = 1$, $h = 0.5$, $\varepsilon_{0N} = 0.05$, (a) $n = 0.5$ and e_2 = (*1*) 0.15, (*2*) 0.2, and (*3*) 0.25; and (b) $n = 0.8$ and e_2 = (*1*) 0.2, (*2*) 0.5, and (*3*) 0.8

Figure 5 shows the effect of the parameter e_1 of graded-index profile (4) on the spatial distribution of the surface wave intensity. At small values of the effective refractive index (Fig. 5a), with increasing e_1, the position of the intensity maximum located in the graded-index medium shifts into the depth of the graded-index medium, and its height decreases. At large values of the effective refractive index (Fig. 5b), with increasing e_1, the position of the intensity maximum located in the nonlinear medium shifts into the graded-index medium, and its height also decreases.

Figure 6 displays the effect of the parameter e_0 of graded-index profile (4) on the spatial distribution of the surface wave intensity. At small values of the effective refractive index (Fig. 6a), with decreasing e_0, the position of the intensity maximum located in the graded-index medium shifts into the depth of the graded-index medium, and its height increases. At large values of the effective refractive index (Fig. 6b), with decreasing e_0, the position of the intensity maximum located in the nonlinear medium shifts into its depth, moving away from the interface between the media, and its height does not change.

Figure 7 demonstrates the effect of the parameter ε_{0N} of nonlinear model (3) on the spatial distribution of the surface wave intensity. At small values of the effective refractive index (Fig. 7a), with increasing ε_{0N}, the position of the intensity maximum located in the graded-index medium does not change, and its height decreases. At large values of the effective refractive index (Fig. 7b), a similar effect is observed when the position of the intensity maximum is located in the nonlinear medium, but the decrease in the height of the maximum is less significant than in the case of small n in the same range of ε_{0N}.

Figure 8 presents the effect of the nonlinearity coefficient α of model (3) on the spatial distribution of the surface wave intensity. At small (Fig. 8a) and large (Fig. 8b) values of the effective refractive index, with increasing α, the position of the intensity maximum does not change, and its height decreases. However, at large values of n, the penetration depth of the field into the nonlinear medium decreases more significantly than at small n.

Figure 9 demonstrates the effect of the optical parameters of the model on the intensity I_0 of the surface wave at the interface between the media (expression (20)). It should be noted that the dependencies of the intensity of the surface wave at the interface on the optical parameters

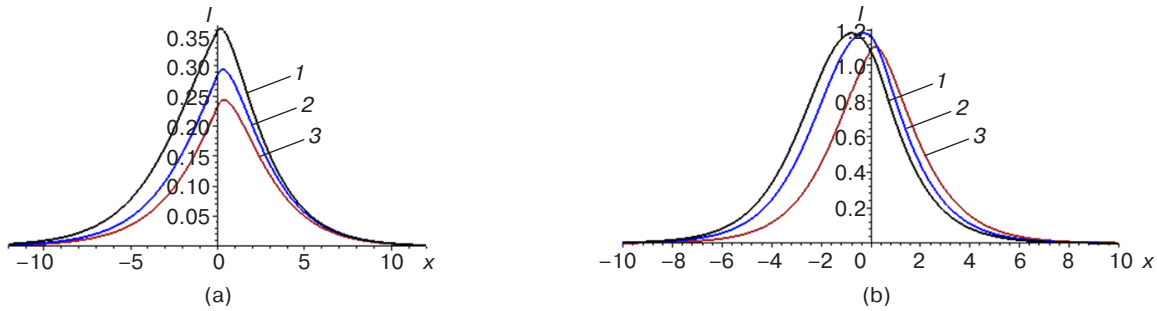


Fig. 5. Spatial distributions of the field intensity in the surface waves in the focusing medium (expression (21)) at the following values of the system parameters (in conventional dimensionless units): $\alpha = 1$, $k_0 = 0.5$, $e_0 = -0.1$, $e_2 = 0.2$, $h = 0.5$, $\varepsilon_{0N} = 0.05$, (a) $n = 0.5$ and $e_1 = (1)$ 0.8, (2) 0.9, and (3) 0.95; and (b) $n = 0.8$ and $e_1 = (1)$ 0.9, (2) 1.1, and (3) 1.5

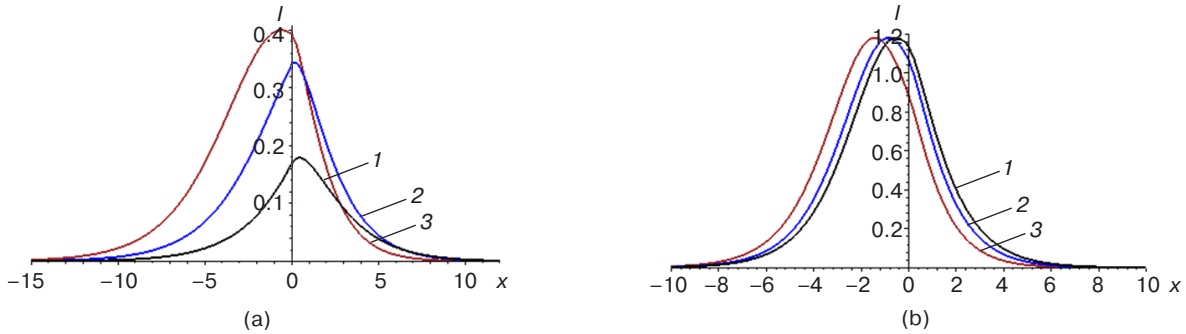


Fig. 6. Spatial distributions of the field intensity in the surface waves in the focusing medium (expression (21)) at the following values of the system parameters (in conventional dimensionless units): $\alpha = 1$, $k_0 = 0.5$, $e_1 = 1$, $e_2 = 0.2$, $h = 0.5$, $\varepsilon_{0N} = 0.05$, $e_0 = (1)$ −0.1, (2) −0.2, and (3) −0.4; and $n =$ (a) 0.5 and (b) 0.8

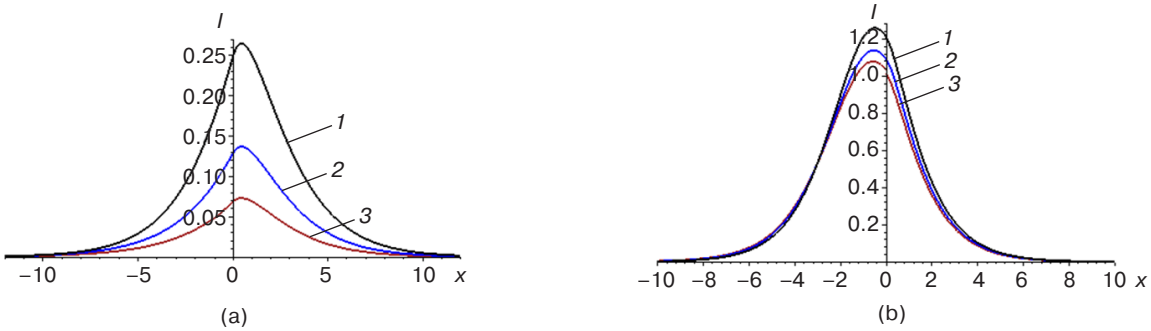


Fig. 7. Spatial distributions of the field intensity in the surface waves in the focusing medium (expression (21)) at the following values of the system parameters (in conventional dimensionless units): $\alpha = 1$, $k_0 = 0.5$, $e_0 = -0.1$, $e_1 = 1$, $e_2 = 0.2$, $h = 0.5$, ε_{0N} = (*1*) 0.01, (*2*) 0.07, and (*3*) 0.1; and n = (a) 0.5 and (b) 0.8

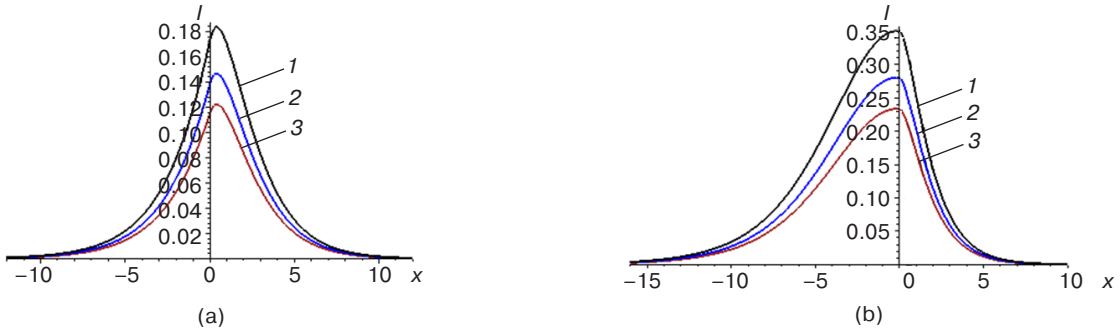


Fig. 8. Spatial distributions of the field intensity in the surface waves in the focusing medium (expression (21)) at the following values of the system parameters (in conventional dimensionless units): $k_0 = 0.5$, $e_0 = -0.1$, $e_1 = 1$, $e_2 = 0.2$, $h = 0.5$, $\varepsilon_{0N} = 0.05$, α = (*1*) 0.8, (*2*) 1, and (*3*) 1.2; and n = (a) 0.5 and (b) 0.8

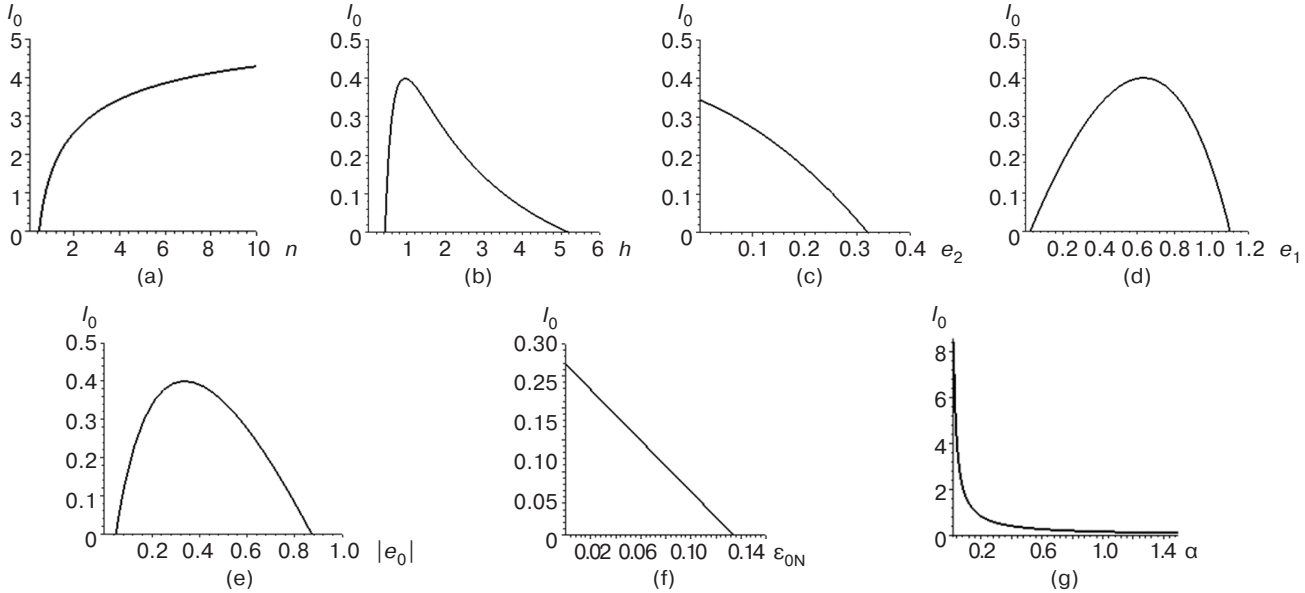


Fig. 9. Field intensity at the interface between the media (expression (20)) at the following values of the system parameters (in conventional dimensionless units): $k_0 = 0.5$; (a) dependence on n at $h = 0.5$, $e_0 = -0.1$, $e_1 = 1$, $e_2 = 0.2$, $\varepsilon_{0N} = 0.05$, and $\alpha = 1$; (b) dependence on h at $n = 0.5$, $e_0 = -0.1$, $e_1 = 1$, $e_2 = 0.2$, $\varepsilon_{0N} = 0.05$, and $\alpha = 1$; (c) dependence on e_2 at $n = 0.5$, $e_0 = -0.1$, $e_1 = 1$, $h = 0.5$, $\varepsilon_{0N} = 0.05$, and $\alpha = 1$; (d) dependence on e_1 at $n = 0.5$, $e_0 = -0.1$, $e_2 = 0.2$, $h = 0.5$, $\varepsilon_{0N} = 0.05$, and $\alpha = 1$; (e) dependence on $|e_0|$ at $n = 0.5$, $e_1 = 1$, $e_2 = 0.2$, $h = 0.5$, $\varepsilon_{0N} = 0.05$, and $\alpha = 1$; (f) dependence on ε_{0N} at $n = 0.5$, $e_0 = -0.1$, $e_1 = 1$, $e_2 = 0.2$, $h = 0.5$, and $\alpha = 1$; and (g) dependence on α at $n = 0.5$, $e_0 = -0.1$, $e_1 = 1$, $e_2 = 0.2$, $h = 0.5$, and $\varepsilon_{0N} = 0.05$

of the model are not all monotonic; i.e., at certain values of a number of optical parameters, I_0 has a maximum. Analysis of the modeling results showed that, with an increase in the effective refractive index, the intensity of the surface wave at the interface between the media increases monotonically (Fig. 9a). The intensity I_0 monotonically decreases with an increase in such optical parameters as e_2 (Fig. 9c), ε_{0N} (Fig. 9f), and α (Fig. 9g). Intensity maxima were found in the dependencies on such optical parameters as h (Fig. 9b), e_1 (Fig. 9d), and e_0 (Fig. 9e).

(2) The case of defocusing nonlinearity ($\alpha < 0$).

Modeling using analytical solution (26) revealed that the surface wave propagating along the interface with the defocusing nonlinear medium is always characterized by a single intensity maximum located in the graded-index medium. As the effective refractive index increases, its height decreases (Fig. 10), the field localization width decreases, and the position of the maximum shifts slightly toward the interface. This effect of the effective refractive index at the interface with the defocusing medium is opposite to the effect observed at the interface with the focusing medium (compare Figs. 3 and 10). Consequently, changing the angle of incidence of the beam exciting the surface wave has different (or, more accurately, opposite) effects on the field distribution in the surface waves in the focusing and defocusing media.

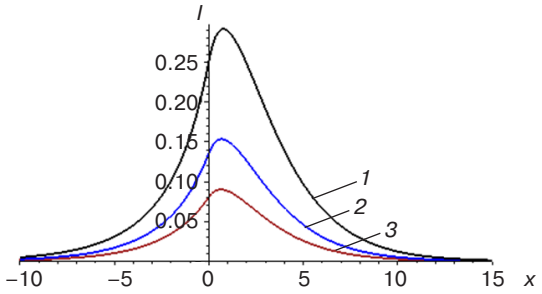


Fig. 10. Spatial distributions of the field intensity in the surface waves in the defocusing medium (expression (26)) at the following values of the system parameters (in conventional dimensionless units): $\alpha = -1$, $k_0 = 0.5$, $h = 0.5$, $e_0 = -0.1$, $e_1 = 1$, $e_2 = 0.2$, $\varepsilon_{0N} = 0.05$, and $n =$ (*1*) 0.42, (*2*) 0.44, and (*3*) 0.45

Figure 11 displays the effect of the parameter h of graded-index profile (4) on the spatial distribution of the intensity of the surface wave determined by solution (26). With increasing h, an effect similar to an increase in the effective refractive index is observed.

Figure 12 illustrates the effect of the parameter e_2 of graded-index profile (4) on the spatial distribution of the intensity of the surface wave determined by solution (26). With increasing e_2, the position of the intensity maximum located in the graded-index medium remains unchanged, its height increases, and the width of the field localization also increases.

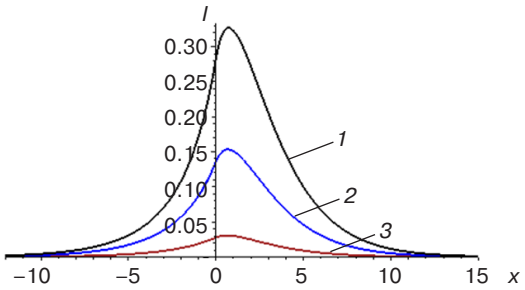


Fig. 11. Spatial distributions of the field intensity in the surface waves in the defocusing medium (expression (26)) at the following values of the system parameters (in conventional dimensionless units): $\alpha = -1$, $k_0 = 0.5$, $e_0 = -0.1$, $e_1 = 1$, $e_2 = 0.2$, $\varepsilon_{0N} = 0.05$, $n = 0.44$, and $h =$ (*1*) 0.45, (*2*) 0.50, and (*3*) 0.55

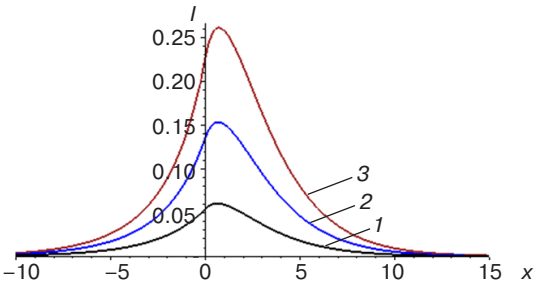


Fig. 12. Spatial distributions of the field intensity in the surface waves in the defocusing medium (expression (26)) at the following values of the system parameters (in conventional dimensionless units): $\alpha = -1$, $k_0 = 0.5$, $e_0 = -0.1$, $e_1 = 1$, $h = 0.5$, $\varepsilon_{0N} = 0.05$, $n = 0.44$, and $e_2 =$ (*1*) 0.15, (*2*) 0.2, and (*3*) 0.25

Figure 13 presents the effect of the parameter e_1 of graded-index profile (4) on the spatial distribution of the intensity of the surface wave determined by solution (26). With increasing e_1, an effect similar to an increase in the parameter e_2 of graded-index profile (4) is observed.

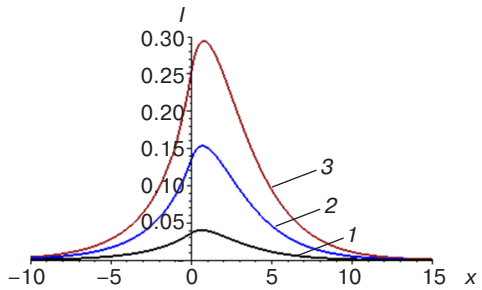


Fig. 13. Spatial distributions of the field intensity in the surface waves in the defocusing medium (expression (26)) at the following values of the system parameters (in conventional dimensionless units): $\alpha = -1$, $k_0 = 0.5$, $e_0 = -0.1$, $e_2 = 0.2$, $h = 0.5$, $\varepsilon_{0N} = 0.05$, $n = 0.5$, and $e_1 =$ (*1*) 0.95, (*2*) 1.0, and (*3*) 1.05

Figure 14 demonstrates the effect of the parameter e_0 of graded-index profile (4) on the spatial distribution of the intensity of the surface wave determined by solution (26). With decreasing e_0, an effect similar to an increase in the effective refractive index and the parameter h is observed.

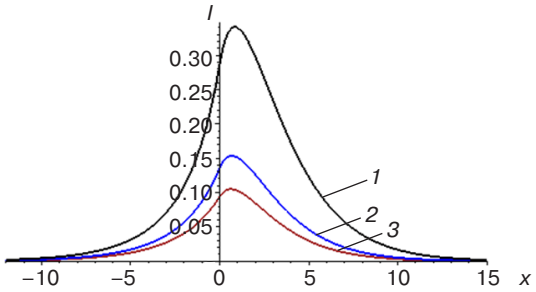


Fig. 14. Spatial distributions of the field intensity in the surface waves in the defocusing medium (expression (26)) at the following values of the system parameters (in conventional dimensionless units): $\alpha = -1$, $k_0 = 0.5$, $e_1 = 1$, $e_2 = 0.2$, $h = 0.5$, $\varepsilon_{0N} = 0.05$, $n = 0.44$, and e_0 = (*1*) −0.07, (*2*) −0.10, and (*3*) −0.11

Figure 15 shows the effect of the parameter ε_{0N} of nonlinear model (3) on the spatial distribution of the surface wave intensity determined by solution (26). With increasing ε_{0N}, an effect similar to an increase in the parameters e_1 and e_2 of graded-index profile (4) is observed.

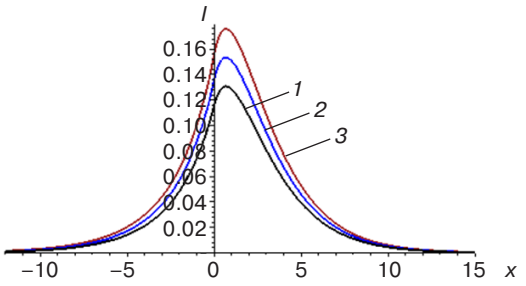


Fig. 15. Spatial distributions of the field intensity in the surface waves in the defocusing medium (expression (26)) at the following values of the system parameters (in conventional dimensionless units): $\alpha = -1$, $k_0 = 0.5$, $e_0 = -0.1$, $e_1 = 1$, $e_2 = 0.2$, $h = 0.5$, $n = 0.44$, and ε_{0N} = (*1*) 0.04, (*2*) 0.05, and (*3*) 0.06

Figure 16 illustrates the effect of the nonlinearity coefficient α of model (3) on the spatial distribution of the intensity of the surface wave determined by solution (26). With a decrease in the Kerr nonlinearity coefficient (i.e., with an increase in the absolute value of α), an effect similar to an increase in the effective refractive index and the parameter h is observed.

