

ISSN 1816-0301 (Print)
ISSN 2617-6963 (Online)

ИНФОРМАТИКА INFORMATICS

TOM 22
VOL.

4 | 2025

ОТ РЕДАКЦИИ

В журнале «Информатика» публикуются оригинальные и обзорные статьи, описывающие результаты фундаментальных и прикладных исследований специалистов академического и вузовского профиля в области информатики и информационных технологий.

Основной целью журнала является публикация наиболее значимых новых результатов в указанной области. Приветствуются статьи, описывающие заключительные результаты научных проектов и докторских исследований, открывающие новые направления исследований, которые находятся на стыке информатики и других наук.

Журнал рассчитан на широкий круг специалистов в области информатики и информационных технологий.

Основные разделы журнала:

- биоинформатика;
- математическое моделирование;
- защита информации и надежность систем;
- информационные технологии;
- логическое проектирование;
- обработка сигналов, изображений, речи, текста и распознавание образов;
- автоматизация проектирования;
- интеллектуальные системы.

Префикс DOI: 10.37661

Условия распространения материалов:

контент доступен под лицензией Creative Commons Attribution 4.0 License

Индексирование:

Высшей аттестационной комиссией Республики Беларусь журнал «Информатика» был включен в список научных изданий для опубликования результатов докторских исследований.

В декабре 2017 г. включен в базу данных Российского индекса научного цитирования (РИНЦ). С помощью инструментов и сервисов, доступных на платформе eLIBRARY (раздел «Личный кабинет»), можно самостоятельно корректировать список своих публикаций и цитирований в РИНЦ.

В июле 2017 г. включен в базу журналов открытого доступа Directory of Open Access Journals (DOAJ).

С помощью поисковых систем Google Scholar, WorldCat, Соционет можно получить свободный доступ к полному тексту научных публикаций журнала.

Адрес редакции:

ул. Сурганова, 6, к. 305, г. Минск, 220012, Беларусь
Тел. +375 (017) 351 26 22

Editorial address:

Surganova str., 6, of. 305, Minsk, 220012, Belarus
Phone +375 (017) 351 26 22

E-mail: rio@newman.bas-net.by

<https://inf.grid.by/jour>

THE EDITOR'S NOTE

The journal "Informatics" is a scientific publication in computer sciences and information technologies which reviews the results in basic and applied research of scientists from the universities and scientific centers.

The journal focuses on the most significant and modern papers of research projects results and PhD/DSc thesis in computer sciences.

The journal is edited for the specialists in IT and computer sciences research and application.

The main sections of the journal:

- bioinformatics;
- mathematical modeling;
- information protection and system reliability;
- information technology;
- logical design;
- signal, image, speech, text processing and pattern recognition;
- computer-aided design;
- artificial intelligence methods.

DOI Prefix: 10.37661

Distribution:

content is distributed under Creative Commons Attribution 4.0 License

Indexation:

the journal "Informatics" is in the list of scientific publications recommended by the Higher Attestation Commission of the Republic of Belarus for scientists to publish the results of PhD/DSc research.

In December 2017 the journal was included in the database of the Russian Science Citation Index (RISC) and provides free access to reviewed electronic scientific paper, improving scientific information traffic and also raising quotation of works of the authors (please use <https://elibrary.ru> or section for authors https://elibrary.ru_author_tools).

In July 2017 included in the database of open access journals Directory of Open Access Journals (DOAJ).

Using the Google Scholar, WorldCat, Соционет search engine, you can get free access to full text of scientific publications of magazine.

ОБЪЕДИНЕНИЙ ИНСТИТУТ ПРОБЛЕМ ИНФОРМАТИКИ
НАЦИОНАЛЬНОЙ АКАДЕМИИ НАУК БЕЛАРУСИ

ИНФОРМАТИКА

Informatika

Том 22, № 4, октябрь-декабрь 2025

Ежеквартальный научный журнал

Издаётся с января 2004 г.

Учредитель и издатель – государственное научное учреждение «Объединенный институт проблем информатики Национальной академии наук Беларусь» (ОИПИ НАН Беларусь)

Главный редактор

Кругликов Сергей Владимирович, д-р воен. наук, канд. техн. наук, доцент,
ОИПИ НАН Беларусь (Минск, Беларусь)

Заместитель главного редактора

Тузиков Александр Васильевич, д-р физ.-мат. наук, проф., чл.-корр. НАН Беларусь,
ОИПИ НАН Беларусь (Минск, Беларусь)

Редакционная коллегия

Абламейко Сергей Владимирович, д-р техн. наук, проф., академик НАН Беларусь, БГУ (Минск, Беларусь)

Анищенко Владимир Викторович, канд. техн. наук, доцент, ООО «СофтКлуб» (Минск, Беларусь)