Thus, it is clear that the effect of the optical parameters of the model on the profiles of the spatial intensity distribution in the surface waves propagating along the interfaces depends significantly on the sign of the nonlinearity coefficient, and the observed effects are often opposite in the focusing and defocusing media.

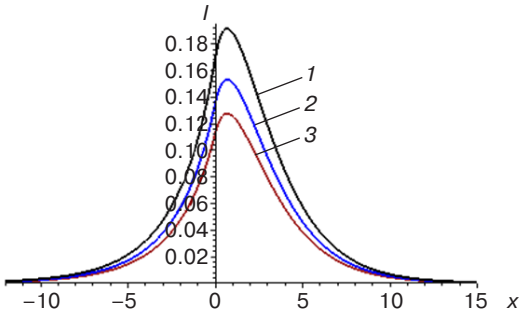


Fig. 16. Spatial distributions of the field intensity in the surface waves in the defocusing medium (expression (26)) at the following values of the system parameters (in conventional dimensionless units): $k_0 = 0.5$, $e_0 = -0.1$, $e_1 = 1$, $e_2 = 0.2$, $h = 0.5$, $\varepsilon_{0N} = 0.05$, $n = 0.44$, and α = (*1*) −0.8, (*2*) −1.0, and (*3*) −1.2

Figure 17 displays the effect of the optical parameters of the model on the intensity I_0 of the surface wave at the interface between the media (expression (25)). It should be noted that the dependencies of the intensity of the surface wave at the interface with the defocusing medium on the optical parameters of the model are all monotonic, unlike the case of contact with the focusing medium. Analysis of the modeling results showed that the surface wave intensity at the interface between the media decreases monotonically with increasing effective refractive index (Fig. 17a), parameter h (Fig. 17b), and absolute value of e_0 (Fig. 17e). The intensity I_0 increases monotonically with increasing optical parameters such as e_2 (Fig. 17c), e_1 (Fig. 17d), and ε_{0N} (Fig. 17f). In these cases, the effects observed in the defocusing medium are opposite to those observed in the focusing medium. The increase in the absolute value of the Kerr nonlinearity coefficient (Fig. 17g) is completely analogous to the increase in α in the focusing medium (Fig. 9g).

CONCLUSIONS

This paper presents the results of analytical modeling of surface shear waves propagating without loss along the interface between a nonlinear and a graded-index nonmagnetic medium. A linear dependence of permittivity on light intensity is chosen as a model of the nonlinearity of the medium to describe the nonlinear optical response of the medium to electric field perturbations. The graded-index medium is modeled using a special form of the spatial profile of permittivity as a function of distance to the interface for which an

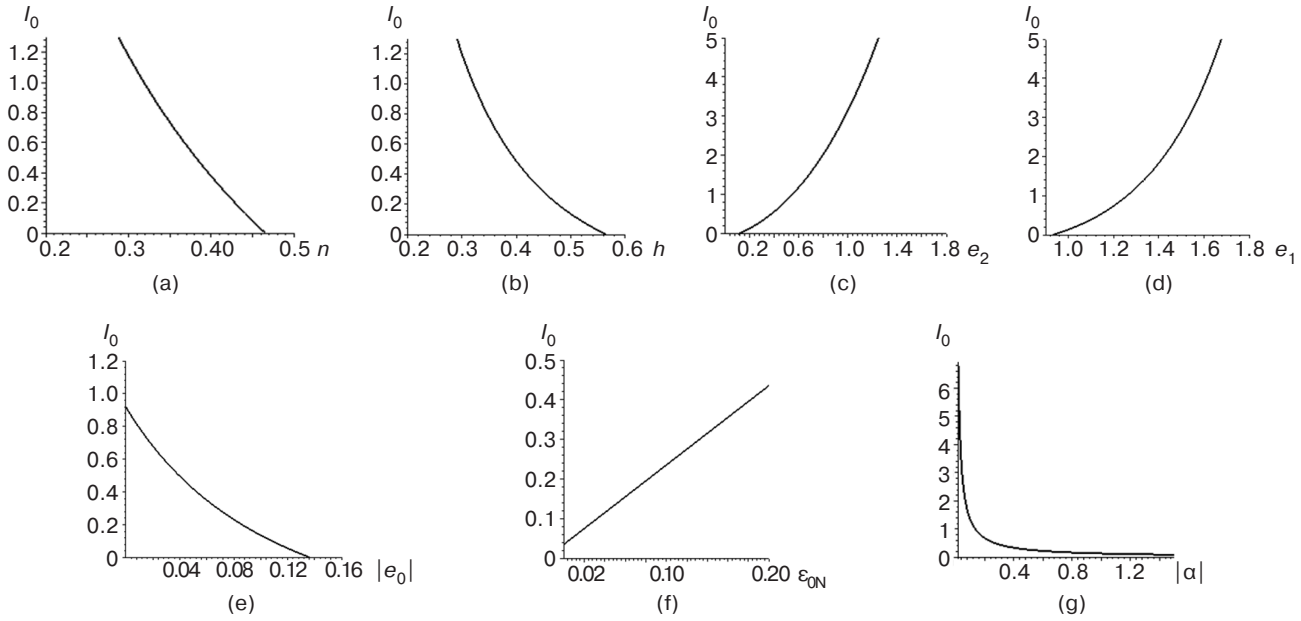


Fig. 17. Field intensity at the interface between the media (expression (25)) at the following values of the system parameters (in conventional dimensionless units): $k_0 = 0.5$;
(a) dependence on n at $h = 0.5$, $e_0 = -0.1$, $e_1 = 1$, $e_2 = 0.2$, $\varepsilon_{0N} = 0.05$, and $\alpha = -1$;
(b) dependence on h at $n = 0.44$, $e_0 = -0.1$, $e_1 = 1$, $e_2 = 0.2$, $\varepsilon_{0N} = 0.05$, and $\alpha = -1$;
(c) dependence on e_2 at $n = 0.44$, $e_0 = -0.1$, $e_1 = 1$, $h = 0.5$, $\varepsilon_{0N} = 0.05$, and $\alpha = -1$;
(d) dependence on e_1 at $n = 0.44$, $e_0 = -0.1$, $e_2 = 0.2$, $h = 0.5$, $\varepsilon_{0N} = 0.05$, and $\alpha = -1$;
(e) dependence on $|e_0|$ at $n = 0.44$, $e_1 = 1$, $e_2 = 0.2$, $h = 0.5$, $\varepsilon_{0N} = 0.05$, and $\alpha = -1$;
(f) dependence on ε_{0N} at $n = 0.44$, $e_0 = -0.1$, $e_1 = 1$, $e_2 = 0.2$, $h = 0.5$, and $\alpha = -1$; and
(g) dependence on $|\alpha|$ at $n = 0.44$, $e_0 = -0.1$, $e_1 = 1$, $e_2 = 0.2$, $h = 0.5$, and $\varepsilon_{0N} = 0.05$

exact analytical solution of the stationary wave equation can be found.

The presented mathematical formulation of the model consists in a conjugation boundary value problem for a nonlinear equation with variable coefficients. Exact analytical solutions to this boundary value problem are found for the cases of focusing and defocusing nonlinearities to describe the spatial distributions of the electric field strength in the direction transverse to the interface.

Analysis of the model revealed significant differences in the spatial distribution of the field intensity in surface waves propagating in the focusing and defocusing media. In the case of defocusing nonlinearity, the intensity maximum is always located in the graded-index medium, while in the case of focusing nonlinearity, it can be located in either the graded-index or the nonlinear medium, but at different parameter values. Furthermore, the light intensity in the surface wave in the defocusing medium is higher than the intensity in the surface wave in the focusing medium at the same distance from the interface and the same model parameter values.

The effect of the values of model parameters characterizing the optical properties of the contacting media on the spatial distribution of the light intensity in the surface waves was analyzed in detail. The modeling results showed that changing the same parameters has different (or, more accurately, opposite) effects on the field distribution in the surface waves in the focusing and defocusing media. Specifically, as the effective refractive index increases, the height of the intensity distribution maximum increases in the focusing medium, while it decreases in the defocusing medium.

In this paper, we chose a different model of contacting optical media than the one considered in our previous studies. The resulting new analytical solutions differ from those obtained previously, leading to differences in the properties of the surface waves they describe, particularly with regard to the sensitivity of their profile shapes to changes in the optical parameters of the media.

The obtained results, which supplement the existing theory of nonlinear and waveguide optics, can be applied in the design of new waveguide structures whose required dispersion properties are determined by the intensity of surface waves and by the optical characteristics of nonlinear and graded-index media. The new solutions obtained expand the class of exactly solvable models of planar waveguide structures with distributed inhomogeneous and nonlinear properties.

Authors' contributions

S.E. Savotchenko—conceptualization, methodology, analytical calculations, visualization, writing the manuscript.

N.O. Afanasyeva—numerical calculations, investigations, visualization, and writing the manuscript.

REFERENCES

1. Bogolyubov A.N., Mosunova N.A., Petrov D.A. Mathematical models of chirowaveguides. *Matematicheskoe modelirovanie = Mat. Model.* 2007;19(5):3–24 (in Russ.). https://www.elibrary.ru/hzwrxp
2. Svendsen B.B., Söderström M., Carlens H., Dalarsson M. Analytical and Numerical Models for TE-Wave Absorption in a Graded-Index GNP-Treated Cell Substrate Inserted in a Waveguide. *Appl. Sci*. 2022;12(14):7097. https://doi.org/10.3390/app12147097
3. Adams M.J. *An Introduction to Optical Waveguides.* Chichester: Wiley; 1981, 401 p.
4. Mihalache D. Localized structures in optical media and Bose-Einstein condensates: An overview of recent theoretical and experimental results. *Rom. Rep. Phys*. 2024;76(2):402. https://doi.org/10.59277/RomRepPhys.2024.76.402
5. Goyal A.K., Husain M., Massoud Y.Y. Analysis of interface mode localization in disordered photonic crystal structure. *J. Nanophoton*. 2022;16(4):046007. https://doi.org/10.1117/1.JNP.16.046007
6. Kubica J.M. Analysis of planar waveguides with a thin overlayer and nonlinear cladding. *Opt. Quant. Electron*. 2023;55(12):137. https://doi.org/10.1007/s11082-022-04390-4
7. Shvartsburg A.B. Dispersion of electromagnetic waves in stratified and nonstationary media (exactly solvable models). *Phys. Usp*. 2000;43(12):1201–1228. https://doi.org/10.1070/PU2000v043n12ABEH000827
[Original Russian Text: Shvartsburg A.B. Dispersion of electromagnetic waves in stratified and nonstationary media (exactly solvable models). *Uspekhi Fizicheskikh Nauk*. 2000;170(12):1297–1324 (in Russ.). https://doi.org/10.3367/UFNr.0170.200012b.1297]
8. Chen C-L. *Foundations for Guided-Wave Optics.* New York: John Wiley & Sons Inc.; 2005, 462 p. https://doi.org/10.1002/0470042222
9. Kivshar Yu.S., Agrawal G.P. *Optical Solitons: From Fibers to Photonic Crystals*. San Diego: Academic Press; 2003, 540 p.
10. Kudryashov N.A. Stationary solitons of the model with nonlinear chromatic dispersion and arbitrary refractive index. *Optik*. 2022;259(7):168888. https://doi.org/10.1016/j.ijleo.2022.168888
11. Liang N., Mihalache D., Ma M., Rao J., Liu Y. The multiple bright soliton pairs of the fully PT-symmetric nonlocal Davey-Stewartson I equation. *Rom. Rep. Phys.* 2024;76(2):106. https://doi.org/10.59277/RomRepPhys.2024.76.106
12. Kadantsev V.N., Goltsov A.N. Lateral proton transport induced by acoustic solitons propagating in lipid membranes. *Russian Technological Journal.* 2025;13(2):111–120. https://doi.org/10.32362/2500-316X-2025-13-2-111-120
13. Chernyaev A. P., Chernyaeva S. A. Degeneration of cnoidal waves into unbounded solutions for the Korteweg-De Vries equation. *Zhurnal radioehlektroniki = Journal of Radio Electronics.* 2018;6:8 (in Russ.). https://doi.org/10.30898/1684-1719.2018.6.5
14. Agrawal G.P. *Physics and Engineering of Graded-Index Media*. New York: Cambridge University Press; 2023, 348 p. https://doi.org/10.1017/9781009282086
15. Touam T., Yergeau F. Analytical solution for a linearly graded-index-profile planar waveguide. *Appl. Opt.* 1993;32(3): 309–312. https://doi.org/10.1364/AO.32.000309
16. Almawgani A.H.M., Taya S.A., Hussein A.J., Colak I. Dispersion properties of a slab waveguide *with* a graded-index core layer and a nonlinear cladding using the WKB approximation method. *J. Opt. Soc. Am. B.* 2022;39(6):1606–1613. https://doi.org/10.1364/JOSAB.458569
17. Savotchenko S.E. Models of waveguides combining gradient and nonlinear optical layers. *Russian Technological Journal*. 2023;11(4):84–93. https://doi.org/10.32362/2500-316X-2023-11-4-84-93
18. Savotchenko S.E. Models of symmetric three-layer waveguide structures with graded-index core and nonlinear optical liners. *Russian Technological Journal.* 2024;12(5):77–89. https://doi.org/10.32362/2500-316X-2024-12-5-77-89
19. Savotchenko S.E. Modeling of surface waves in photonic crystal structures with a refractive index profile decreasing with distance from the surface. *Russian Technological Journal.* 2026;14(1):91–102. https://doi.org/10.32362/2500-316X-2026-14-1-91-102
20. Savotchenko S.E. New surface waves in a hyperbolic graded-index crystal. *Rom. Rep. Phys*. 2024;76(4):406. https://doi.org/10.59277/RomRepPhys.2024.76.406
21. Savotchenko S.E. Surface waves in a medium with spatial monotonic attenuation of the refractive index. *Rom. Rep. Phys*. 2025;77(1):402. https://doi.org/10.59277/RomRepPhys.2025.77.402
22. Singh B.K., Bijalwan A., Pandey P.C., Rastogi V. Photonic bandgaps engineering in double graded hyperbolic, exponential and linear index materials embedded one-dimensional photonic crystals. *Eng. Res. Express*. 2019;1(2):025004. https://doi.org/10.1088/2631-8695/ab48a0
23. Singh B.K., Bambole V., Rastogi V., Pandey P.C. Multi-channel photonic bandgap engineering in hyperbolic graded index materials embedded one-dimensional photonic crystals. *Opt. Laser Technol*. 2020;129(17):106293. https://doi.org/10.1016/j.optlastec.2020.106293
24. Andrews G.E., Askey R., Roy R. *Special Functions*. UK: Cambridge University Press; 1999, 664 p. https://doi.org/10.1017/CBO9781107325937

СПИСОК ЛИТЕРАТУРЫ

1. Боголюбов А.Н., Мосунова Н.А., Петров Д.А. Математические модели киральных волноводов. *Матем. моделирование.* 2007;19(5):3–24. https://www.elibrary.ru/hzwrxp
2. Svendsen B.B., Söderström M., Carlens H., Dalarsson M. Analytical and Numerical Models for TE-Wave Absorption in a Graded-Index GNP-Treated Cell Substrate Inserted in a Waveguide. *Appl. Sci*. 2022;12(14):7097. https://doi.org/10.3390/app12147097
3. Adams M.J. *An Introduction to Optical Waveguides.* Chichester: Wiley; 1981, 401 p.
4. Mihalache D. Localized structures in optical media and Bose-Einstein condensates: An overview of recent theoretical and experimental results. *Rom. Rep. Phys*. 2024;76(2):402. https://doi.org/10.59277/RomRepPhys.2024.76.402

5. Goyal A.K., Husain M., Massoud Y.Y. Analysis of interface mode localization in disordered photonic crystal structure. *J. Nanophoton*. 2022;16(4):046007. https://doi.org/10.1117/1.JNP.16.046007
6. Kubica J.M. Analysis of planar waveguides with a thin overlayer and nonlinear cladding. *Opt. Quant. Electron*. 2023;55(12):137. https://doi.org/10.1007/s11082-022-04390-4
7. Шварцбург А.Б. Дисперсия электромагнитных волн в слоистых и нестационарных средах (точно решаемые модели). *Успехи физических наук* (*УФН*). 2000;170(12):1297–1324. https://doi.org/10.3367/UFNr.0170.200012b.1297
8. Chen C-L. *Foundations for Guided-Wave Optics.* New York: John Wiley & Sons Inc.; 2005, 462 p. https://doi.org/10.1002/0470042222
9. Kivshar Yu.S., Agrawal G.P. *Optical Solitons: From Fibers to Photonic Crystals*. San Diego: Academic Press; 2003, 540 p.
10. Kudryashov N.A. Stationary solitons of the model with nonlinear chromatic dispersion and arbitrary refractive index. *Optik*. 2022;259(7):168888. https://doi.org/10.1016/j.ijleo.2022.168888
11. Liang N., Mihalache D., Ma M., Rao J., Liu Y. The multiple bright soliton pairs of the fully PT-symmetric nonlocal Davey-Stewartson I equation. *Rom. Rep. Phys.* 2024;76(2):106. https://doi.org/10.59277/RomRepPhys.2024.76.106
12. Каданцев В.Н., Гольцов А.Н. Латеральный протонный транспорт, индуцированный распространением акустических солитонов в липидных мембранах. *Russian Technological Journal*. 2025;13(2):111–120. https://doi.org/10.32362/2500-316X-2025-13-2-111-120
13. Черняев А.П., Черняева С.А. Вырождение кноидальных волн в неограниченные решения для уравнения Кортевега – Де Фриза. *Журнал радиоэлектроники*. 2018;6:8. https://doi.org/10.30898/1684-1719.2018.6.5
14. Agrawal G.P. *Physics and Engineering of Graded-Index Media*. New York: Cambridge University Press; 2023, 348 p. https://doi.org/10.1017/9781009282086
15. Touam T., Yergeau F. Analytical solution for a linearly graded-index-profile planar waveguide. *Appl. Opt.* 1993;32(3): 309–312. https://doi.org/10.1364/AO.32.000309
16. Almawgani A.H.M., Taya S.A., Hussein A.J., Colak I. Dispersion properties of a slab waveguide *with* a graded-index core layer and a nonlinear cladding using the WKB approximation method. *J. Opt. Soc. Am. B.* 2022;39(6):1606–1613. https://doi.org/10.1364/JOSAB.458569
17. Савотченко С.Е. Модели волноводов, сочетающих градиентные и нелинейно-оптические слои. *Russian Technological Journal.* 2023;11(4):84–93. https://doi.org/10.32362/2500-316X-2023-11-4-84-93
18. Савотченко С.Е. Модели симметричных трехслойных волноводных структур с градиентной сердцевиной и нелинейно-оптическими обкладками. *Russian Technological Journal.* 2024;12(5):77–89. https://doi.org/10.32362/2500-316X-2024-12-5-77-89
19. Савотченко С.Е. Моделирование поверхностных волн в фотонных кристаллических структурах с профилем показателя преломления, убывающим с расстоянием от поверхности. *Russian Technological Journal.* 2026;14(1):91–102. https://doi.org/10.32362/2500-316X-2026-14-1-91-102
20. Savotchenko S.E. New surface waves in a hyperbolic graded-index crystal. *Rom. Rep. Phys*. 2024;76(4):406. https://doi.org/10.59277/RomRepPhys.2024.76.406
21. Savotchenko S.E. Surface waves in a medium with spatial monotonic attenuation of the refractive index. *Rom. Rep. Phys*. 2025;77(1):402. https://doi.org/10.59277/RomRepPhys.2025.77.402
22. Singh B.K., Bijalwan A., Pandey P.C., Rastogi V. Photonic bandgaps engineering in double graded hyperbolic, exponential and linear index materials embedded one-dimensional photonic crystals. *Eng. Res. Express*. 2019;1(2):025004. https://doi.org/10.1088/2631-8695/ab48a0
23. Singh B.K., Bambole V., Rastogi V., Pandey P.C. Multi-channel photonic bandgap engineering in hyperbolic graded index materials embedded one-dimensional photonic crystals. *Opt. Laser Technol*. 2020;129(17):106293. https://doi.org/10.1016/j.optlastec.2020.106293
24. Andrews G.E., Askey R., Roy R. *Special Functions*. UK: Cambridge University Press; 1999, 664 p. https://doi.org/10.1017/CBO9781107325937

About the Authors

Sergey E. Savotchenko, Dr. Sci. (Phys.-Math.), Associate Professor, Professor, High Mathematics Department – 3, Institute for Advanced Technologies and Industrial Programming, MIREA – Russian Technological University (78, Vernadskogo pr., Moscow, 119454 Russia). E-mail: savotchenkose@mail.ru. Scopus Author ID 6603577988, RSCI SPIN-code 2552-4344, https://orcid.org/0000-0002-7158-9145

Nadezhda O. Afanasyeva, Postgraduate Student, High Mathematics and Physics Department, Sergo Ordzhonikizdze Russian State University for Geological Prospecting (MGRI) (23, Miklukho-Maklaya ul., Moscow, 117997 Russia). E-mail: noafanasieva@mail.ru. https://orcid.org/0009-0009-0044-7930

Об авторах

Савотченко Сергей Евгеньевич, д.ф.-м.н., доцент, профессор кафедры высшей математики – 3, Институт перспективных технологий и индустриального программирования ФГБОУ ВО «МИРЭА – Российский технологический университет» (119454, Россия, Москва, пр-т Вернадского, д. 78). E-mail: savotchenkose@mail.ru. Scopus Author ID 6603577988, SPIN-код РИНЦ 2552-4344, https://orcid.org/0000-0002-7158-9145

Афанасьева Надежда Олеговна, аспирант, кафедра высшей математики и физики, ФГБОУ ВО «Российский государственный геологоразведочный университет имени Серго Орджоникидзе» (МГРИ) (117997, Россия, Москва, ул. Миклухо-Маклая, д. 23). E-mail: noafanasieva@mail.ru. https://orcid.org/0009-0009-0044-7930

Translated from Russian into English by K. Nazarov
Edited for English language and spelling by Thomas A. Beavitt

Economics of knowledge-intensive and high-tech enterprises and industries. Management in organizational systems

Экономика наукоемких и высокотехнологичных предприятий и производств. Управление в организационных системах

UDC 338.2
https://doi.org/10.32362/2500-316X-2026-14-3-131-144
EDN TTSRRP

RESEARCH ARTICLE

Planning of reproduction processes in electronics for a strategic perspective

Vasily V. Shpak @

MIREA – Russian Technological University, Moscow, 119454 Russia
@ *Corresponding author, e-mail: morser@yandex.ru*

• **Submitted:** 07.04.2026 • **Revised:** 19.04.2026 • **Accepted:** 27.04.2026

Abstract

Objectives. The work sets out to develop new approaches to strategic planning methodology in the electronics industry that take into account the current global multi-vector economic transformation and corresponding acceleration of technological changes. The most general framework for the further development of the electronics industry is outlined along with a discussion of its likely impact on the socioeconomic structure of future society.

Methods. Along with classical methods for studying socioeconomic relationships pertaining to complex systems such as the electronics industry, systems analysis, foresight technologies, production process modeling, and strategic planning methods are applied. In order to evaluate the explanatory power of different development models, particular attention is given to the multiple comparison method.

Results. The study identifies a need for a radical restructuring of economic and industrial sectors in which the electronics industry is accorded a leading role as a result of its development transcending traditional economic rationalizations to acquire a civilizational character. The establishment of the electronics industry as the main driving force for the entire economy, along with a corresponding restructuring of personnel training, thus represents a key strategic element in Russia's continuing development.

Conclusions. In the ongoing race for technological (and therefore economic, military, and political) sovereignty, the creation and continuous development of a strategic planning and management system for the electronics industry acquires a strategic character. The establishment of a genuinely independent national macroeconomic region can be achieved by ensuring self-sufficiency in both the production and consumption of electronic products.

Keywords: electronics industry, strategic planning, digitalization, artificial intelligence (AI), technological singularity, production processes, digital transformation

For citation: Shpak V.V. Planning of reproduction processes in electronics for a strategic perspective. *Russian Technological Journal.* 2026;14(3):131–144. https://doi.org/10.32362/2500-316X-2026-14-3-131-144, https://www.elibrary.ru/TTSRRP

Financial disclosure: The author has no financial or proprietary interest in any material or method mentioned.

The author declares no conflicts of interest.

НАУЧНАЯ СТАТЬЯ

Планирование воспроизводственных процессов в электронике на стратегическую перспективу

В.В. Шпак [@]

МИРЭА – Российский технологический университет, Москва, 119454 Россия
[@] *Автор для переписки, e-mail: morser@yandex.ru*

• **Поступила:** 07.04.2026 • **Доработана:** 19.04.2026 • **Принята к опубликованию:** 27.04.2026

Резюме

Цели. Цель работы – разработка новых подходов к методологии стратегического планирования в электронной промышленности с учетом глобальной многовекторной трансформации экономики и ускорения технологических изменений. Настоящее исследование показывает наиболее общие рамки дальнейшего развития электронной промышленности и рост ее влияния на социально-экономическое устройство будущего общества.

Методы. Наряду с классическими методами исследования социально-экономических отношений сложной системы, которой является электронная промышленность, применялись системный анализ, форсайт-технологии, моделирование производственных процессов и методы стратегического планирования. Особое внимание уделялось методу множественного сравнения для оценки объяснительной способности различных моделей развития.

Результаты. Проведенное исследование объективно подводит к необходимости новой реструктуризации отраслей экономики и промышленности с особым статусом для электронной промышленности, т.к. ее развитие выходит за рамки экономической рациональности и приобретает цивилизационный характер. Постановка электронной промышленности в качестве локомотива всей экономики наряду с реструктуризацией подготовки кадров – это залог стратегического успеха развития нашей страны.

Выводы. Создание и непрерывное развитие системы стратегического планирования и управления электронной промышленностью – единственная возможность «срезать углы» в развернувшейся гонке за технологическим, а значит экономическим, военным и политическим суверенитетом. Особое стратегическое значение приобретает необходимость построения собственного макрорегиона, самодостаточного как с точки зрения производства, так и потребления электронной продукции.

Ключевые слова: электронная промышленность, стратегическое планирование, цифровизация, искусственный интеллект, технологическая сингулярность, производственные процессы, цифровая трансформация

Для цитирования: Шпак В.В. Планирование воспроизводственных процессов в электронике на стратегическую перспективу. *Russian Technological Journal*. 2026;14(3):131–144. https://doi.org/10.32362/2500-316X-2026-14-3-131-144, https://www.elibrary.ru/TTSRRP

Прозрачность финансовой деятельности: Автор не имеет финансовой заинтересованности в представленных материалах или методах.

Автор заявляет об отсутствии конфликта интересов.

INTRODUCTION

Fundamental transformations occurring in the contemporary market economy are driven in part by changing consumption patterns and the crisis in the existing financial system. These developments present both challenges and new opportunities for the development of Russia's electronics industry. On January 22, 2026, the President of Russia, Vladimir Putin, held a meeting dedicated to the development of the Russian electronics industry, including issues related to the development and production of electronic component base and integrated circuits.[1] Following that meeting, the President of Russia decreed the creation of an interdepartmental commission to improve the efficiency of managing the development of the Russian electronics industry, which is headed by First Deputy Prime Minister D. Manturov and Presidential Aide A. Fursenko. Due to the strategic focus of the decisions to be adopted, executive bodies involved in developing the Russian economy will create a tiered planning system for the timely adjustment of strategic plans in light of scientific and technological achievements, which in practice emerge faster than strategic forecasts and foresight studies.

The key indicators reflecting the current state and development trends of the Russian electronics industry can be summarized in a few parameters. The gross output of the radio-electronic industry (REI) reached RUR 3.5 trillion in 2025. Despite a 34% increase in production, Russian electronics continued to lag behind major competitors by 25–35 years in terms of key technological characteristics. The most significant factor hindering the industry's development is a significant shortage of skilled personnel.

THEORETICAL FOUNDATIONS OF STRATEGIC PLANNING IN THE ELECTRONICS INDUSTRY

Methodological framework of the study

The contemporary economy is a multifaceted phenomenon characterized by a high degree of interconnectivity among its structures and components, including its supersystem, comprising the aggregate of natural factors and civilizational socioeconomic relations in their broadest sense. Thus, the wider economy serves as a supersystem relative to the electronics industry, which in turn is composed of various subsystems and subsectors. Therefore, a systemic approach to strategic planning and management of production processes in the electronics industry requires a coordinated analysis of the following factors: the crisis of the existing financial system, the regionalization of the economy, the formation of new economic blocs, changes in trade routes, the decline in effective demand, mass migration, the struggle for resources, etc. The necessity for—and approaches to—adapting production to these objective processes will be analyzed below.

A systems approach to the study of complex systems reveals the full set of their parameters, including interactions along the "supersystem–system–subsystems" axis. Since the fundamental task of strategic planning and management consists in monitoring the state of a complex system over time and responding proactively to challenges and changes, a multiple comparison method should be applied both to the entire set of parameters characterizing the system and to its dynamics. The essence of this approach consists in comparing selected parameters of the controlled system, which do not always have a strictly metric form, in a vector format at set intervals, such as a month, quarter, or year. With this approach, not only is the resulting indicator visible—such as changes in the industry's gross revenue or the dynamics of the topological component of the designed electronics—but also the factors that determined the overall change. These may include changes in the number of enterprises in the industry, the structure and physical volumes of output, price changes, etc. System analysis using the multiple comparison method can be effectively combined with artificial intelligence and big data technologies.

The rapid pace of change in both the hardware infrastructure of electronics, along with the constantly improving algorithms used to control it, underscores the critical importance of strategic foresight research, including both established and entirely new technological trends.

Current changes in strategic planning and management methodology can be summarized in comparison to established methods in several key areas. Since the digitization of management processes will become comprehensive in the very near future, the accuracy of forecasts and risk assessments can be expected to experience dramatic improvements. The widespread implementation of artificial intelligence systems will free management personnel from performing routine operations and significantly reduce the risk of errors. Meanwhile, the use of big-data technologies will allow decision-makers to operate not merely with selective aggregate indicators, but also taking into account primary production parameters available to sensors and cyber-physical systems in real time. Thus, with the optimization of industry-specific processes, management professionals will need to focus instead on solving new heuristic problems that arise.

[1] Meeting on the development of microelectronics. Vladimir Putin chaired a meeting on the development of Russian integrated electronics. http://www.kremlin.ru/events/president/transcripts/79027. (In Russ.). Accessed April 01, 2026.

Trends and factors driving the development of the electronics industry

The global transformation unfolding before our eyes is heightening the role of geostrategic factors that affect every aspect of civilized life. Among the most significant developments, particular attention should be paid to the global transformation of the division of labor and expertise, as well as the emergence of cooperation on the scale of large regional technology hubs. Compounding all this is the rapid development of digital technologies and their penetration into every aspect of socioeconomic relations.

Most experts agree that these changes have already begun [1]. The current economic crisis is objective and irreversible due to the market model of self-reproducing capital and aggressive financing having reached growth limits on a global scale. This is evident in the gradual dismantling of the global fiat dollar system. Today, money with its classic five functions no longer exists. As well as depriving Russia of the ability to use so-called reserve currencies, sanctions have revealed their fiat nature to other countries. Due to sanctions, Russia cannot settle payments using its sovereign funds, which it had placed with counterparties to diversify risks. Meanwhile, many friendly countries have been presented with incontrovertible evidence of the unsustainability of a global financial system controlled from a single country. As a result, regional payment systems are actively developing.

This, in turn, creates objective conditions for the regionalization of the real economy. While the global division of labor that emerged in the global electronics industry was economically efficient, it proved to too fragile to cope with systemic shocks. The emergence in the global economy of relatively self-sufficient currency-technology zones implies a clustering into separate macroregions, in which resources, markets, production, and technologies are starting to be concentrated. This process is accompanied by the practice of strict moratoriums on the export of high-tech solutions outside their zones. Since it is already clear that global international cooperation in the field of high technology, including electronics, is effectively impossible, the only viable way forward involves increased reliance on domestic strengths and cooperation models, including with reliable regional partners.

Thus, the strategic goal of Russia and its partners consists in the creation of a macroregion based on existing natural resources and domestic scientific expertise, as well as creative and industrial potential, that is maximally self-sufficient in terms of utilizing productive resources, and in which the population produces and consumes its own goods. Meanwhile, the Western approach to regional transformation is being implemented in an increasingly aggressive, forceful manner. This includes claims to the resources and markets of Latin America, Mexico, Canada, Greenland, and the Middle East, as well as ongoing attempts to maintain a military-monopolistic global hegemony.

Conversely, the framework of our Eurasian region is being shaped on a fundamentally different basis. This involves recognizing the objective interdependence of the economies of various countries and seeking avenues for integration based on the principles of mutual benefit, including strategic predictability and security. The Shanghai Cooperation Organization and BRICS[2] has shown itself to be capable of coordinating harmonious relations among the member states of these international associations to shape a new landscape of relations and create the necessary conditions for the emergence of independent macroregions (currency-technology zones) by leading countries (e.g., Russia, India, and China).

Forms of mutually beneficial cooperation are aimed at mitigating negative socioeconomic changes, such as a global decline in demand, the inevitable decline of the mass production model, the growing scarcity of all types of resources including even drinking water, as well as widespread environmental pollution. The formerly dominant market-based system of reproduction centered on mass consumption, the short life cycle of consumer goods, and the race for ostentatious status symbols, is clearly in the final stages of its existence. From the ruins of its gigantic globalist model, a new world and a new economy are already emerging.

A careful consideration of these factors along with other macroeconomic and geostrategic developments, which include changed migration patterns, includes a reassessment of the direction and depth of technological trends. Information and communication technologies in general—and microelectronics in particular—have already become a prerequisite for establishing a sovereign macroeconomic zone. However, the negative impact of geostrategic factors on the markets for new scientific and technological fields, such as neural networks, new communication technologies, and unmanned transport systems, continues to hold back their development as compared to the more optimistic predictions of experts [2, 3].

However, this can be explained in terms of new approaches to modeling production processes not yet having reached their full potential. Having accustomed themselves to operating under the conditions of fierce international competition, it is economically and psychologically difficult for manufacturers to shift toward cooperation. While competition according to

[2] BRICS—an acronym for Brazil, Russia, India, China, and South Africa—is an informal group of countries with rapidly growing economies.

the winner-takes-all approach has led to significant overconsumption of resources, cooperation, especially when combined with planning and coordination aimed at integrating various actors to achieve a synergistic result, can be applied to ensure reliable and effective industry connectivity. As a result, the private autonomy of manufacturers is increasingly giving way to transparency and traceability of the life cycle of electronic devices, from design and production to their sale, disposal, and subsequent recycling. This will be facilitated by the decision to mandate the labeling of all Russian electronic products, which will enable real-time visibility of the entire process.[3]

As experts from the Club of Rome[4] warned as early as the last century [4], the extensive phase of the market economy's development has now reached its limits. The top priority in the transition to a new, intensive economic model for the country consists in formulating strategic goals for the electronics industry and the economy as a whole. However, the formulation of such strategic goals involves several factors inherent in objective processes of production intensification.

The inevitable shift in production paradigms that is currently underway involves a movement away from large-scale, mass-market assembly-line production of electronic goods, which was focused on the mass consumer and the rapid obsolescence of electronic devices, towards flexible, reconfigurable automated systems and durable devices built on the principles of platform-modular architecture.

The paradigm shift involving the emergence and development of a qualitatively new socioeconomic model was described in detail in the works of the Soviet scholar N.D. Kondratiev [5]. The new post-capitalist paradigm must introduce a number of fundamental changes into the reproductive process. Industrial mass-production technologies should be replaced by those inspired by nature that improve the quality of life of citizens, including the production of high-quality goods that retain their consumer properties for a long time to ensure the rational and efficient use of resources at the same time as respecting the natural environment [6].

The strategic trend toward standardizing products and technologies based on open architectures must become a priority for government agencies. Such an approach will enhance the sovereignty of our future macroregion, strengthen cooperative ties, and significantly reduce the full lifecycle costs of the products being developed.

It is already clear that a revolution is also on the horizon in the refinement of manufacturing processes. While hundreds of operations are currently required to go from crystal growth to a finished electronic product, the transition to manufacturing finished electronic modules using various additive processes, as well as 3D printing of blocks and assemblies, is already being developed through research and development efforts by industry leaders and leading global research centers.

Electronics form the foundation of digital technologies, which in turn influence every stage of industrial production process, from strategic planning to the design of specific products. The development of digital technologies can be viewed as an intrinsic catalyst for the advancement of the electronics industry and electronic products.

An integral aspect of these and other objective socioeconomic processes consists in a concomitant transformation of consumer behavior. The ongoing economic crisis is exacerbated by various negative social shifts, including mass immigration, which in many countries has already become unmanageable and may turn out in the long term to be fatal. This includes a massive decline in the well-being of the so-called middle class and the general population. As purchasing power declines, manufacturers must shift to more flexible technologies that are responsive to scaling back production in line with actual effective demand. Meanwhile, mass immigration and the fierce global struggle for resources create conditions of social oppression and apathy, which have deleterious economic consequences.

The importance of the reputational assets of electronics manufacturers is rapidly growing. Given the global nature of world trade, rational consumer choice based on price-quality criteria is complemented by the equally important criterion of trust in the quality, durability, and efficiency of competing brands—or, more broadly, technological platforms. The strategic success of manufacturers will ultimately depend on consumer choice and their trust. Currently, competition is rapidly developing among several international technological platforms, which are backed by states forming sovereign technology zones. All of these platforms will need to become self-sufficient across the entire value chain, from the production of materials, chemicals, and design and manufacturing tools, to the development of digital intelligent technologies for collecting, processing, transmitting, storing, and analyzing information. Such challenges can be effectively addressed by the development of electronics products.

Meanwhile sanctions, local military conflicts, and artificially created tensions in various regions of the

[3] On Confirmation of Production of Russian Industrial Products. Resolution of the Government of the Russian Federation of July 17, 2015, No. 719. http://government.ru/docs/all/102816/. (In Russ.). Accessed April 01, 2026.

[4] The Club of Rome is an international public organization uniting representatives of the world's political, financial, cultural, and scientific elite.

world have disrupted supply chains, in some cases bringing production and consumption to a standstill. While this is a large, independent topic that requires separate consideration, we outline it here as a factor in the systematic understanding of current global trends.

Russian President Vladimir Putin's remark to journalists from the Financial Times[5] in 2019 that the liberal idea had outlived its purpose turned out to be prophetic: the currently unfolding global political crisis has compounded the already latent economic and social crises. Strategic planning in Russia must be based on the assumption that the economic, social, and political world will never be the same.

TECHNOLOGICAL ASPECTS OF STRATEGIC DEVELOPMENT OF THE ELECTRONICS INDUSTRY

The presented analysis confirms the necessity to develop new approaches to the technological aspects of production. Despite the rapid pace of technological development, the physical structure of global microchip consumption has remained virtually unchanged over the past five years, with the share of traditional solutions at the 350-nm level and above still standing at approximately 50%. The advanced level (below 28 nm) currently occupies approximately 16% of the market, which share has remained virtually unchanged since 2019. By 2030, it is projected to increase to only 17%. Thus, currently and in the coming years, most practical problems have been—and will continue to be—solved using topological standards of 350 nm and above. This, in turn, will impact the structure and quality of the industry's productive forces.

The increase in the average microchip production cycle is indicative. Up until 2020, the average cycle was approximately 30 days. However, with the decision to move beyond topological standards below 7 nm, the cycle began to lengthen to reach 69 days in 2025. This is evidence not only of the complexity of advanced technologies, but also their inherent instability.

Some experts believe that the minimal fabs concept developed by Shiro Hara [7] make it possible to compete with electronics giants, including by reducing the size of wafers. Such approaches are expected to gain momentum and become more competitive. This is due to the fact that the cost of modern giant silicon fabs, which focus on increasing wafer diameters and reducing design topology standards, reaches tens of billions of dollars and continues to rise. As products become more miniaturized, the cost of photomasks and lithographic processing equipment for wafers also increases. The cost of developing process technologies for new submicron technologies is also growing at an accelerated pace. At the present stage, a new superfab needs to sell wafers worth almost USD 20 million weekly for 10 years to ensure a return on investment. According to analysts at PricewaterhouseCoopers[6], the average cost of a chip created using 3-nm technology is 8 times higher than that of a 45-nm chip.

Since the Russian electronics market, especially under the conditions of sanctions, currently lacks the potential for global sales, a promising solution for the Russian electronics industry could involve the creation of low-cost and flexible production lines for processing wafers with diameters of 100–200 mm. Reducing the wafer diameter will lead to reduced fab productivity and increased processing costs per unit of wafer area. However, such solutions are becoming highly competitive for medium- and small-scale production, which currently does not fit into the superfab concept, as the cost of the line itself is reduced by orders of magnitude.

This type of production requires equipment that is modular and can operate in one or two shifts rather than continuously, which provides significant investment and operational savings. Currently, the uptime of advanced lithography systems from the Dutch company ASML[7] is approximately 50–60%, the remainder of the time being spent on maintenance and repairs. Since the cost of such machines is approximately USD 300–400 million, production costs are significantly impacted.

The mini-factory concept significantly lowers the barrier to entry, making this solution potentially attractive for export-focused development. Equipment for assembling integrated circuits and semiconductor devices in various package types can be easily integrated into the wafer production cycle. This results in highly efficient and flexible production systems, including a track for applying and developing photoresists, the corresponding test and measurement equipment, and a range of photoresists, developers, and anti-reflective coatings, to enable a complete cycle from wafer to packaged device. Accordingly, materials and chemistry, as well as design tools and other necessary industrial software, should be included in a packaged proposal for establishing such a facility, along with standard design and engineering solutions.

At the same time, electronic devices themselves are acquiring a platform-modular architecture, which enhances the competitiveness of such factories. As such

[5] http://www.kremlin.ru/events/president/news/60836. (In Russ.). Accessed April 01, 2026.

[6] International network of audit and consulting firms. https://www.pwc.com/gx/en.html. Accessed April 01, 2026.

[7] Manufacturer of lithographic equipment for the microelectronics industry. https://www.asml.com/en. Accessed April 01, 2026.

hardware solutions evolve, their workload becomes more predictable.

Electronic engineering, which became a strategic priority in 2021, already provides the foundation of the productive forces of the Russian electronics industry. For example, in 2024, testing of the first 350-nm projection lithography system, which was created in partnership with the Belarusian company Planar[8], was successfully completed. Two contracts for serial deliveries of these machines have already been signed. The next generation of 130-nm topology lithographs that use a Russian excimer laser should be ready by the end of this year. Meanwhile development of a 90-nm lithograph is already planned. In 2025, development of cluster plasma-chemical etching and deposition machines capable of handling 300-mm-diameter wafers was completed. The industrial strategy, which can be summarized as technology–equipment–chemistry–material, provides the foundation of the sovereignty of the Russian electronics industry.

Particular attention is being paid to the recycling of end-of-life products. The development of technologies for the renewal and recycling of resources holds significant promise. For example, the value of valuable materials recovered from electronic scrap could reach up to RUR 350 billion per year. The advantages of this form of organizing electronic production are significantly greater, and all of them must be considered within the context of strategic planning. The gradual digitalization of all production and management processes has entered the phase of industrial application of artificial intelligence and the creation of smart, adaptable production systems. Such technological breakthroughs are complemented by very significant organizational phenomena.

While artificial intelligence [8] is currently only a new, more advanced form of algorithmic processing of the accumulated knowledge of mankind, the prospects for its future application in production processes are difficult to fully imagine. It can become a comprehensive navigator, freeing specialists from processing already known facts and solutions, and opening up space for creative human activity. A striking historical analogy is the Luddite movement against the use of machines. However, history teaches that it was not machines that "crushed" workers, but workers who began to use machines for the purposes of improving their well-being. A similar situation is now developing with robotics. While the introduction of a fourth element—the "control device"—into the classical structure of the machine, consisting of engine, transmission device, and working element, also alarms many with prospect of the loss of their jobs, it is already clear that the robotization of production processes in the real sector of production will free up "human capital" for to create innovative solutions to the problems of human development arising within the sixth technological paradigm [9] or in Industry 4.0 [10].

A clearer understanding of the difference between artificial and "natural" human intelligence has now emerged. Artificial intelligence has received the adjective "artificial" because it currently processes the vast amounts of information accumulated by humanity in an efficient manner. Human natural intelligence, which has also been forced to engage in this routine work, has sought to obtain synergistic results from existing knowledge. While natural human abilities are already far inferior to computer algorithms in inductive fields, individual scientists have from time to time used deductive reasoning, which relies on an understanding of the subject of research in its entirety and diversity, to penetrate areas of fundamentally new and previously unknown knowledge. The geniuses of scientific history such as Hermes Trismegistus [11], Aristotle, Plato, Kepler, Einstein, Mendeleev, lay in their formulation of such insights in the familiar language of formulas and three-dimensional models. This example clearly demonstrates the division of functions between artificial and natural intelligence.

Thus, the strategic goal of electronics development can be formulated as the comprehensive development of computational and analytical computer programs for artificial intelligence to expand the scope of human creative activity while maintaining absolute control over the algorithms being created [12]. This technological trend will lead to a qualitatively new level of automation of the full cycle of production processes that optimizes the potential of cooperative links within the industry and between related industries. Overall production efficiency will increase thanks to multiparameter forecasting of both technological development and the market for final electronic products. Modern digital technologies and product labeling practices will lead to a leap in the quality of electronic products. Future "intelligent" self-reproducing systems will lead to the transformation of not only production, but also the entire set of socioeconomic relations.

STRATEGIC OBJECTIVES OF INDUSTRY MANAGEMENT TODAY

The methodological challenges of addressing strategic technological development objectives for the electronic industry are detailed in a dedicated monograph [12]. In practice, however, a tremendous amount of work lies ahead, the goal of which is to transform the industry management system into

[8] A developer of specialized technological equipment for the implementation of technologies in microelectronics. https://planar.by/. (In Russ.). Accessed April 01, 2026.

a system that can be broadly described as "Digital Twins of Enterprises and Organizations in the Electronic and Radio-Electronic Industries on the Platform of the Digital Twin of the Electronic Industry."

Modern digital information and analytical systems operate as network-centric systems with a distributed data storage protocol. This information exchange structure has proven its relevance and effectiveness. The ultimate goal of developing a system for collecting, storing, and processing information for the needs of government agencies should be the gradual creation of digital twins of enterprises and organizations in the real economy and digital twins of complex products on the industry digital twin platform. In such a system, all recurring information processes will be automated using the M2M (machine-to-machine) protocol, i.e., machine-to-machine communication.

It should be immediately noted that the transition to digital twin technologies in no way diminishes the importance of highly qualified managers. The Digital Twins hardware and software complex will undoubtedly free up a significant amount of time that managers currently have to spend on routine operations, ranging from collecting primary information to verifying and structuring it for the purpose of developing management decisions. Modern managers are sorely lacking this time to implement a comprehensive, all-encompassing optimization of industry reproduction and the formation of dynamic plans for solving the industry's strategic problems, such as eliminating the technological lag of Russian production, effective import substitution, and the timely implementation of the results of applied research and development, etc. In this area of work, neither the dominance of artificial intelligence nor the danger of technological singularity is visible. Without the participation of highly qualified managers, the management system at the mesolevel [13] will be nonviable.

At the base of the management pyramid will be digital twins of enterprises and organizations in the industry. Both human managers and digital twins will operate with arrays of production, business, and financial information on production in the electronics industry. Moreover, if we compare the arrays of information available for processing even by the most qualified managers with the arrays of information that a digital twin can handle, they are incomparably larger. And the gap will grow exponentially, but this is not a problem; in fact, it will be quite beneficial for decision makers, as they will be increasingly freed from routine tasks.

Previously, the object of strategic planning was the country's national economy. The reproduction cycle was calculated in two main aggregate formats: sectoral and territorial. This enabled, without the need for significant computational power, the calculation of two planned vectors of economic indicators, the comparison of which enabled for the verification or refinement of the methods and source data used. Repeated iterative recalculations were conducted, ultimately yielding an acceptable and plausible version of the strategic goal in specific indicators. This methodology was necessitated by the lack of modern digital tools. Currently, a significantly wider range of reproduction relations can be selected as the object of strategic planning. In its most general form, such an object can be called the "Electronics and Radioelectronic Industry Ecosystem."

The REI ecosystem can be conceptualized as a complex network of software-based and digitized actors, processes, technologies, and infrastructure that interact to create, produce, and distribute electronic products, including electronic components and REI. The ecosystem includes companies, research organizations, government institutions, suppliers, consumers, and other stakeholders.

A digital twin of an industrial sector, in this case the REI, is a computer model that enables planning, modeling, and optimization of the operations of enterprises in the industry and their interactions, including technological competition and production and business cooperation, and also ensures the dynamic formation of new supply chains for the fulfillment of complex orders [14, 15]. In computer terms, this is a platform that is formed on a powerful hardware and software complex, onto which digital twins of control objects are integrated.

A digital twin of a controlled object is a cyber-physical system augmented by a computer model of the controlled object, reflecting its state at any given moment and synchronized with the actual state of the physical object. A digital twin of a controlled object can be used for proactive planning and forecasting of the object's future behavior, as well as for modeling the object's behavior under any scenario. Types of digital twins of controlled objects are rapidly evolving from digital shadows of a real statistically monitored object to the formation of cognitive digital twins of enterprises and REI organizations, built on the basis of big data and artificial intelligence technologies.

A digital twin of an enterprise is an ontological twin of a real enterprise or organization, based on an ontological knowledge base.

Real electronics factories and their digital twins are linked into technological production chains, beginning with raw materials and ending with finished products. In the digital version, such chains should be represented as product graphs. This is a cybernetic object that digitally represents the relationships between product components located at the graph's vertices (objects) through edges (production and business relationships). The introduction of comprehensive labeling of

electronic products[9], coupled with the construction of product graphs, will not only enable quality control of the final product but also immediately identify those responsible for defects. This approach will enable the rapid integration of partial results of completed research and development (R&D) projects within product graphs, greatly accelerating the technological development of the industry and guaranteeing the quality of the REI products.

Thus, the digital system for preparing industry management decisions can be broadly termed the "Digital Model of the REI," which includes three interconnected modules: an information and reference system, an analytical simulation modeling unit, and a digital "intelligent" voice interface. As this system develops, intensive and extensive improvement processes are implemented.

This model is now generally referred to as a cyber-physical system. It is a computer system that integrates tools for obtaining primary data from databases and sensors; computing tools for performing computer calculations, transmitting and processing data, and controlling physical processes using actuators; and communication tools for interaction with users and decision-makers.

ORGANIZATIONAL CHANGES

Strategic planning for such a specific industry as the electronics industry has another important aspect. Russian industry is an extremely complex reproduction system, and traditional manual management methods do not provide a synergistic increase in efficiency or a reduction in the complexity of management information flows. Therefore, quantitative changes have accumulated that must be transformed into a qualitative leap in the meso- and macroeconomic management system [16].

Within the complex Russian economy, distinct sectors (transportation, communications, finance, etc.) have cross-cutting functionalities across the entire economy. The electronics industry has similar functions, but it is integrated into the conglomerate known as industry and trade.

The macroeconomic role of electronics has increased immeasurably compared to the 1990s, when it was consolidated into an independent ministry. In the strategic perspective, electronics will become a dominant factor in all spheres of government. These trends indicate the need to separate the REI into an independent industry, as was done in the Soviet Union's economy in the 1960s. At that time, this organizational measure, coupled with the development and approval of the General Plan for the Development and Allocation of Productive Forces for the Strategic Perspective, allowed the country's economy to achieve parity with world leaders in electronics [17]. This experience should not be neglected.

Industry self-determination of the electronics industry will enable, within the framework of big data technology, the creation of dynamically updated classifiers, starting from a closed list of enterprises and organizations of the REI, as well as classifiers of equipment, personnel, and products both in a subindustry context and with an understanding of the localization of their production thanks to the product labeling system currently being deployed.

SOME ASPECTS OF STRATEGIC PLANNING AND MANAGEMENT METHODS

Data collection and processing system

The implementation of such a large-scale management system must be preceded by a full-scale inventory of enterprises and organizations in the industry [18]. As a reliable database and knowledge base are created in the industry, a digital twin of the REI should be created and approved for training. As REI enterprise managers master digital management technologies, REI enterprises and organizations of all types of ownership and legal entities will be connected to this platform. The synergistic effect of such organizational transformations will be achieved already at the stage of creating a system incorporating 50–100 of the largest enterprises in the industry, and will increase as the management and information integration of REI enterprises progresses.

In existing regulatory documents defining the foundation of economic management, the concept of information, which is the starting point of any management decision, is used too generally, without a strict topology (the format of each indicator) and metrics (attributes reflecting the volume or dynamics of the processes characterized by a specific indicator). Therefore, it is necessary to clarify that information for the purposes of modeling decision-making processes in the industry must be *reliable*, *comparable*, and *equatable*, and the methodology for obtaining information from the primary source must be the same for all aggregated indicators [19].

[9] Resolution of the Government of the Russian Federation of November 28, 2025 No. 1954 "On approval of the Rules for marking certain types of radio-electronic products with identification means and the specifics of the implementation of the state information system for monitoring the circulation of goods subject to mandatory marking with identification means, in relation to certain types of radio-electronic products." http://publication.pravo.gov.ru/document/0001202511290036. (In Russ.). Accessed April 01, 2026.

Without fulfilling these basic mandatory conditions, the aggregation of indicators of enterprises and organizations into general industry indicators is inadequate from a scientific point of view and contains systemic errors, the multiplication of which makes information aggregated from several noncomparable indicators unreliable.

Without delving into the intricacies of statistical science, we will highlight only the most general, yet absolutely essential, requirements for the collected production, economic, and financial information. Primary information is information obtained directly from the object of statistical observation. Previously, it was collected by the relevant departments of enterprises and primary statistical offices throughout the country. Now, it is possible to automate the collection of any primary production, economic, and financial information using so-called cyber-physical agents.

Only aggregated information that can be verified based on the primary data obtained directly from the object of statistical observation can be considered *reliable*. For the needs of industry management, primary information is aggregated into summary indicator formats, but the "ancestral trail" of each summary indicator can always be traced back to the enterprise or organization level. An aggregated industry indicator must be equal to the sum of similar indicators for enterprises and organizations within the industry, down to the last decimal point. If a deviation occurs, a special review of the causes of the discrepancy is required. This is possible. Therefore, the indicator is called aggregated, not summary. However, the deviation and its causes must be rigorously statistically assessed and, if necessary, adjusted.

Aggregated information obtained using a unified methodology for collecting and processing primary data can be considered *comparable*. It is strictly forbidden to aggregate, for example, indicators obtained directly through statistical observation and indicators obtained through an analytical method or a random survey method. This creates systematic statistical errors, which are unacceptable when making management decisions in general and strategic decisions in particular.

Comparability of information characterizes the inherent properties—the quality of the information itself—while comparability of aggregated information demonstrates the admissibility and correctness of comparing similarly named indicators in a time series. This is necessary for recognizing extensive and intensive factors influencing aggregated indicators. For example, the number of electronics companies changes from year to year, and the performance of each change. Therefore, it is necessary to identify the parts of the indicator change (increase or decrease) associated with changes in the number of objects of statistical observation and with the factor of improvement or deterioration in the performance of enterprises. Furthermore, all aggregated indicators must be recorded on the same date of receipt of the primary information and accounted for a comparable accounting period. These are only the most general, unavoidable requirements for management information. More precise statistical settings require a more thorough analysis of the primary indicators.

Statistical observation has many other important limitations and criteria, which we will leave to specialists. Here, we will simply highlight the problem that for planning, especially at a strategic depth, the indicators, and indeed all information used in the simulation model, must be as relevant as possible to the actual state of each enterprise and the industry as a whole, in both metric and dynamic terms.

Knowledge base and ontologies

The ability of modern algorithms (so-called large language models) to recognize and process not only numbers but also text is based on the scientific method of information ontology. This is a methodology for the ontological description of the qualitative parameters of phenomena, processes, or structures, such as equipment, enterprises, and organizations in the REI, with an emphasis on key properties in the form of actors, collections, structure, and attributes, which are used in subsequent modeling [20].

Agents of a multiagent model

Digital technologies make it possible to consider an ever-increasing number of factors, parameters, external disturbances, etc., integrating them into a strategic planning model by connecting an ever-increasing number of modules (agents). An agent (module) is autonomous software capable, within the framework of its assigned functions, of analyzing collected information, optimizing decisions, interacting and negotiating with other agents (M2M), informing decision-makers of the results of its actions, issuing recommendations, and incorporating feedback from all human and automated counterparties. An agent is capable of not only performing once-programmed tasks but also of continuously learning. Digital technologies make it possible to automate the entire cycle of routine and repetitive factors and, through combinatorial analysis, calculate a multitude of possible final values.

Visualization tools

Having almost all the metric and ontological information about the reproduction process in the industry, the control subsystem can generate visual

graphs that will also be used in the procedures of strategic management of the REI.

ANTICIPATED RESULTS

As a result of the phased implementation of the digital twin of the REI, the following results are expected:

1. The system will be secure thanks to a layered user authorization subsystem.
2. A digital industry corporate platform will be created for interaction between all authorized participants in reproduction in electronics, which can be collectively referred to as the State Information System of Industry Electronics (SISIE).
3. A dynamically maintained digital classifier of REI enterprises will be created, whose relevance will be continuously maintained by an "intelligent" agent.
4. A digital classifier of REI will be created, broken down into the main subsectors, and linked to the mechanism for labeling products and their components.
5. A classifier of REI technologies will reveal the relative technological levels of Russian and foreign manufacturers and their dynamics online.
6. A dynamically supported classifier of REI personnel will be created, linked by digital algorithms to existing systems for training personnel, improving their qualifications, and personal career plans.
7. A reliable and comprehensive statistical database on REI will be created, along with a tool for its "intelligent" verification, processing, and visualization.
8. A digital "intelligent" agent will be put into operation, which will verify all information in the system online 24/7.
9. The analytical unit in the format of a digital twin model of the REI will calculate all possible consequences of management and technological decisions and propose several scenarios of strategic actions to optimize the final result.
10. Thanks to digitalized product graphs, the digital twin model of the REI will continuously record the balance/imbalance not only for the industry as a whole, but also for individual production items.
11. Product graphs will not only enable the current competitive cost of REI products to be maintained through strict alignment of physical production proportions, but also support online cost-effectiveness analysis of price versus quality, which will reduce the burden on the budget.
12. By expanding the digital twin model of the REI to function as a general plan for the development and deployment of productive forces, a layered linkage of R&D and production over a time horizon of 10–15 years will be facilitated.
13. All industry aggregated indicators will be programmatically linked to the actual primary indicators of enterprises and organizations of the REI. This will enable the targeted identification of those responsible for production failures and the elimination of the causes.
14. A comprehensive, general indicator of the efficiency of the electronic product reproduction cycle will not only include production, but also the entire socially significant chain of "product development (resource and material support for its reproduction) that best meets consumer requirements, including operational and repair support, as well as environmentally safe disposal."

CONCLUSIONS

Following its implementation, the conceptual version of the presented strategic planning and management model developed in Russia, which proactively responds to emerging risks and challenges, will provide a means for overcoming the technological backwardness of Russian electronics, enable the development of the industry, tackle the shortage of highly qualified personnel, and move away from dependence on imports of critical electronics in the context of sanctions aggression.

The general plan for the development and deployment of REI's productive forces will enable effective R&D and the development of in-house technologies that are integrated with the training of qualified personnel focused on new technological advances. It will create conditions for the formation, development, and strengthening of a macroeconomic region of friendly countries, facilitating the emergence of alternative production and supply chains, and thereby increasing the scale and efficiency of production. The described model for the digitalization of industry has the potential to be scaled up for the purposes of managing the entire economy.

REFERENCES

1. Schwab K., Malleret T. *Covid-19*: *Velikaya perezagruzka* (*Covid-19*: *The Great Reset*). Forum; 2020, 262 p. (In Russ.).
2. Vinge V. *Singulyarnost'* (*Singularity*): transl. from Engl. Moscow: AST; 2019, 224 p. (In Russ.). ISBN 978-5-17-114349-7
3. Kurzweil R. *Revolyutsiya razuma*: *na podstupakh k Singulyarnosti. Kak tekhnologii izmenyat obshchestvo i soznanie* (*Revolution of the Mind*: *on the Approaches to the Singularity. How Technologies Will Change Society and Consciousness*). Moscow: Eksmo, BOMBORA; 2025. 661 p. (In Russ.). ISBN 978-5-04-228765-7
4. Peccei A. The Challenge of the 1970s to the Modern World. Topic for Discussion. In: Gvishiani D.M. (Ed.). *The Club of Rome. History of Creation, Selected Papers and Speeches, Official Proceedings*. Moscow: URSS; 1997, 384 p. (In Russ.). ISBN 5-88417-100-5
5. Kondrat'ev N.D. *Bol'shie tsikly kon"yunktury i teoriya predvideniya* (*Large Cycles of Economic Conditions and the Theory of Foresight*): International Fund of Kondratiev N.D., et al. Moscow: Ehkonomika; 2002, 550 p. (In Russ.). ISBN 5-282-02181-1
6. Koval'chuk M.V. *Ideologiya prirodopodobnykh tekhnologii* (*Ideology of Nature-Inspired Technologies*). Moscow: Fizmatlit; 2021, 329 p. (In Russ.). ISBN 978-5-9221-1931-3
7. Hara Sh. *Minimal FAB. Fab System Research Consortium*. AIST, Japan. 2010.
8. Malenkov Yu.A., Davydova A.E. Possibilities of applying digital technologies in strategic planning and forecasting of sustainable development of organizations. *Molodoi uchenyi = Young Scientist*. 2021;14(356):84–89 (in Russ.). https://elibrary.ru/zrjnia
9. Glaz'ev S.Yu. New approaches to the organization of enterprises in the context of changing technological and world economic modes. In: Kleiner G.B. (Ed.). *Strategic Planning and Evolution of Enterprises:* Materials of the 20th Russian Symposium. Moscow: CEMI RAS; 2019. P. 7–18 (in Russ.).
10. Shpak V.V. *Razvitie elektronnoi promyshlennosti v usloviyakh menyayushchegosya mira* (*Development of the Electronics Industry in a Changing World*). Moscow: Tekhnosfera; 2024, 128 p. (In Russ.). ISBN 978-5-94836-708-8
11. Dlyasin G.G. *Azbuka Germesa Trismegista ili molekulyarnaya tainopis' myshleniya* (*The ABCs of Hermes Trismegistus, or the Molecular Secret Writing of Thinking*). Moscow: Belye al'vy; 2002, 141 p. (In Russ.). ISBN 5-7619-0126-9
12. Shpak V.V. Strategic planning in modern conditions. In: *Public Administration in New Geopolitical and Geoeconomic Conditions*: Proceedings of the 19th International Conference. Moscow State University, December 5–9, 2022. Moscow: KDU, Dobrosvet; 2023. P. 130–136 (in Russ.). https://elibrary.ru/lehfav
13. Maevskii V.I. (Ed.). *Mezoekonomika*: *sostoyanie i perspektivy*: monografiya (*Mesoeconomics: State and Prospects*: monograph). Moscow: Institute of Economics, RAS; 2018, 314 p. (In Russ.). ISBN 978-5-9940-0642-9
14. Skobelev P.O. Multi-agent technology for industrial applications: towards 20 years anniversary of Samara scientific school of multi-agent systems. *Mekhatronika, Avtomatizatsiya, Upravlenie*. 2010;12:33–46 (in Russ.). https://elibrary.ru/nblbbh
15. Shpak V.V. Fundamentals of a multi-agent planning and logistics model for managing the reproduction cycle of the industry in the "online" mode. *Fundamental'nye issledovaniya = Fundamental Research*. 2022;3:146–155 (in Russ.). https://doi.org/10.17513/fr.43229
16. Shpak V.V. Formation of the organizational and managerial model for the implementation of the "Strategy for the Development of the Electronics Industry of the Russian Federation for the period up to 2030." *Vestnik Chelyabinskogo gosudarstvennogo universiteta = Bulletin of Chelyabinsk State University*. 2021;3(449):10–23 (in Russ.). https://doi.org/10.47475/1994-2796-2021-10302
17. Shokin A.A. *Ministr neveroyatnoi promyshlennosti SSSR* (*Minister of Incredible Industry of the USSR*). Moscow: Tekhnosfera; 2007, 455 p. (In Russ.). ISBN 978-5-94836-151-2
18. Shpak V.V. Electronics inventory as the foundation of breakthrough into global leadership. *Russian Technological Journal.* 2025;13(4):135–148. https://doi.org/10.32362/2500-316X-2025-13-4-135-148
19. Simchera V.M. *Izoblichaya lzhivye tsifry. Etyudy o pirrovykh pobedakh i poteryakh Rossii* (*Exposing False Figures. Studies on Russia's Pyrrhic Victories and Losses*). Izdatel'skie resheniya; 2019, 357 p. (In Russ.). ISBN 978-5-4496-4788-7
20. Shvedin B.Ya. *Ontologiya predpriyatiya: ehkspirientologicheskii podkhod. Tekhnologiya postroeniya ontologicheskoi modeli predpriyatiya na osnove analiza i strukturirovaniya zhivogo opyta* (*Enterprise Ontology: An Experiential Approach. Technology for Constructing an Enterprise Ontological Model Based on the Analysis and Structuring of Lived Experience*). Moscow: LENAND; 2010, 240 p. (In Russ.). ISBN 978-5-9710-0308-3

СПИСОК ЛИТЕРАТУРЫ

1. Шваб К., Маллере Т. *Covid-19*: *Великая перезагрузка*. Форум; 2020, 262 с.
2. Виндж В. *Сингулярность*: пер. с англ. М.: АСТ; 2019, 224 с. ISBN 978-5-17-114349-7
3. Курцвейл Р. *Революция разума*: *на подступах к Сингулярности. Как технологии изменят общество и сознание*. М.: Эксмо, БОМБОРА; 2025, 661 с. ISBN 978-5-04-228765-7
4. Печчеи А. Вызов 70-х годов современному миру. Тема для дискуссии. В сб.: *Римский клуб. История создания, избранные доклады и выступления, официальные материалы;* под ред. Д.М. Гвишиани. М.: УРСС; 1997, 384 с. ISBN 5-88417-100-5
5. Кондратьев Н.Д. *Большие циклы конъюнктуры и теория предвидения:* Междунар. фонд Н.Д. Кондратьева и др. М.: Экономика; 2002, 550 с. ISBN 5-282-02181-1
6. Ковальчук М.В. *Идеология природоподобных технологий*. М.: Физматлит; 2021, 329 с. ISBN 978-5-9221-1931-3
7. Hara Sh. *Minimal FAB. Fab System Research Consortium*. AIST, Japan. 2010.
8. Маленков Ю.А., Давыдова А.Э. Возможности применения цифровых технологий в стратегическом планировании и прогнозировании устойчивого развития организаций. *Молодой ученый*. 2021;14(356):84–89. https://elibrary.ru/zrjnia

9. Глазьев С.Ю. Новые подходы к организации предприятий в условиях смены технологического и мирохозяйственного укладов. В сб.: *Стратегическое планирование и развитие предприятий:* пленарные доклады Двадцатого Всероссийского симпозиума; под ред. чл.-корр. РАН Г.Б. Клейнера. М.: ЦЭМИ РАН; 2019. С. 7–18.
10. Шпак В.В. *Развитие электронной промышленности в условиях меняющегося мира*. М.: Техносфера; 2024, 128 с. ISBN 978-5-94836-708-8
11. Длясин Г.Г. *Азбука Гермеса Трисмегиста или молекулярная тайнопись мышления*. М.: Белые альвы; 2002, 141 с. ISBN 5-7619-0126-9
12. Шпак В.В. Стратегическое планирование в современных условиях. В сб.: *Государственное управление в новых геополитических и геоэкономических условиях:* Материалы XIX международной конференции. МГУ, 5–9 декабря 2022 г. М.: КДУ, Добросвет; 2023. С. 130–136. https://elibrary.ru/lehfav
13. *Мезоэкономика: состояние и перспективы*: монография; под ред. Маевского В.И. М.: Институт экономики Российской академии наук; 2018, 314 с. ISBN 978-5-9940-0642-9
14. Скобелев П.О. Мультиагентные технологии в промышленных применениях: к 20-летию основания Самарской научной школы мультиагентных систем. *Мехатроника, автоматизация, управление*. 2010;12:33–46. https://elibrary.ru/nblbbh
15. Шпак В.В. Основы мультиагентной планово-логистической модели управления воспроизводственным циклом отрасли в режиме «онлайн». *Фундаментальные исследования*. 2022;3:146–155. https://doi.org/10.17513/fr.43229
16. Шпак В.В. Формирование организационно-управленческой модели реализации «Стратегии развития электронной промышленности Российской Федерации на период до 2030 года». *Вестник Челябинского государственного университета*. 2021;3(449):10–23. https://doi.org/10.47475/1994-2796-2021-10302
17. Шокин А.А. *Министр невероятной промышленности СССР*. М.: Техносфера; 2007, 455 с. ISBN 978-5-94836-151-2
18. Шпак В.В. Инвентаризация электроники – фундамент прорыва в лидеры. *Russian Technological Journal*. 2025;13(4):135–148. https://doi.org/10.32362/2500-316X-2025-13-4-135-148
19. Симчера В.М. *Изобличая лживые цифры. Этюды о пирровых победах и потерях России*. Издательские решения; 2019, 357 с. ISBN 978-5-4496-4788-7
20. Шведин Б.Я. *Онтология предприятия*: *экспириентологический подход. Технология построения онтологической модели предприятия на основе анализа и структурирования живого опыта*. М.: ЛЕНАНД; 2010, 240 с. ISBN 978-5-9710-0308-3

About the Author

Vasily V. Shpak, Cand. Sci. (Econ.), Associate Professor, Department of Nanoelectronics, Institute for Advanced Technologies and Industrial Programming, MIREA – Russian Technological University (78, Vernadskogo pr., Moscow, 119454 Russia). E-mail: morser@yandex.ru. https://orcid.org/0009-0002-3548-1070

Об авторе

Шпак Василий Викторович, к.э.н., доцент, кафедра наноэлектроники, Институт перспективных технологий и индустриального программирования, ФГБОУ ВО «МИРЭА – Российский технологический университет» (119454, Россия, Москва, пр-т Вернадского, д. 78). E-mail: morser@yandex.ru. https://orcid.org/0009-0002-3548-1070

Translated from Russian into English by L. Bychkova
Edited for English language and spelling by Thomas A. Beavitt

Philosophical foundations of technology and society

Мировоззренческие основы технологии и общества

UDC 304.5, 327, 339.9
https://doi.org/10.32362/2500-316X-2026-14-3-145-153
EDN UHPJPA

RESEARCH ARTICLE

Conceptualizing the global environmental challenges of our time: Technological and sociocultural aspects

Nikita Yu. Ryabtchenko [1], Artem P. Fedorov [2, @]

[1] *D.F. Ustinov Baltic State Technical University "VOENMEH", Saint Petersburg, 190005 Russia*
[2] *MIREA – Russian Technological University, Moscow, 119454 Russia*
[@] *Corresponding author, e-mail: z333e@yandex.ru*

• **Submitted:** 01.11.2025 • **Revised:** 23.12.2025 • **Accepted:** 22.03.2026

Abstract

Objectives. Nowadays, the global environmental agenda has become a significant factor in world politics and international relations. Along with climate change, growing demographic and technological pressures on the environment, the degradation of ecosystems, and the depletion of nonrenewable natural resources represent a single set of environmental challenges for humanity as a whole, which will remain relevant throughout the 21st century. The objective of this article is to consider the technological and sociocultural aspects of modern global environmental challenges in terms of their interconnectedness.

Methods. The article is based on the systems history method and the cycles and waves approach to studying global dynamics.

Results. Considering the main features of humanity's interaction with the environment, the factors and nature of changes in the anthropogenic burden are analyzed to reveal the stages according to which environmentalism is shaped as part of the global political agenda. Based on a comparison of the climatic picture of the past and modernity, priorities for steering the environmental agenda as a key component of the emerging world system are identified.

Conclusions. The overall anthropogenic impact on the environment is determined by a number of sociocultural, economic and technological factors. Although the degree of this impact has been increasing throughout history, climate change as such does not have a pronounced anthropogenic nature, being driven mainly by solar and geophysical cycles. Meanwhile, the very fact that climate change has become a key element of the global environmental agenda often leads to imposing regulatory requirements and technological standards that slow down the pace of industrialization and modernization in developing countries. In the context of the emerging world system architecture, it becomes possible to rethink the environmental agenda and repurpose it to take account of the priorities of climate forecasting and adaptation to ensure environmental safety while developing environmentally appropriate nature-similar technologies.

Keywords: global environmental challenges, environmental agenda, anthropogenic pressures, technosphere, spectrums of needs, nature-similar technologies, national security, the World Majority

For citation: Ryabtchenko N.Yu., Fedorov A.P. Conceptualizing the global environmental challenges of our time: Technological and sociocultural aspects. *Russian Technological Journal.* 2026;14(3):145–153. https://doi.org/10.32362/2500-316X-2026-14-3-145-153, https://www.elibrary.ru/UHPJPA

Financial disclosure: The authors have no financial or proprietary interest in any material or method mentioned.

The authors declare no conflicts of interest.

НАУЧНАЯ СТАТЬЯ

Осмысление глобальных экологических вызовов современности: технологические и социокультурные аспекты

Н.Ю. Рябченко [1], А.П. Федоров [2, @]

[1] *Балтийский государственный технический университет «ВОЕНМЕХ» им. Д.Ф. Устинова, Санкт-Петербург, 190005 Россия*
[2] *МИРЭА – Российский технологический университет, Москва, 119454 Россия*
[@] *Автор для переписки, e-mail: z333e@yandex.ru*

• **Поступила:** 01.11.2025 • **Доработана:** 23.12.2025 • **Принята к опубликованию:** 22.03.2026

Резюме

Цели. В настоящее время глобальная экологическая повестка стала значимым фактором мировой политики и международных отношений. Происходящие в мире климатические изменения, нарастающее демографическое и техногенное давление на среду обитания, деградация биогеоценозов и истощение невозобновляемых природных ресурсов представляют собой единый комплекс экологических вызовов для всего человечества, который останется актуальным в течение всего XXI века. Целью настоящей статьи является рассмотрение технологических и социокультурных аспектов глобальных экологических вызовов современности в их взаимосвязи.

Методы. В статье используются системно-исторический метод и циклически-волновой подход к изучению мировой динамики.

Результаты. Рассмотрены особенности взаимодействия человечества с природной средой, проанализированы факторы и характер изменения антропогенных нагрузок в исторической ретроспективе, показаны этапы становления глобально-политической повестки экологизма, проведено сравнение климатической картины прошлого и наших дней, обозначены приоритеты управления экологической повесткой в новом мироустройстве.

Выводы. Общее антропогенное влияние на природную среду определяется рядом социокультурных, экономических и технологических факторов. Хотя степень этого влияния на протяжении истории повышается, однако именно климатические изменения не имеют выраженной антропогенной природы, будучи подчинены главным образом солнечным и геофизическим циклам. Наряду с этим, тема климатических изменений стала одним из ключевых элементов глобальной экологической повестки, которая в ряде случаев приводит к постановке регуляторных требований и технологических стандартов, затрудняющих темпы индустриализации и модернизации в развивающихся странах. В условиях формирования новой системы международных отношений появляется возможность переосмысления экологической повестки с дальнейшим перенаправлением ее на приоритеты климатического прогнозирования и адаптации, обеспечения экологической безопасности и развития природоподобных технологий.

Ключевые слова: глобальные экологические вызовы, экологическая повестка, антропогенные нагрузки, техносфера, спектры потребностей, природоподобные технологии, национальная безопасность, Мировое большинство

Для цитирования: Рябченко Н.Ю., Федоров А.П. Осмысление глобальных экологических вызовов современности: технологические и социокультурные аспекты. *Russian Technological Journal*. 2026;14(3):145–153. https://doi.org/10.32362/2500-316X-2026-14-3-145-153, https://www.elibrary.ru/UHPJPA

Прозрачность финансовой деятельности: Авторы не имеют финансовой заинтересованности в представленных материалах или методах.

Авторы заявляют об отсутствии конфликта интересов.

INTRODUCTION

Currently, the global environmental agenda has become a significant factor in world politics and international relations. Global climate change, growing demographic and technogenic pressures on the environment, degradation of biogeocenoses, and depletion of nonrenewable natural resources represent a single set of environmental challenges for humanity as a whole, which will remain relevant throughout the 21st century. Meanwhile, the frequent use of the environmental agenda as a tool of international regulation reveals contradictions between different groups of countries in terms of their levels of technological development and long-term economic growth priorities. In light of the latter circumstance, it becomes important to identify the causes of the global biosphere-ecological crisis observed today, for which it is necessary to refer to objective biosphere-wide, species-specific, and sociocultural patterns in light of some relevant aspects of the global historical process.

THE *SAPIENS* PHENOMENON: FEATURES OF INTERACTION WITH THE NATURAL ENVIRONMENT

Homo sapiens is the only biological species in the Earth's biosphere that engages in the conscious, purposeful transformation of the natural environment to obtain various intermediate resources, material products, and other goods that did not previously exist in its habitat [1, 2].

However, like other biological species, humans also have their own initial, predetermined ecological niche, whose depletion poses a risk to the further existence of the species. Nevertheless, as culture (including the technosphere), material practices, and energy capabilities have developed, the ecological niche of *Homo sapiens* has constantly expanded [3]. At the same time, there is always a certain set of parameters for humanity's natural habitat, the significant violation of whose boundary values results in even the most extensive habitats becoming unsuitable for the further reproduction, survival, and development of the species.

Along with that, the human population has another important difference from the rest of the Earth's biosphere: while all other biological species have a deterministic spectrum of needs that remains unchanged over long periods of time, the spectrum of human needs is inherently variable, hierarchically complex, and throughout history has tended to expand as culture develops in terms of self-knowledge, self-organization, and material practice [3, 4].

ANTHROPOGENIC IMPACTS IN HISTORICAL RETROSPECTIVE

In the most general terms, four interrelated factors can be identified that determine the level of anthropogenic pressure on the natural environment: demographics, spectrum of needs, energy base, and level of development of productive forces (including technological development). Throughout human history, the dynamics of these factors have been uneven and nonlinear; moreover, their abrupt changes have always led to the transition of civilization to a qualitatively new state, characterized by the concept of "phase of development" [5].

The first such phase transition (from the archaic to the traditional phase of development) known as the Neolithic Revolution is associated with the emergence of a productive economy. At that time, the development of crop rotation, early construction technologies, and metallurgy led to a multiple increase in the human population and the first surge in anthropogenic pressure on the Earth's ecosystem. However, since human energy consumption was still incomparably lower than the energy capacity of natural global processes at the initial stage of civilization's development, the increased ecological footprint of humanity occurring at that time remained negligible [5, 6].

With the transition from the traditional to the industrial phase of development during the 18th century, humanity began to use technogenic energy for the first time, which then increasingly replaced biogenic energy in all production cycles, exponentially increasing the energy intensity of the *Homo sapiens* species.

The growth of anthropogenic pressure during the first industrial revolution was also associated with a new combination of sociocultural factors that included the secular ideology of bourgeois liberalism and the capitalist mode of production. This combination led to a significant expansion of the mass spectrum of needs, an increase in production volumes, and a corresponding reduction in production cycles, which for the first time in history began to have a noticeable impact on the natural environment [7, 8].

By the beginning of the first scientific and technological revolution at the turn of the 19th and 20th centuries, the world's population had reached 1 billion, while the aforementioned sociocultural factors continued to exert their influence and even intensified. Under the new technological conditions, this manifested itself in the beginning of the era of conveyor-belt production of mass-market goods.

A separate important phenomenon of that time was the emergence of the mass communications industry, which, as it developed, became a significant factor in the artificial expansion of the spectrum of degradative needs. This last point requires clarification: unlike basic, demographically determined personal, family, and infrastructural needs, degradative needs are artificially cultivated needs, whose satisfaction objectively harms the consumer, society, and the environment [4, 9].

Throughout the 20th century and up to the present day, under the influence of sociocultural factors in the world as a whole, especially in developed countries, there has been a predominance of such artificially created needs. This, combined with a manifold increase in population to 8 billion people and accelerating technological change, has resulted in a historically unprecedented increase in anthropogenic pressures on the environment [10].

ECOLOGISM: THE EMERGENCE OF A GLOBAL POLITICAL AGENDA

During the second half of the 20th century, the West began to show genuine concern for preserving the ecological niche necessary for humanity, while at the same time viewing environmental challenges as an incentive to consolidate efforts and strengthen its own technological and institutional advantages [11]. Based on this motivation and relying on their collective intellectual, technological, and (most importantly) institutional leadership, Western countries began to systematically form a new global political agenda centered on environmental issues and, in particular, climate change.

The milestones of this agenda are the report to the Club of Rome entitled "The Limits to Growth" (Meadows spouses report, 1972), the report of the United Nations World Commission on Environment and Development published under the heading of "Our Common Future" (the Brundtland Report, 1987), the adoption of Agenda 21 and the United Nations Framework Convention on Climate Change at the Rio de Janeiro conference (1992), the signing of the Kyoto Protocol (1997), the adoption of the Millennium Declaration at the United Nations General Assembly meeting in New York (2000), the adoption of the 2030 Agenda and the United Nations Sustainable Development Goals, the signing of the Paris Agreement (2015), the latest COP[1] conferences in Glasgow, Sharm El Sheikh, and Dubai (2021–2023), and the United Nations Summit of the Future held in September 2024 [12].

All of the above policy documents and measures within this agenda are united by the fundamental idea that climate change is predominantly anthropogenic in nature, is primarily linked to the greenhouse effect (global warming), and will have catastrophic consequences for humanity if left unchecked.

The climate agenda, which has been institutionalized and actively implemented in international political practice, has acquired the quality of a modern quasi-ideology that can be defined as political environmentalism. In its radical version, this quasi-ideology is already taking on radical features, including placing *Homo sapiens* on a par with the animal world and even with the plant kingdom, thus diminishing the value of human life and all cultural achievements.

Public debates on climate policy sometimes reveal selective use of scientific data, reflecting the diversity of positions and interests in this area. Another sign of the purely political and instrumental nature of this quasi-ideology is the unevenness of its media coverage: in some years, there is massive media promotion of relevant ideas, events, and personalities, while in others, the climate agenda virtually "fades away" due to the fact that global politics is being pursued by other means during that period.

A telling example of the politicization of environmental discourse is the work of Swedish eco-activist Greta Thunberg, whose public appearances have been interpreted as a tool for creating emotional pressure to promote the climate agenda. It is noteworthy that her entire family has consciously switched to veganism, viewing it as a personal contribution to reducing their carbon footprint. In this way, environmental issues are taking on the characteristics

[1] Discussion on the environmental agenda and the outcomes of the COP28 climate conference in Dubai. Valdai Discussion Club. https://ru.valdaiclub.com/multimedia/video/diskussiya-po-ekologicheskoy-povestke-i-itogam-cop28/. Accessed September 06, 2025. (In Russ.).

of mass ideological mobilization, in which individual practices and values are transformed into elements of global influence.

At the same time, it is important to understand that the environmental agenda is increasingly intertwined with issues of economic, energy, and technological policy [11]. From a non-western perspective, especially the countries of the Global South, it may often seem that the strategic goals of this policy are to increase global socioeconomic inequality and exacerbate disparities in terms of access to technologies and resources necessary for the development modernization.[2] At the same time, international environmental policy practice shows a tendency for industrialized countries to maintain their technological leadership and priority access to financial and innovative resources in the process of energy transition [13].

MULTIFACTORIAL NATURE OF CLIMATE CHANGE: FROM THE PAST TO THE PRESENT

Given that the current global environmental agenda focuses particularly on anthropogenic factors of climate change, it is important to note that there are other scientific approaches that view climate dynamics as the result of complex interactions between natural and anthropogenic processes. In this context, it is appropriate to briefly consider the nature and factors of climate change, based on both paleoclimatological data and the results of contemporary research in the field of climate science [14].

In general, the planet's climate dynamics represent a system of interconnected cyclical processes of varying duration. The main factors determining the formation of the Earth's climate are: (1) the power of absorbed solar radiation; (2) the power of the Earth's own radiation from its interior; and (3) the spatial relationship between the world's oceans and the geostructures of the land.

Thus, geological timespans measured in millions of years can be characterized by the alternation of cryoeras and thermoeras, as determined by the possibility of continuous heat and mass transfer in the equatorial plane of the planet. Accordingly, since in our geological era the equator is "covered" by two continents, humanity as a whole has invariably lived and continues to live in a climatic cryoera. During thermoeras, however, the Earth's temperature was much higher.

Furthermore, over tens of thousands of years, glacial and interglacial periods regularly alternate, including the alternation of over centuries of minor ice ages and so-called climatic optimums. Thus, the fall of the Roman Empire and the crisis of the ancient world partially coincided with a phase of climatic cooling known as the Late Antique Little Ice Age. In turn, the relative warming in the 11th–12th centuries, known as the Medieval Climate Optimum, was accompanied by favorable climatic conditions in the North Atlantic, which, in particular, allowed the Norse civilization to colonize the southern coastal areas of Greenland, which is now covered by a glacier.

Five hundred years later, at the turn of the 17th century, another minor ice age brought ice to the Moskva River in summer and the well-known years of poor harvests that caused famine in medieval Tsarist Russia. It is important to note here that the overall temperature difference between the climatic situations of the 12th and 17th centuries was from 1.5 to 2°C, i.e., a range that is actively discussed in international climate agreements and serves as a benchmark for assessing the possible consequences of global warming.[3] At the same time, it is obvious that the described differences between the climatic situation of the 12th and 17th centuries could not have been related to human activity, given the vanishingly small energy intensity and anthropogenic footprint of global civilization in that pre-industrial era.

Today, we are witnessing another phase of climate fluctuations caused by natural geophysical and astronomical cycles, along with anthropogenic factors [14].

Finally, according to a number of Russian and foreign studies, the natural factors influencing modern climate warming include a change in the angle of the Earth's axis of rotation, which affects the insolation of polar and equatorial regions and, as a result, the intensity of meridional heat and mass transfer [15–17]. At the same time, most climate models point to the significant role of anthropogenic factors—primarily the increase in greenhouse gas concentrations—in shaping current climate trends, which remains the subject of active scientific debate.

Moreover, various hypotheses explaining the processes of atmospheric ozone destruction are discussed in scientific literature. Along with anthropogenic sources, in particular chlorofluorocarbon emissions, which played a key role in the formation of ozone "holes" at the end of the 20th century, natural geophysical factors are also

[2] Barabanov O., Bespalov A., Ibragimova K., Koltashov V., Poletaev D., Savorskaya E. *The social dimension of the "global commons": can inequality in the world be overcome?* Valdai Discussion Club. https://ru.valdaiclub.com/a/reports/sotsialnoe-izmerenie-globalnogo-dostoyaniya/. Accessed September 05, 2025. (In Russ.).

[3] Special report: global warming of 1.5°C. IPCC. https://www.ipcc.ch/sr15/. Accessed September 10, 2025.

being studied, including degassing processes from the Earth's interior, volcanic activity, and solar variability. Studying the relationships between these processes opens up opportunities for more accurate forecasting of meteorological changes and assessment of climate risks [18].

Thus, past and present climate change is mainly caused by natural factors. Anthropogenic impact only serves to amplify the overall climate process [19]. However, unlike previous generations, humanity now has sufficient scientific and technological capabilities not only to predict adverse climate change, but also to adapt its environment to these changes in advance using the state-of-the-art technologies and governance principles [10, 12, 14].

STEERING THE ENVIRONMENTAL AGENDA IN THE NEW WORLD SYSTEM ARCHITECTURE: THE ROLES OF SCIENCE AND MULTILATERAL DIPLOMACY

Apart from objective macro-processes beyond human control, genuine environmental challenges arising at the present time are caused by the cumulative anthropogenic impact on the natural environment, which is primarily determined by technological and sociocultural factors. Therefore, given that the main sociocultural factor is the artificially formed spectrum of needs already described, the technological factor should be considered in more detail.

Throughout human history, the technosphere has developed as a means of protection from the natural environment and, in many ways, as a counterbalance to it. The logical consequence of this approach has been the primitive principles of operation of the technologies available to us, which are inadequate in terms of the complexity and sustainability of the natural processes observed.

The majority of global energy sector is still based on burning fossil fuels. Even modern nuclear power plants are based on traditional principles of thermal power engineering, where the key process remains the conversion of thermal energy into mechanical and electrical energy through the formation of steam. However, nature itself demonstrates the prevalence of biochemical and thermonuclear fusion reactions characterized by energy release.

In this connection, a promising direction involves the development of nature-similar technologies[4] developed by the Kurchatov Institute[5] in Russia, which now has a branch in Belarus.

These are technologies that model the principles of ecological systems to reproduce natural processes in the form of technical solutions.[6] Nature-similar technologies are based on the use of renewable resources, i.e., they do not disrupt the cycles of substances in nature, and are also highly energy efficient.[7]

The interconnection between technological and environmental aspects of global dynamics is particularly important from the perspective of ensuring the comprehensive security of sovereign states. Thus, according to the current version of the National Security Strategy of the Russian Federation, a number of strategic national priorities are linked to scientific and technological development, environmental security, and rational use of natural resources. In addition, taking into account long-term trends in Russia and worldwide, the national interests of the Russian Federation include environmental protection, conservation of natural resources, adaptation to climate change, and ensuring the sustainable development of the Russian economy on a new technological basis.[8]

At the same time, the Climate Doctrine of the Russian Federation rightly notes that problems related to climate change, in particular, ensuring a balance between economic efficiency and social justice, eliminating potential conflicts of interest in connection with extreme manifestations of climate change (heat waves, floods, droughts, and other phenomena) cannot be solved by scientific methods alone. Under such conditions, the necessary balance can only be achieved through comprehensive management decisions.[9] When developing these decisions, it is advisable to take

[4] Decree of the President of the Russian Federation No. 818 dated November 2, 2023, "On the Development of Nature-Inspired Technologies in the Russian Federation." Official publication of legal acts. http://publication.pravo.gov.ru/document/0001202311020021. Accessed September 25, 2025. (In Russ.).

[5] Thematic session "Nature-like technologies: a new era of human development" as part of the Third Congress of Young Scientists. Roscongress Foundation. https://roscongress.org/sessions/kmu-2023-prirodopodobnye-tekhnologii-novaya-era-razvitiya-chelovechestva/about/#. Accessed September 22, 2025. (In Russ.).

[6] Decree of the President of the Russian Federation No. 145 dated February 28, 2024, "On the Strategy for Scientific and Technological Development of the Russian Federation." Government of Russia. http://government.ru/docs/all/152305/. Accessed September 30, 2025. (In Russ.).

[7] Nature-like technologies. Kurchatov Institute Research Center. https://nrcki.ru/catalog/nauka/fundamentalnye-i-prikladnye-nauchnye-issledovaniya/nbiks-prirodopodobnye-tekhnologii/. Accessed October 02, 2025. (In Russ.).

[8] Decree of the President of the Russian Federation No. 400 dated July 2, 2021, "On the National Security Strategy of the Russian Federation." President of Russia. http://www.kremlin.ru/acts/bank/47046/page/1. Accessed October 02, 2025. (In Russ.).

[9] Decree of the President of the Russian Federation No. 812, dated October 26, 2023, "On the Approval of the Climate Doctrine of the Russian Federation." Official publication of legal acts. http://publication.pravo.gov.ru/document/0001202310260009. Accessed October 04, 2025. (In Russ.).

into account the fact that the vector of development of the global economy—and, accordingly, scientific research—is currently shifting from robotization and automation (Industry 4.0) to the development of sociocentric and environmentally sustainable technologies [20].

CONCLUSIONS

Thus, when speaking about global environmental challenges in general, the understanding of this topic outlined above opens up opportunities for interested and responsible actors in global politics to form and promote common positions in this area within the framework of new formats for interaction between the countries of the World Majority and specialized organizations of the United Nations system that are oriented toward a more balanced approach to global development that takes into account the interests of all groups of countries. Such common positions may primarily concern the following issues:

- adaptation to expected climate change, climate forecasting tools, and forward-looking energy policy;
- principles of prudent management of scientific and technological progress, priority of developing nature-like technologies;
- the need to change the paradigm of economic and cultural development in order to minimize the range of degrading needs;
- principles of responsible production, fair distribution, and moderate consumption in the global economy.

In such a context, the updated "environmental agenda of the World Majority" could become a real unifying direction for cooperation within international organizations, transcending any ethnic and religious differences.

Authors' contributions

N.Yu. Ryabtchenko—conceptualization, methodology, writing (original draft preparation).

A.P. Fedorov—literature review, manuscript editing, revising the research content.

All authors have read and agreed to the published version of the manuscript.

REFERENCES

1. Stepin V.S. A radical change in civilization. *Ekologiya i zhizn' = Ecology and Life.* 2012;5:10–14 (in Russ.).
2. Milovzorova M.N. *Aksiologicheskie osnovaniya bezopasnogo razvitiya sotsial'nykh sistem v usloviyakh globalizatsii* (*Axiological Foundations for Secure Development of Social Systems in the Context of Globalization:* monograph). Saint Petersburg: Asterion; 2021, 266 p. (In Russ.). ISBN 978-5-00188-109-4
3. Rakitskii B.V. *Nauka o bytii chelovechestva* (*obshchestvo i chelovek v obshchestve*) (*Science on Humanity in Being* (*Society and a Person within Society*). Moscow: Probel-2000; 2020, 152 p. (In Russ.). ISBN 978-5-98604-527-6
4. Velichko M.V., Efimov V.A., Zaznobin V.M. *Ehkonomika innovatsionnogo razvitiya: upravlencheskie osnovy ehkonomicheskoi teorii* (*Economics for Innovative Development: Managerial Foundations of Economic Theory*). Moscow: Kontseptual; 2017, 571 p. (In Russ.). ISBN 978-5-906867-59-9
5. Pereslegin S.B., Pereslegina E. "Dikie karty" budushchego. Fors-mazhor dlya chelovechestva (*"Wild Cards" of the Future. Force Majeure for Humanity*). Moscow: Algoritm; 2015, 480 p. (In Russ.). ISBN 978-5-4438-0983-0
6. Pereslegin S.B. Technological order VI: The space of opportunities. *Ekonomicheskie strategii.* 2019;21(3-161):24–34 (in Russ.). https://elibrary.ru/malwfm
7. Maslov G.A. Theory of technologic orders: reconsidering the change of globalization waves. *Problemy sovremennoi ekonomiki = Problems of Modern Economics.* 2021;1(77):65–70 (in Russ.). https://elibrary.ru/jqclai
8. Glazyev S.Yu. The modern theory of long waves in economic development. *Ekonomicheskaya nauka sovremennoy Rossii = Economics of Contemporary Russia*. 2012;2(57):27–42 (in Russ.). https://www.elibrary.ru/ozejdt
9. Kovalev D.V. A transformation of consciousness influenced by technological factors of globalization. *Tendentsii Razvitiya Nauki i Obrazovaniya*, 2019;(56-14):37–41 (in Russ.). https://doi.org/10.18411/lj-11-2019-305, https://www.elibrary.ru/kdcdfj
10. Sadovnichiy V.A., Akaev A.A., Il'in I.V., Korotaev A.V., Malkov S.Yu. Modeling and forecasting of global dynamics in the 21st century. *Vestnik Moskovskogo universiteta. Seriya 27. Globalistika i geopolitika = Bulletin of Moscow University. Series 27. Global Studies and Geopolitics.* 2022;1:5–35 (in Russ.). https://doi.org/10.56429/2414-4894-2022-39-1-5-35
11. Yakovenko A.V. (Ed.). *Kartina narozhdayushchegosya mira: bazovye cherty i tendentsii* (*A Picture of the Emerging World: Basic Features and Trends*): Report. Moscow; 2024, 62 p. (In Russ.).
12. Sadovnichiy V.A. (Ed.). *Preodolevaya predely rosta. Osnovnye polozheniya doklada dlya Rimskogo kluba* (*Overcoming the Limits to Growth. Main Provisions of the Report to the Club of Rome*): monograph. Moscow: Moscow University Press; 2023, 99 p. (In Russ.). ISBN 978-5-19-011875-9
13. Dynkin A.A. World order transformation: economy, ideology, technology. *Polis. Politicheskie issledovaniya = Polis. Political Studies*. 2024;5:8–23 (in Russ.). https://doi.org/10.17976/jpps/2024.05.02
14. Glaz'ev S.Yu., Bezrukov L.B., Dolgolaptev A.V., Larin N.V., Syvorotkin V.L., Fedorov V.M. Climate change and energy transition. *Ehkonomicheskie strategii.* 2023;6:16–29 (in Russ.). https://www.elibrary.ru/ugbxrb
15. Smirnov B.M. *Fizika global'noi atmosfery. Parnikovyi ehffekt, atmosfernoe ehlektrichestvo, ehvolyutsiya klimata* (*Physics of Global Atmosphere. Greenhouse Effect, Atmospheric Electricity, Evolution of Climate*). Moscow: Intellekt; 2017, 256 p. (In Russ.).

16. Riser S.C., Freeland H.J., Roemmich D., et al. Fifteen years of ocean observations with the global Argo array. *Nature Climate Change.* 2016;6:145–153. https://doi.org/10.1038/nclimate2872
17. Keeling C.D. The concentration and isotopic abundances of carbon dioxide in the atmosphere. *Tellus.* 1960;12(2):200–203. https://doi.org/10.3402/tellusa.v12i2.9366
18. Syvorotkin V.L. *Glubinnaya degazatsiya i global'nye katastrofy* (*Deep Degassing and Global Disasters*). Moscow: Geoinformtsentr; 2002, 250 p. (In Russ.). ISBN 5-900357-80-5
19. Monin A.S., Sonechkin D.M. *Kolebaniya klimata po dannym nablyudenii. Troinoi solnechnyi i drugie tsikly* (*Climate Fluctuations Based on Observation Data. Triple Solar and Other Cycles*). Moscow: Nauka; 2005, 190 p. (In Russ.).
20. Karpukhina N.N., Mityakov E.S., Pronin A.Yu. Industrial revolutions: From Industry 3.0 to Industry 5.0 in the context of the Russian economy. *Russian Technological Journal.* 2025;13(4):123–134. https://doi.org/10.32362/2500-316X-2025-13-4-123-134

СПИСОК ЛИТЕРАТУРЫ

1. Стёпин В.С. Коренной перелом цивилизации. *Экология и жизнь*. 2012;5:10–14. https://www.elibrary.ru/pdfdqh
2. Миловзорова М.Н. *Аксиологические основания безопасного развития социальных систем в условиях глобализации:* монография. СПб.: Астерион; 2021, 266 с. ISBN 978-5-00188-109-4
3. Ракитский Б.В. *Наука о бытии человечества* (*общество и человек в обществе*). М.: Пробел-2000; 2020, 152 с. ISBN 978-5-98604-527-6
4. Величко М.В., Ефимов В.А., Зазнобин В.М. *Экономика инновационного развития: управленческие основы экономической теории*. М.: Концептуал; 2017, 571 с. ISBN 978-5-906867-59-9
5. Переслегин С.Б., Переслегина Е. «*Дикие карты» будущего. Форс-мажор для человечества.* М.: Алгоритм; 2015, 480 с. ISBN 978-5-4438-0983-0
6. Переслегин С.Б. VI технологический уклад: пространство возможностей. *Экономические стратегии.* 2019;21(3-161): 24–34. https://elibrary.ru/malwfm
7. Маслов Г.А. Теория технологических укладов: осмысливая смены волн глобализации. *Проблемы современной экономики*. 2021;1(77):65–70. https://elibrary.ru/jqclai
8. Глазьев С.Ю. Современная теория длинных волн в развитии экономики. *Экономическая наука современной России* (*ЭНСР*). 2012;.2(57):27–42. https://www.elibrary.ru/ozejdt
9. Ковалёв Д.В. Трансформация сознания под влиянием технологических факторов глобализации. *Тенденции развития науки и образования*. 2019;(56-14):37–41. https://doi.org/10.18411/lj-11-2019-305, https://www.elibrary.ru/kdcdfj
10. Садовничий В.А., Акаев А.А., Ильин И.В., Коротаев А.В., Малков С.Ю. Моделирование и прогнозирование глобальной динамики в XXI веке. *Вестник Московского университета. Серия 27. Глобалистика и геополитика.* 2022;1:5–35. https://doi.org/10.56429/2414-4894-2022-39-1-5-35
11. *Картина нарождающегося мира: базовые черты и тенденции: доклад:* под ред. А.В. Яковенко. М.: Дипломатическая академия МИД России; 2024, 62 с.
12. *Преодолевая пределы роста. Основные положения доклада для Римского клуба*: монография: под ред. В.А. Садовничего. М.: Изд-во Московского университета; 2023, 99 с. ISBN 978-5-19-011875-9
13. Дынкин А.А. Трансформация мирового порядка: экономика, идеология, технологии. *Полис. Политические исследования*. 2024;5:8–23. https://doi.org/10.17976/jpps/2024.05.02
14. Глазьев С.Ю., Безруков Л.Б., Долголаптев А.В., Ларин Н.В., Сывороткин В.Л., Федоров В.М. Климатические изменения и энергопереход. *Экономические стратегии*. 2023;6:16–29. https://www.elibrary.ru/ugbxrb
15. Смирнов Б.М. *Физика глобальной атмосферы. Парниковый эффект, атмосферное электричество, эволюция климата*. М.: Изд-во «Интеллект»; 2017, 256 с.
16. Riser S.C., Freeland H.J., Roemmich D., et al. Fifteen years of ocean observations with the global Argo array. *Nature Climate Change.* 2016;6:145–153. https://doi.org/10.1038/nclimate2872
17. Keeling C.D. The concentration and isotopic abundances of carbon dioxide in the atmosphere. *Tellus.* 1960;12(2):200–203. https://doi.org/10.3402/tellusa.v12i2.9366
18. Сывороткин В.Л. *Глубинная дегазация и глобальные катастрофы*. М.: Геоинформцентр; 2002, 250 с. ISBN 5-900357-80-5
19. Монин А.С., Сонечкин Д.М. *Колебания климата по данным наблюдений. Тройной солнечный и другие циклы.* М.: Наука; 2005, 190 с.
20. Карпухина Н.Н., Митяков Е.С., Пронин А.Ю. Промышленные революции: от Индустрии 3.0 к Индустрии 5.0 в контексте российской экономики. *Russian Technological Journal.* 2025;13(4):123–134. https://doi.org/10.32362/2500-316X-2025-13-4-123-134

About the Authors

Nikita Yu. Ryabtchenko, Postgraduate Student, D.F. Ustinov Baltic State Technical University "Voenmeh" (1, 1-ya Krasnoarmeiskaya ul., Saint Petersburg, 190005 Russia). E-mail: latempetedeneige@yandex.ru. https://orcid.org/0009-0001-9269-7612

Artem P. Fedorov, Cand. Sci. (Polit.), Associate Professor, Department of Information and Analytical Systems of Cybersecurity, Institute for Cybersecurity and Digital Technologies, MIREA – Russian Technological University (78, Vernadskogo pr., Moscow, 119454 Russia). E-mail: z333e@yandex.ru. RSCI SPIN-code 7872-8580, https://orcid.org/0009-0004-1772-7254

Об авторах

Рябченко Никита Юрьевич, аспирант, ФГБОУ ВО Балтийский государственный технический университет «Военмех» им. Д.Ф. Устинова (БГТУ «Военмех» им. Д.Ф. Устинова) (190005, Россия, Санкт-Петербург, ул. 1-я Красноармейская, д. 1). E-mail: latempetedeneige@yandex.ru. https://orcid.org/0009-0001-9269-7612

Федоров Артем Павлович, к.полит.н., доцент, кафедра КБ-2 «Информационно-аналитические системы кибербезопасности», Институт кибербезопасности и цифровых технологий, ФГБОУ ВО «МИРЭА – Российский технологический университет» (119454, Россия, Москва, пр-т Вернадского, д. 78). E-mail: z333e@yandex.ru. SPIN-код РИНЦ 7872-8580, https://orcid.org/0009-0004-1772-7254

Translated from Russian into English by L. Bychkova
Edited for English language and spelling by Thomas A. Beavitt

Philosophical foundations of technology and society

Мировоззренческие основы технологии и общества

UDC 372.881.1
https://doi.org/10.32362/2500-316X-2026-14-3-154-165
EDN VWGQAB

RESEARCH ARTICLE

Didactic modeling for teaching technological university students the rules of foreign language texts reading

Nadezhda I. Chernova [@], Ekaterina A. Ivanova, Nataliya V. Katakhova

MIREA – Russian Technological University, Moscow, 119454 Russia
[@] Corresponding author, e-mail: chernova@mirea.ru

• **Submitted:** 01.12.2025 • **Revised:** 12.12.2025 • **Accepted:** 27.03.2026

Abstract

Objectives. The work set out to develop didactic models for teaching French reading rules to first-year students from scratch, followed by a choice of the appropriate model to use in teaching.

Methods. The application of didactic models developed using tools such as regression analysis and mathematical theory of learning along with an experiment on the application of the obtained models in groups of French language learners “from scratch.”

Results. Three obtained models of studying 48 French reading rules over 4, 8, and 16 weeks with 12, 6, and 3 rules per lesson respectively, along with a review of the rules learned in the previous lesson, are presented for comparison and analysis. The influence of various factors such as students’ linguistic abilities, their levels of anxiety, previous language learning experience, etc., on the effectiveness of all three models were also taken into account. Based on the results of these analyses, the optimal learning model was chosen.

Conclusions. The average student learned reading rules more efficiently during an eight-week course when the rules were regularly reviewed at each lesson until the end of the semester. When studying reading rules for 16 weeks, students failed to review some material at the end of the semester, with fewer rules being reviewed in lesson. During a four-week study, students confused the rules and had difficulties to remember them due to cognitive overload. In the presence of adverse factors (weak ability and motivation, lack of independent work, etc.) training results were low regardless of the model chosen. On the contrary, under favorable conditions (good abilities, motivation, etc.) the learning model was turned out to be unimportant.

Keywords: didactic model, reading rules, French language, technological university

For citation: Chernova N.I., Ivanova E.A., Katakhova N.V. Didactic modeling for teaching technological university students the rules of foreign language texts reading. *Russian Technological Journal.* 2026;14(3):154–165. https://doi.org/10.32362/2500-316X-2026-14-3-154-165, https://www.elibrary.ru/VWGQAB

Financial disclosure: The authors have no financial or proprietary interest in any material or method mentioned.

The authors declare no conflicts of interest.

НАУЧНАЯ СТАТЬЯ

Дидактическое моделирование обучения студентов технологического университета правилам чтения иноязычных текстов

Н.И. Чернова [@], **Е.А. Иванова**, **Н.В. Катахова**

МИРЭА – Российский технологический университет, Москва, 119454 Россия
[@] *Автор для переписки, e-mail: chernova@mirea.ru*

• **Поступила:** 01.12.2025 • **Доработана:** 12.12.2025 • **Принята к опубликованию:** 27.03.2026

Резюме

Цели. Целями работы являются разработка дидактических моделей процесса преподавания правил чтения французского языка, изучаемого студентами первого курса с нуля, и выбор оптимальной модели для дальнейшего использования в обучении.

Методы. Применение дидактических моделей, разработанных с привлечением таких инструментов, как регрессионный анализ, математическая теория обучения и эксперимент по применению полученных моделей в группах, обучающихся французскому языку с нуля.

Результаты. Для сравнения и анализа представлены 3 полученные модели изучения 48 правил чтения во французском языке в течение 4, 8 и 16 недель по 12, 6 и 3 правила за одно занятие, соответственно, с повторением правил, изученных на предыдущем занятии. Учитывалось также влияние различных индивидуальных факторов на эффективность всех 3 моделей (лингвистические способности студента, уровень тревожности, предыдущий опыт изучения языков и т.п.). В результате была выбрана оптимальная модель обучения.

Выводы. Среднестатистический студент эффективнее усвоил правила чтения при восьминедельном обучении, когда последующие 8 недель (до конца семестра) правила регулярно повторялись на каждом занятии. При изучении правил чтения в течение 16 недель студенты не успели повторить некоторый материал в конце семестра, меньше правил повторялось на занятиях. При четырехнедельном изучении студенты путали правила, плохо их запоминали из-за когнитивной перегрузки. При наличии неблагоприятных факторов (слабые способности и мотивация, отсутствие самостоятельной работы и т.п.) результаты обучения оказались низкими независимо от выбранной модели. Напротив, при благоприятных условиях (хорошие способности, мотивация и т. п.) оказалось, что модель обучения не играет решающей роли.

Ключевые слова: дидактическая модель, правила чтения, французский язык, технологический университет

Для цитирования: Чернова Н.И., Иванова Е.А., Катахова Н.В. Дидактическое моделирование обучения студентов технологического университета правилам чтения иноязычных текстов. *Russian Technological Journal*. 2026;14(3):154–165. https://doi.org/10.32362/2500-316X-2026-14-3-154-165, https://www.elibrary.ru/WWGQAB

Прозрачность финансовой деятельности: Авторы не имеют финансовой заинтересованности в представленных материалах или методах.

Авторы заявляют об отсутствии конфликта интересов.

INTRODUCTION

The modern educational paradigm is undergoing a transformation that involves a comprehensive and coordinated overhaul of key aspects of learning, including learning objectives, curricula, teaching methodology, and the structure of educational activities. On the one hand, a digital transformation is taking place, which refers to the integration of digital technologies, the creation of digital educational platforms, and the introduction of artificial intelligence into the teaching process in order to optimize the management of educational processes [1–3]. On the other hand, the development of mathematical and didactic models, including those that replace experimental research, along with the introduction of mathematical methods and didactic modeling methods into the humanities, signify a transformation in professional training at universities, which is particularly relevant in the teaching of foreign languages at technical universities, where the educational emphasis is on the study of core disciplines. Here, the determination of the structure of the curriculum, work program, teaching materials, and mathematically verified prediction of learning outcomes for each study group and individual student are designed to optimize the processes of teaching and learning a foreign language. Modern didactic models provide an equivalent description of objects, their relationships and dynamics, as well as the parameters of functional dependencies. Thus, didactic modeling provides tools for objective assessment and modeling of foreign language teaching.

The purpose of the present study is to model the process of teaching French reading rules to first-year students starting from scratch and to determine the optimal model.

The main research method is the use of didactic models, which are developed using various methods including regression analysis and mathematical learning theory. The models obtained in this way were experimentally tested in different groups of RTU MIREA students who had commenced a course of French language study.

The novelty of this research consists in the current apparent lack of any formal-logical pedagogical studies involving comparable models.

The practical significance of the study lies in the prospect of managing the educational work of students at a technological university and selecting the optimal educational trajectory. This is achieved by analyzing the influence of various parameters on the learning process and identifying their interrelationships.

LITERATURE REVIEW

The study was conducted within the framework of educational analytics aimed at studying the educational process and the quality of education. In previously published works, educational analytics tends to be considered solely from the point of view of the application of information platforms [1], relevant information support for teachers and students, assessment of the latter's success [4], and the development of information systems that predict students' academic performance based on three parameters [5]. The digital transformation of education involves the use of educational analytics to create a system for managing the educational process [1, 3]. In the case under consideration, an example of such a system is based on mathematical calculations, which can then be converted into digital format and posted on university learning platforms. When building a model using information platforms and mathematical calculations, various types of data are taken into account, which can be constant or variable. On this basis, the educational process management system is built and learning analytics tools are developed, leading to the selection of optimal curricula and work programs.

Works [6–8] present developed mathematical and computer models for researching the educational process based on general systems theory and the axioms of mathematical learning theory. Modeling of weakly formalized systems covering aspects of the selection of educational resources, teaching methods, and criteria for evaluating their effectiveness is possible only within the framework of fuzzy cognitive modeling [8–10]. There are works devoted to a stochastic model in which the probability of data can vary.[1] Here methods are also presented for estimating the parameters of this model and examples of its application to various types of data related to the learning process.

Work [10] describes algorithms for creating teaching materials along with training models based thereupon, as well as the "Training Trajectory" program for any discipline and the "Group Trajectory" program. In addition, the work provides an insight into the process of training teachers in the use of mathematical modeling methods.

The review by Tishina [11] notes that the educational process is a complex system determined by numerous and constantly changing factors, which requires constant adjustment of the methods used with the help of a specific algorithm and mathematical models, which should be experimentally verified, since models "cannot be absolutely adequate to the object." For the effective selection and compilation of a mathematical model of learning, the author recommends using analytical methods developed within the framework of graph theory.

[1] Bush R.R., Mosteller F.A *Stochastic Model with Applications to Learning*. https://projecteuclid.org/download/pdf_1/euclid.aoms/1177728914. Accessed April 24, 2024.

Researchers Kiriy and Chan Van An [6] propose a mathematical model designed for studying the process of foreign language acquisition. This model, based on Kolmogorov's equations, establishes the dependence of a student's level of language proficiency on their initial training, learning speed, and forgetting speed.

In his study, Mayer [7] provides a fairly detailed analysis of a teacher-student interaction system, examining both linear and nonlinear approaches to learning modeling, as well as models describing the process of knowledge acquisition and subsequent information loss.

Study [12], which describes the use of mathematical methods as part of the educational process, examines in detail key factors such as student learning ability, professional competence of teachers, etc. The author identified the relationships between these factors, for example, between technical equipment and the organization of educational material. Concerning the initial parameters, the definition and analysis of student learning ability, as reflected in study [13], appears to be significant. Thanks to the created entrance test and its productive implementation by the authors, this parameter can be considered a "clear" variable, unlike most parameters in the humanities.

Reichle [9] has developed computational models that promote understanding of the cognitive processes involved in word recognition, sentence analysis, coherent text construction, and eye movement control.

Barack and Tsodyks [14] demonstrate various types of learning (classical conditioning, associative memory, free recall) using specific examples. The work examines modeling levels in terms of their connection to empirical data and mathematical concepts that are often found in learning models from an optimization perspective. It is important to note that this work belongs to the field of neuromodeling, rather than purely mathematical modeling.

Study [15] proposes a mathematical model that assesses the completeness of knowledge acquired in all course topics along with the dynamics of learning educational elements over a certain period. It also presents a project for the phased organization of students' educational activities.

The study presented in [16] analyzes various learning models in the context of long-term observation. The results showed that some cognitive abilities can be preserved even in the absence of regular practice. However, this study mainly focuses on the psychological modeling of learning.

Study [17] reflects how averaging individual learning and forgetting curves transforms the overall curves, leading to a conclusion that learning rates may follow different distributions. The authors analyze the skewness of these distributions to exclude cases of low learning rates.

In [18], a knowledge graph is used to study intelligent adaptive learning, which is interpreted as the leading tool for acquiring, managing, and optimizing the learning process. A proposed knowledge graph model of the "knowledge-resources-goals" type is based on a strong connection between knowledge, learning materials, and educational tasks. The author also proposes a bidirectional model that takes into account both long-term and short-term memory. The study confirms that such a model can be successfully adapted to learning objectives, reorganizes knowledge and resources, optimizes learning trajectories, offers recommendations for selecting materials, and visualizes the cognitive state of learners, thereby increasing the effectiveness of intelligent adaptive learning.

Study [19] proposes a system of personalized recommendations based on subject-specific knowledge graphs, learner profiles (obtained using xAPI[2]), and feedback on the learning process. The work [20] is dedicated to solving the problem of diversity in generative language theories by applying formal mathematical methods to language analysis.

One of the objectives of the study [21] was to identify the links between the factors that motivate students to continue studying a language. Multiple linear regression analysis was used to test the hypotheses and confirm the validity of the proposed model. This analysis mathematically models aspects such as "the relevance of teaching materials and technologies to learning objectives and available technologies," "the compatibility of the technologies used with students' individual preferences," "the openness and authority of educational institutions offering multimodal language learning," and "students' personal investment in multimodal language education."

Work[3] is devoted to the development of the following topics: "Algorithms for language learning using reinforcement," "Asymmetry of differences in language acquisition between adults and children," and "Mathematical modeling of language development within a population."

Paper [22] presents a mathematical model that examines the impact of educational programs. In particular, it considers a model for university students learning English that takes into account variables such as

[2] Experience Application Programming Interface is a standard for data exchange between educational content and distance learning systems.

[3] Rische J.L. *Mathematical Modeling of Language Learning*: *Dissertation.* University of California. Irvine. 2014, 89 p. https://escholarship.org/uc/item/0kb837r3. Accessed August 17, 2024.

the total number of students, the number of students who have chosen a modern learning format, the number of students who have chosen a traditional learning format, and the number of students who have successfully completed the course. The results of the analysis of this model indicate that attempts to solve problems in education can lead to unexpected negative consequences, such as reverse bifurcation and complication of the situation, which implies the possibility of failing to resolve existing educational problems.

Paper [23] considers a mathematical model based on multiple linear regression analysis. This model establishes relationships between the language learning strategies used and the level of English proficiency. The analysis of the model confirmed that it is not necessary to include all existing language learning strategies in the model in order to accurately predict language proficiency.

In the logic of the system-structural approach, the work [24] examines the quantitative characteristics of the educational process, which include learning, system state, the degree of learning, and the accuracy of learning. The proposed mathematical model of the learning process is implemented using software tools.

The software developed and presented in [25] for modeling the teaching of a foreign language analyzes, predicts, and monitors the level of proficiency in a non-native language, obtains approximations to the parameters of a mathematical model of the process of teaching a foreign language, constructs a graph of the progressive change in the probabilities of three states selected by the authors, and facilitates testing with various parameters in order to optimize their values to achieve the target level of proficiency in a foreign language.

In the paper [26], the authors identify previously unknown linguistic features to which certain formal attributes need to be assigned in order to create reinforcement learning algorithms for teaching spoken language.

Unfortunately, as in most similar studies, the mentioned works consider mathematical models of learning in a general form, without their specific application to the process of learning foreign languages.

Among the existing studies, algorithms and models are also presented that work in the field of studying the process of acquiring a foreign language [6, 21, 23, 26].[4] However, these works determine the level of proficiency in a foreign language, motivation, and language development rather than addressing specific models of teaching specific topics within the framework of studying foreign languages in general, and the French language in particular.

Thus, the analysis revealed that numerous studies are being conducted in the field of educational process modeling, which aim to strictly formalize vague didactic concepts, objects, and categories. It should be noted that the reviewed studies do not include a didactic model based on teaching French reading rules, designed to identify the optimal learning time and the amount of material covered in a single lesson.

MATERIALS AND METHODS

The study material, which represented an example of modeling educational processes, was based on the reading rules taught to first-year students of all training programs at RTU MIREA, who were studying French from scratch.

The research methods used included regression data analysis[5] (number of reading rules studied per lesson; number of reading rules repeated per lesson; learning ability of a specific cohort of students) and the construction of a didactic model for selecting the optimal number of rules studied per lesson along with the total number of the lessons devoted to learning and repeating the rules by a specific cohort of students. This didactic model thus describes the relationship between variables and the dependencies between them.

It must be recognized that pedagogical structures have a probabilistic nature and depend on many parameters that cannot be fully, strictly and objectively described [15]. Consequently, the model proposed in this paper is based only on the most essential characteristics of the educational activity system.

DEVELOPMENT OF A DIDACTIC MODEL

Developing a didactic model describing pronunciation rules in French represents a challenging task. The main difficulty lies in the significant discrepancy between the spelling and pronunciation of words. The goal of the model is to formalize the rules for the correspondence between graphic elements and phonetic units.

The rules of the following type were created:

$$L_n + (L_{n+1}\ (+L_{n+2})) = S_m\ (+S_k),$$

where L_n (L from French *lettre*—letter) is the first graphic element in the letter combination; L_{n+1} is the second graphic element; L_{n+2} is the third graphic element, if

[4] Rische J.L. *Mathematical Modeling of Language Learning*: *Dissertation.* University of California. Irvine. 2014, 89 p. https://escholarship.org/uc/item/0kb837r3. Accessed August 17, 2024.

[5] Dyachuk A.A. *Mathematical Methods in Psychological and Pedagogical Research*. Study Guide. Krasnoyarsk; 2013, 347 p. (In Russ.).

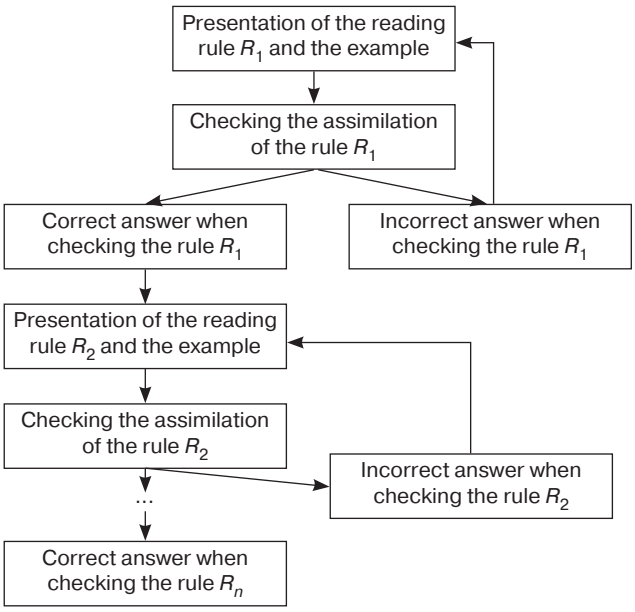


Fig. 1. Sequence of steps for studying the rules of reading

three graphic elements create a letter combination that is read in a special way; $S_{m/k}$ (S from French *son*—sound) is a phonetic unit (less often a combination of phonetic units) corresponding to one or two sounds.

For example, $L_1 = S_1$: *ç* = [*s*]; $L_2 + L_3 + L_4 = S_2$: *eau* = [*o*]; $L_3 + L_4 = S_2$: *au* = [*o*]; $L_5 + L_6 = S_3 + S_4$: *oi* = [*wa*]; $L_2 + L_7 + L_8 = S_0$: *-ent* as the ending of verbs in the 3rd person plural is not pronounced at all, i.e., S_0 is the absence of sounds.

After presenting students with one of the 48 reading rules, their understanding of it is tested. If they have not understood it, the rule is repeated in all subsequent lessons.

When modeling the didactic situation under consideration in this work, we examine a model for teaching the rules for writing and pronouncing French graphemes and their combinations, changing only the number of rules studied per week. Other parameters related to the characteristics of the teacher, teaching materials, psychological and mnemonic characteristics of students, etc., are not considered in this work. However, this variable seems to be the most important in teaching at the initial stage, since, on the one hand, without sufficiently rapid mastery of the rules of reading, it is impossible to effectively progress through the curriculum, and on the other hand, the average human memory is usually not capable of simultaneously perceiving and transferring large amounts of information into long-term memory. The hypothesis was that students at a technical university, who have a technical mindset and poorer linguistic memory than humanities students, should be taught more slowly, with the rules presented in portions. The calculations in this case were to find the optimal number of lessons and portions of reading rules taught to first-year students at RTU MIREA. Figure 1 describes the steps involved in teaching phonetics and reading. The rule R_n (R from *règle*) is the rule number n from a list consisting of 48 rules.

Each ith lesson covers q rules and repeats $q \cdot i$ rules, except for the first lesson, where no rules are repeated, and only the first set of rules is covered. Since students only repeat what they have already learned for the rest of the semester, this formula remains in effect until all rules have been covered.

In order to determine the optimal model for teaching French letter combinations, the effects of three teaching situations were studied: intensive study of all 48 rules over four weeks, slow study over the entire semester, and study at an average pace over half the semester. The most effective model was considered to be the one for which the share of correct answers when reading French words was the highest compared to the other models.

RESULTS

Model 1. When teaching French pronunciation rules for certain combinations of graphemes, six new rules are introduced in each lesson over the course of half a semester; in the subsequent, lesson they are reviewed along with six new ones, and so on, until all the reading rules have been covered. As can be seen in Fig. 2, by the ninth lesson, i.e., after studying all 48 rules and repeating most of them, the average student has still not sufficiently mastered the rules for reading letter combinations. Each subsequent lesson increases the share of correct answers given by students when reading

words containing a particular combination of graphemes pronounced in a specific way. Thus, by the end of the semester, the success rate in mastering the rules has increased to 93%.

Model 2. During intensive training in French pronunciation rules for certain grapheme combinations, 12 new rules are taught in each lesson over the course of a quarter; in the next lesson, these are reviewed along with 12 new ones, and so on, until all the reading rules have been covered. As can be seen in Fig. 3, by the fifth lesson, the rules for reading letter combinations have been poorly mastered. Each subsequent lesson again increases the share of correct answers from students; this is due to all rules being repeated multiple times. Thus, by the end of the semester, the success rate in mastering the rules has increased to 81%.

Model 3. When teaching French pronunciation rules slowly, three new rules are introduced in each lesson throughout the semester; in the subsequent lesson, they are repeated along with three new ones, and so on, until the end of the semester. Thus, the last three rules are studied in the very last lesson of the semester, as a result of which they cannot be repeated in subsequent lessons. As can be seen from Fig. 4, by the end of the semester, despite repeated repetition of a number of rules, not all rules could be fully mastered, which led to the total number of correct answers not reaching the required level.

Thus, intensive and slow learning in a non-language university led to lower performance at the end of the semester. Fewer students learned to read correctly than at the average learning speed, i.e., when all the rules were presented during the first half of the semester and repeated during the second half of the semester.

The proposed models illustrate average data regarding French language learning at the initial stage at RTU MIREA. However, it is important to take into account the high variability of both individual learning trajectories and the average level within a specific group, since, as experience shows, groups have different levels and characteristics. This variability is determined by a number of factors (F), including cognitive and linguistic abilities, psychological characteristics, motivational attitudes, workload, educational conditions, etc. [14]. Within the framework of constructing a prognostically correct model, students' mastery of French reading rules will be determined by the additive sum of the influence of all possible factors and the level of mastery of each specific rule. Consequently, the effectiveness indicator of the lesson (Fig. 5) is expressed in the number of correct answers recorded after mastering the reading rules, in accordance with the formula:

$$Ef = \sum_{i=1}^{n} F_i + \sum_{R=1}^{m} Z_R,$$

where F_i is one of n significant factors characterizing students' ability to master the rules of reading and pronouncing French grapheme combinations; Z_R is the accumulated knowledge of the student N regarding rules of reading (rules R from 1 to m).

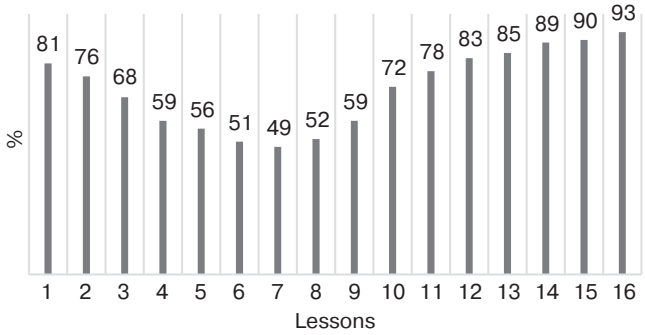


Fig. 2. Share of correct answers in lessons during the semester when studying 6 rules over 8 weeks (%)

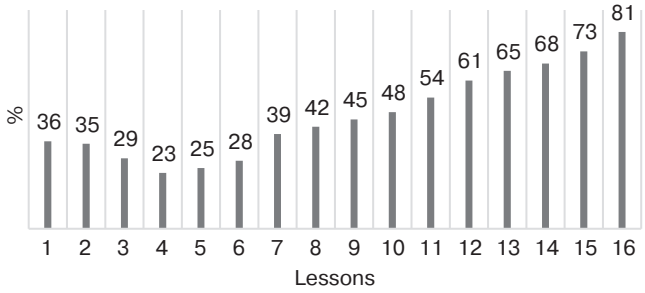


Fig. 3. Share of correct answers in lesson during the semester when studying 12 rules over 4 weeks (%)

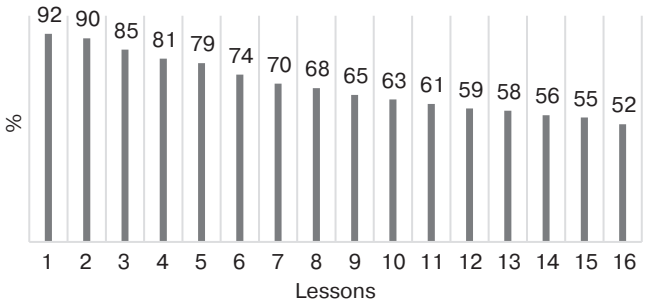


Fig. 4. Share of correct answers when studying 3 rules over 16 weeks (%)

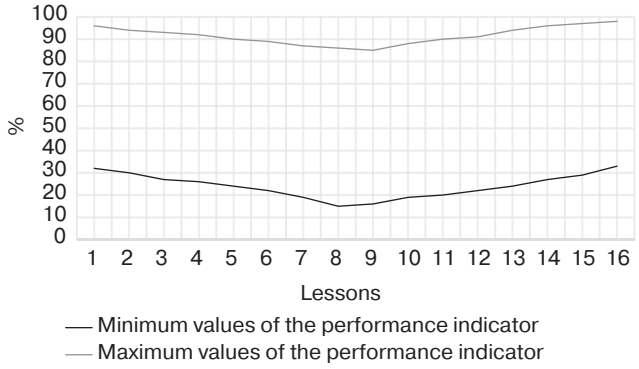


Fig. 5. Lesson effectiveness graph showing the share of correct answers over 16 weeks based on the effectiveness indicator when using Model 1

SUMMARY

The optimal approach to teaching French to beginners turned out to be a program in which reading rules are taught only in the first half of the semester and then regularly repeated in each lesson until the end. The study showed that students who studied using this model learned the material better than those who studied using the other two methods: slow learning throughout the semester and intensive learning of all the rules over four lessons.

In the first case, due to the extended study period, students did not have time to review some of the material at the end of the semester; moreover, fewer rules were reviewed in lesson. For example, Rules 1, 2, and 3 were practiced throughout the semester in all variants, while Rules 46, 47, and 48 were practiced for just over half the semester when studying the rules over 8 weeks and only during one lesson when studying the rules throughout the semester.

In the second case, intensive study led to confusion about the rules and cognitive overload, resulting in poorer memory retention [27].

Moreover, it is important to consider that various factors [28] influence the effectiveness of training, such as:

1) the student's intellectual and linguistic abilities;
2) level of anxiety and adaptability;
3) speed of information perception;
4) motivation (internal and external);
5) group size;
6) lesson schedule;
7) amount of independent work;
8) previous language learning experience.

In the presence of unfavorable factors (poor abilities, high anxiety, etc.), learning outcomes were low regardless of the model chosen. Conversely, under favorable conditions (good abilities, motivation, etc.), the learning model does not play a decisive role.

Considering the average level of students, the eight-week model (with learning the rules in the first half of the semester and then reviewing them) proved to be the most effective, so it can be recommended for use in the educational process.

CONCLUSIONS

The theoretical and practical significance of the study lies in the solution of the tasks set and the well-reasoned conclusions based on its results:

1) variables for formalizing the educational process have been defined;
2) options for models acceptable for assessing the educational aspects of French language teaching in a technological university have been selected;

3) certain models describing the process of mastering the pronunciation of French letter combinations have been analyzed;
4) it has been shown that the optimal model for studying the pronunciation rules of French letter combinations is one in which students learn all the rules in eight lessons, with repetition in subsequent lessons.

Didactic modeling of elements of the educational process enables the optimal planning of students' educational activities, including foreign language learning. The creation of a series of models covering all potential objects, concepts, forms, characteristics, patterns, parameters, categories, and factors of learning, and the generalization of the results obtained, make it possible to modernize curricula and plans, methods, and technologies.

A promising area for further research is to study the potential of the developed methodology as a basis for the development of didactic, mathematical, and computer models for other thematic blocks of the curriculum, as well as universal, adaptive, and effective forms, methods, and technologies for teaching foreign languages in non-language educational institutions.

Authors' contributions

N.I. Chernova—justification for the study concept, developing the experimental work algorithm, research findings synthesis.

E.A. Ivanova—analysis and synthesis of the literature, study of Russian and foreign publications on the issue, drafting the manuscript.

N.V. Katakhova—formulation of the conclusions, interpretation of the study results.

REFERENCES

1. Kytmanov A.A., Gorelova Yu.N., Zykova T.V., Pikhtilkova O.A., Pronina E.V. A conceptual approach to digital transformation of the educational process at a higher education institution. *Russian Technological Journal*. 2024;12(5):98–110. https://doi.org/10.32362/2500-316X-2024-12-5-98-110
2. Sysoev P.V. Personalized Learning Based on Artificial Intelligence: How Ready Are Modern Students for New Educational Opportunities. *Vysshee obrazovanie v Rossii = Higher Education in Russia*. 2025;34(2):51–71 (in Russ.). https://doi.org/10.31992/0869-3617-2025-34-2-51-71
3. Vonog V.V. Digital transformation of the foreign language training system in terms of educating students of engineering profile. *Vestnik Permskogo natsional'nogo issledovatel'skogo politekhnicheskogo universiteta. Problemy yazykoznaniya i pedagogiki* (*Vestnik PNIPU. Problemy yazykoznaniya i pedagogiki*) = *PNRPU Linguistics and Pedagogy Bulletin.* 2024;2:150–162 (in Russ.). Available from URL: https://cyberleninka.ru/article/n/tsifrovaya-transformatsiya-sistemy-inoyazychnoy-podgotovki-v-protsesse-obucheniya-studentov-inzhenernogo-profilya. Accessed April 24, 2026.
4. Teasley S.D., Kay M., Elkins S., Hammond J. User-Centered Design for a Student-Facing Dashboard Grounded in Learning Theory. In: Sahin M., Ifenthaler D. (Eds.). *Visualizations and Dashboards for Learning Analytics. Advances in Analytics for Learning and Teaching*. Cham, Switzerland: Springer; 2021. P. 191–212. https://doi.org/10.1007/978-3-030-81222-5_9
5. Arnold K.E., Pistilli M.D. Course Signals at Purdue: Using learning analytics to increase student success. In: *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge* (*LAK'12*). 2012. P. 267–270. https://doi.org/10.1145/2330601.2330666
6. Kiriy V.G., Chan Van An. About one Mathematical Model of Ambivalent System of Teaching in the Foreign Language. *Vestnik NGU. Seriya: Informatsionnye tekhnologii = Vestnik NSU. Series*: *Information Technologies*. 2010;8(1):45–53 (in Russ.). https://www.elibrary.ru/lalalb
7. Mayer R.V. Application of Mathematical and Computer Models in Learning Theory. In: *Theory and Practice of Pedagogical Activity*: *Experiments and Innovations:* Proceedings of the 17th International In-Person and Correspondence Scientific and Practical Conference. Moscow: 2019. P. 79–88 (in Russ.). https://elibrary.ru/rfbiii
8. Flegontov A.V., Dyuk V.A., Fomina I.K. Soft Knowledge and Fuzzy Systemology of the Humanities. *Programmnye produkty i sistemy = Software & Systems*. 2008;3:97–102 (in Russ.). https://elibrary.ru/jwdpvt
9. Reichle E.D. *Computational Models of Reading:* A Handbook. Oxford University Press; 2021, 608 p.
10. Kiseleva O.M. Mathematical Methods to Formalize Elements of the Educational Process. *Kontsept = Concept.* 2013;2:51–57 (in Russ.). https://elibrary.ru/pvxpov
11. Tishina E.M. Methodological Basis of the Principles of the Mathematical Model in the Learning Process. *Obrazovanie i nauka v sovremennom mire. Innovatsii = Education and Science in the Modern World. Innovations*. 2018;5(18):28–33 (in Russ.). https://elibrary.ru/xwezkx
12. Nainish L.A., Tishina E.M. Improvement of Training Process Efficiency by Mathematical Modeling Methods. *Vestnik Nizhegorodskogo universiteta im. N.I. Lobachevskogo = Vestnik of Lobachevsky University of Nizhni Novgorod*. 2008;2: 27–31 (in Russ.). https://elibrary.ru/ujxnbd
13. Chernova N.I., Ivanova E.A., Bogush N.B., Katakhova N.V. Technology for determining non-humanities university students' cognitive-and-psychological characteristics. *Russian Technological Journal*. 2023;11(3):104–116. https://doi.org/10.32362/2500-316X-2023-11-3-104-116
14. Barak O., Tsodyks M. Mathematical models of learning and what can be learned from them. *Curr. Opin. Neurobiol*. 2023;80:102721. https://doi.org/10.1016/j.conb.2023.102721

15. Kulikova O.V., Chuev N.P. Learning process planning with the help of mathematical modeling the quality study. *Fundamental'nye issledovaniya = Fundamental Research.* 2014;8-7:1658–1662 (in Russ.). https://elibrary.ru/swoibv
16. Kumar A., Benjamin A.S., Heathcote A., Steyvers M. Comparing models of learning and relearning in large-scale cognitive training data sets. *npj Sci. Learn*. 2022;7(1):24. https://doi.org/10.1038/s41539-022-00142-x
17. Murre J.M.J. How averaging individual curves transforms their shape: Mathematical analyses with application to learning and forgetting curves. *J. Math. Psychology*. 2023;117(5):102816. https://doi.org/10.1016/j.jmp.2023.102816
18. Zhu Y. A knowledge graph and BiLSTM-CRF-enabled intelligent adaptive learning model and its potential application. *Alexandria Eng. J*. 2024;91:305–320. https://doi.org/10.1016/j.aej.2024.02.011
19. Duan S., Chen K., Yang Y., Shi S. Research on Personalized Learning Recommendation Based on Subject Knowledge Graphs and Learner Portraits. In: Gan J., Pan Y., Zhou J., Liu D., Song X., Lu Z. (Eds.). *Computer Science and Educational Informatization. CSEI 2023. Communications in Computer and Information Science*. 2024. V. 1900. P. 367–374. https://doi.org/10.1007/978-981-99-9492-2_31
20. O'Donnell T.J., Hauser M.D., Tecumseh Fitch W. Using mathematical models of language experimentally. *Trends in Cogn. Sci*. 2005;9(6):284–289. https://doi.org/10.1016/j.tics.2005.04.011
21. Huang Y., Xu W., Sukjairungwattana P., Yu Zh. Learners' continuance intention in multimodal language learning education: An innovative multiple linear regression model. *Heliyon*. 2024;10(6):e28104. https://doi.org/10.1016/j.heliyon.2024.e28104
22. Ghasemabadi A., Soltanian N. Qualitative properties of mathematical model of English language education. *Adv. Differ. Equations*. 2021;202115. https://doi.org/10.1186/s13662-020-03180-0
23. Kiram J.J., Sulaiman J., Swanto S., Din W.A. Modeling the language learning strategies and English language proficiency of pre-university students in *UMS*. A case study. *AIP Conf. Proc*. 2015;1682(1):050004. https://doi.org/10.1063/1.4932495
24. Askarova A.H., Svetlov M.S., Svetlova M.K. Mathematical Model for the Process of Training. In: *Problems of Control, Processing and Transmission of Information:* Collection of works of the 4th International Scientific Conference in 2 v. Saratov: 2015. V. 2. P. 199–204 (in Russ.). https://elibrary.ru/woxvhh
25. Le V.Kh., Chernenkaya L.V., Nguyen T.T.Z. *Program for Solving the Inverse Problem of Restoring Parameters in a Mathematical Model of the Process of Teaching a Foreign Language:* Certificate of state registration of computer software No. 2023617166 Russian Federation. Publ. 06.04.2023 (in Russ.).
26. Ukhova T.V., Ukhov A.V. The Overview of Mathematical Models for a Foreign Language Learning. In: *Current Trends in Teaching Foreign Languages at Non-Linguistic Universities:* Proceedings of the 17th All-Russian Scientific and Methodological Conference with International Participation. Krasnoyarsk: 2023. P. 89–93 (in Russ.).
27. Ivanova E.A. Regulation of Cognitive Load in Teaching a Foreign Language at a Technological University. *Mir nauki. Pedagogika i psikhologiya = World of Science. Pedagogy and Psychology*. 2024;12(1) (in Russ.). Available from URL: https://mir-nauki.com/PDF/73PDMN124.pdf. Accessed April 24, 2024.
28. Bogush N.B., Ivanova E.A. Difficulties in the formation of Foreign Language Competences of Foreign Language Competences of a Technological University Students due to Non-Linguistic Factors. *Obrazovanie i pravo = Education and Law*. 2023;3:230–238 (in Russ.). https://doi.org/10.24412/2076-1503-2023-3-230-238

СПИСОК ЛИТЕРАТУРЫ

1. Кытманов А.А., Горелова Ю.Н., Зыкова Т.В., Пихтилькова О.А., Пронина Е.В. Концептуальный подход к цифровой трансформации образовательного процесса в вузе. *Russian Technological Journal*. 2024;12(5):98–110. https://doi.org/10.32362/2500-316X-2024-12-5-98-110
2. Сысоев П.В. Персонализированное обучение на основе технологий искусственного интеллекта: насколько готовы современные студенты к новым возможностям получения образования. *Высшее образование в России*. 2025;34(2): 51–71. https://doi.org/10.31992/0869-3617-2025-34-2-51-71
3. Воног В.В. Цифровая трансформация системы иноязычной подготовки в процессе обучения студентов инженерного профиля. *Вестник Пермского национального исследовательского политехнического университета. Проблемы языкознания и педагогики*. 2024;2:150–162. URL: https://cyberleninka.ru/article/n/tsifrovaya-transformatsiya-sistemy-inoyazychnoy-podgotovki-v-protsesse-obucheniya-studentov-inzhenernogo-profilya. Дата обращения 24.04.2026.
4. Teasley S.D., Kay M., Elkins S., Hammond J. User-Centered Design for a Student-Facing Dashboard Grounded in Learning Theory. In: Sahin M., Ifenthaler D. (Eds.). *Visualizations and Dashboards for Learning Analytics. Advances in Analytics for Learning and Teaching*. Cham, Switzerland: Springer; 2021. P. 191–212. https://doi.org/10.1007/978-3-030-81222-5_9
5. Arnold K.E., Pistilli M.D. Course Signals at Purdue: Using learning analytics to increase student success. In: *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge* (*LAK'12*). 2012. P. 267–270. https://doi.org/10.1145/2330601.2330666
6. Кирий В.Г., Чан Ван Ан. Об одной математической модели амбивалентной системы обучения неродному языку. *Вестник Новосибирского государственного университета. Серия: Информационные технологии* (*Вестник НГУ. Серия: Информационные технологии*). 2010;8(1):45–53. https://www.elibrary.ru/lalalb
7. Майер Р.В. Применение математических и компьютерных моделей в теории обучения. В сб: *Теория и практика педагогической деятельности*: *эксперименты и инновации:* сборник материалов XVII Международной очно-заочной научно-практической конференции. М.: НИЦ «Империя»; 2019. С. 79–88. https://elibrary.ru/rfbiii
8. Флегонтов А.В., Дюк В.А., Фомина И.К. Мягкие знания и нечеткая системология гуманитарных областей. *Программные продукты и системы*. 2008;3:97–102. https://elibrary.ru/jwdpvt

9. Reichle E.D. *Computational Models of Reading:* A Handbook. Oxford University Press; 2021, 608 p.
10. Киселева О.М. Использование математических методов для формализации элементов образовательного процесса. *Научно-методический электронный журнал «Концепт»*. 2013;2:51–57. https://elibrary.ru/pvxpov
11. Тишина Е.М. Методологические основы использования принципов математического моделирования в учебном процессе. *Образование и наука в современном мире. Инновации*. 2018;5(18):28–33. https://elibrary.ru/xwezkx
12. Найниш Л.А., Тишина Е.М. Повышение эффективности процесса обучения методами математического моделирования. *Вестник Нижегородского университета им. Н.И. Лобачевского*. 2008;2:27–31. https://elibrary.ru/ujxnbd
13. Чернова Н.И., Иванова Е.А., Богуш Н.Б., Катахова Н.В. Технология определения когнитивно-психологических особенностей студентов негуманитарного вуза. *Russian Technological Journal*. 2023;11(3):104–116. https://doi.org/10.32362/2500-316X-2023-11-3-104-116
14. Barak O., Tsodyks M. Mathematical models of learning and what can be learned from them. *Curr. Opin. Neurobiol*. 2023;80:102721. https://doi.org/10.1016/j.conb.2023.102721
15. Куликова О.В., Чуев Н.П. Проектирование учебного процесса на основе математического моделирования качества освоения дидактических единиц. *Фундаментальные исследования*. 2014;8-7:1658–1662. https://elibrary.ru/swoibv
16. Kumar A., Benjamin A.S., Heathcote A., Steyvers M. Comparing models of learning and relearning in large-scale cognitive training data sets. *npj Sci. Learn*. 2022;7(1):24. https://doi.org/10.1038/s41539-022-00142-x
17. Murre J.M.J. How averaging individual curves transforms their shape: Mathematical analyses with application to learning and forgetting curves. *J. Math. Psychology*. 2023;117(5):102816. https://doi.org/10.1016/j.jmp.2023.102816
18. Zhu Y. A knowledge graph and BiLSTM-CRF-enabled intelligent adaptive learning model and its potential application. *Alexandria Eng. J*. 2024;91:305–320. https://doi.org/10.1016/j.aej.2024.02.011
19. Duan S., Chen K., Yang Y., Shi S. Research on Personalized Learning Recommendation Based on Subject Knowledge Graphs and Learner Portraits. In: Gan J., Pan Y., Zhou J., Liu D., Song X., Lu Z. (Eds.). *Computer Science and Educational Informatization*. *CSEI 2023*. *Communications in Computer and Information Science*. 2024. V. 1900. P. 367–374. https://doi.org/10.1007/978-981-99-9492-2_31
20. O'Donnell T.J., Hauser M.D., Tecumseh Fitch W. Using mathematical models of language experimentally. *Trends in Cogn. Sci*. 2005;9(6):284–289. https://doi.org/10.1016/j.tics.2005.04.011
21. Huang Y., Xu W., Sukjairungwattana P., Yu Zh. Learners' continuance intention in multimodal language learning education: An innovative multiple linear regression model. *Heliyon*. 2024;10(6):e28104. https://doi.org/10.1016/j.heliyon.2024.e28104
22. Ghasemabadi A., Soltanian N. Qualitative properties of mathematical model of English language education. *Adv. Differ. Equations*. 2021;202115. https://doi.org/10.1186/s13662-020-03180-0
23. Kiram J.J., Sulaiman J., Swanto S., Din W.A. Modeling the language learning strategies and English language proficiency of pre-university students in *UMS.* A case study. *AIP Conf. Proc*. 2015;1682(1):050004. https://doi.org/10.1063/1.4932495
24. Аскарова А.Х., Светлов М.С., Светлова М.К. Математическая модель процесса обучения иностранным языкам. В сб.: *Проблемы управления, обработки и передачи информации:* сборник трудов IV Международной научной конференции в 2 т. Саратов, 22–25 сентября 2015 г. Т. 2. Саратов: Райт-Экспо; 2015, с. 199–204. https://elibrary.ru/woxvhh
25. Лэ В.Х., Черненькая Л.В., Нгуен Т.Т.З. *Программа решения обратной задачи восстановления параметров в математической модели процесса обучения неродному языку:* Свидетельство о государственной регистрации программы для ЭВМ № 2023617166 Российская Федерация. Заявка № 2023615774: заявл. 29.03.2023: опубл. 06.04.2023.
26. Ухова Т.В., Ухов А.В. Обзор математических моделей обучения иностранному языку. В сб.: *Современные тенденции в преподавании иностранных языков в неязыковом вузе:* Сборник статей XVII Всероссийской научно-методической конференции с международным участием. Красноярск, 18 мая 2023 г. Красноярск: Сибирский государственный университет науки и технологий имени академика М.Ф. Решетнева; 2023. С. 89–93.
27. Иванова Е.А. Регулирование когнитивной нагрузки при обучении иностранному языку в технологическом вузе. *Мир науки. Педагогика и психология*. 2024;12(1). URL: https://mir-nauki.com/PDF/73PDMN124.pdf. Дата обращения 24.04.2026.
28. Богуш Н.Б., Иванова Е.А. Трудности формирования иноязычных компетенций студентов технологического вуза, обусловленные нелингвистическими факторами. *Образование и право*. 2023;3:230–238. https://doi.org/10.24412/2076-1503-2023-3-230-238

About the Authors

Nadezhda I. Chernova, Dr. Sci. (Pedagog.), Professor, Head of the Foreign Languages Department, Institute of Radio Electronics and Informatics, MIREA – Russian Technological University (78, Vernadskogo pr., Moscow, 119454 Russia). E-mail: chernova@mirea.ru. Scopus Author ID 57194042371, RSCI SPIN-code 7391-4040, https://orcid.org/0000-0002-5685-9733

Ekaterina A. Ivanova, Cand. Sci. (Philolog.), Associate Professor, Foreign Languages Department, Institute of Radio Electronics and Informatics, MIREA – Russian Technological University (78, Vernadskogo pr., Moscow, 119454 Russia). E-mail: ivanova@mirea.ru. Scopus Author ID 57216646627, RSCI SPIN-code 8980-5591, https://orcid.org/0000-0001-6891-4966

Nataliya V. Katakhova, Cand. Sci. (Pedagog.), Associate Professor, Foreign Languages Department, Institute of Radio Electronics and Informatics, MIREA – Russian Technological University (78, Vernadskogo pr., Moscow, 119454 Russia). E-mail: katakhova@mirea.ru. Scopus Author ID 57204175929,RSCI SPIN-code 2552-5380, https://orcid.org/0000-0002-9416-5011

Об авторах

Чернова Надежда Ивановна, д.пед.н., профессор, заведующий кафедрой иностранных языков, Институт радиоэлектроники и информатики, ФГБОУ ВО «МИРЭА – Российский технологический университет» (119454, Россия, Москва, пр-т Вернадского, д. 78). E-mail: chernova@mirea.ru. Scopus Author ID 57194042371, SPIN-код РИНЦ 7391-4040, https://orcid.org/0000-0002-5685-9733

Иванова Екатерина Александровна, к.филол.н, доцент, кафедра иностранных языков, Институт радиоэлектроники и информатики, ФГБОУ ВО «МИРЭА – Российский технологический университет» (119454, Россия, Москва, пр-т Вернадского, д. 78). E-mail: ivanova@mirea.ru. Scopus Author ID 57216646627, SPIN-код РИНЦ 8980-5591, https://orcid.org/0000-0001-6891-4966

Катахова Наталия Владимировна, к.пед.н., доцент, кафедра иностранных языков, Институт радиоэлектроники и информатики, ФГБОУ ВО «МИРЭА – Российский технологический университет» (119454, Россия, Москва, пр-т Вернадского, д. 78). E-mail: katakhova@mirea.ru. Scopus Author ID 57204175929, SPIN-код РИНЦ 2552-5380, https://orcid.org/0000-0002-9416-5011

Translated from Russian into English by L. Bychkova
Edited for English language and spelling by Thomas A. Beavitt

MIREA – Russian Technological University.
78, Vernadskogo pr., Moscow, 119454 Russian Federation.
Publication date May 29, 2026.
Not for sale.

МИРЭА – Российский технологический университет.
119454, РФ, г. Москва, пр-т Вернадского, д. 78.
Дата опубликования 29.05.2026 г.
Не для продажи.

https://www.rtj-mirea.ru