Бибило Петр Николаевич, д-р техн. наук, проф., ОИПИ НАН Беларусь (Минск, Беларусь)

Бобов Михаил Никитич, д-р техн. наук, проф., БГУИР (Минск, Беларусь)

Долгий Александр Борисович, д-р техн. наук, проф., Высшая инженерная школа Бретани (Нант, Франция)

Дудин Александр Николаевич, д-р физ.-мат. наук, проф., БГУ (Минск, Беларусь)

Карпов Алексей Анатольевич, д-р техн. наук, доцент, СПИИРАН (Санкт-Петербург, Россия)

Ковалев Михаил Яковлевич, д-р физ.-мат. наук, проф., чл.-корр. НАН Беларусь, ОИПИ НАН Беларусь
(Минск, Беларусь)

Килин Сергей Яковлевич, д-р физ.-мат. наук, проф., академик НАН Беларусь, Центр «Квантовая оптика и квантовая информатика» Института физики им. Б. И. Степанова НАН Беларусь (Минск, Беларусь)

Краснопрошин Виктор Владимирович, д-р техн. наук, проф., БГУ (Минск, Беларусь)

Крот Александр Михайлович, д-р техн. наук, проф., ОИПИ НАН Беларусь (Минск, Беларусь)

Лиходед Николай Александрович, д-р физ.-мат. наук, проф., БГУ (Минск, Беларусь)

Матус Петр Павлович, д-р физ.-мат. наук, проф., Институт математики НАН Беларусь (Минск, Беларусь)

Скляров Валерий Анатольевич, д-р техн. наук, проф., Университет Авейру (Авейру, Португалия)

Сотков Юрий Назарович, д-р физ.-мат. наук, проф., ОИПИ НАН Беларусь (Минск, Беларусь)

Харин Юрий Семенович, д-р физ.-мат. наук, проф., академик НАН Беларусь, НИИ ППМИ БГУ
(Минск, Беларусь)

Черемисинова Людмила Дмитриевна, д-р техн. наук, проф., ОИПИ НАН Беларусь (Минск, Беларусь)

Чернявский Александр Федорович, д-р техн. наук, проф., академик НАН Беларусь, НИИ ПФП
им. А. Н. Севченко БГУ (Минск, Беларусь)

Ярмолик Вячеслав Николаевич, д-р техн. наук, проф., БГУИР (Минск, Беларусь)

Редакционный совет

Ефанов Дмитрий Викторович, Российский университет транспорта (Московский институт инженеров транспорта) (Москва, Россия)

Кумари Мадху, Университетский центр исследований и разработок, Университет Чандигарха (Мохали, Пенджаб, Индия)

Лазарев Александр Алексеевич, Институт проблем управления им. В. А. Трапезникова РАН (Москва, Россия)

Лай Цунг-Чьян, Азиатский университет в Тайчжуне (Китайская Народная Республика, Тайвань)

Марина Нинослав, Университет информационных наук и технологий им. Св. апостола Павла (Охрид, Македония)

Меликян Вазген Шаваршович, Национальный политехнический университет Армении (Ереван, Армения)

Пеш Эрвин, Зигенский университет (Зиген, Германия)

Сингх Таджиндер, Институт инженерии и технологий Сант Лонговал (Лонговал, Пенджаб, Индия)

Ходаченко Максим Леонидович, Институт космических исследований Австрийской академии наук (Грац, Австрия)

Чиулла Карло, Университет Эпока (Тирана, Албания)

Штейнберг Борис Яковлевич, Институт математики, механики и компьютерных наук Южного федерального университета (Ростов-на-Дону, Россия)

ИНФОРМАТИКА

Том 22, № 4, октябрь-декабрь 2025

Ответственный за выпуск *Мойсейчик Светлана Сергеевна*

Редактор *Гончаренко Галина Борисовна*

Компьютерная верстка *Бутевич Ольга Борисовна*

Сдано в набор 16.11.2025. Подписано в печать 18.12.2025. Формат 60×84 1/8. Бумага офсетная.
Гарнитура Таймс. Ризография. Усл. печ. л. 11,4. Уч.-изд. л. 11,2. Тираж 40 экз. Заказ 15.

Государственное научное учреждение «Объединенный институт проблем информатики
Национальной академии наук Беларусь».

Свидетельство о государственной регистрации издателя, изготовителя, распространителя печатных изданий
№ 1/274 от 04.04.2014. ЛП № 38200000016516 от 18.12.13. Ул. Сурганова, 6, 220012, Минск, Беларусь.

ISSN 1816-0301 (Print)
ISSN 2617-6963 (Online)

THE UNITED INSTITUTE OF INFORMATICS PROBLEMS
OF THE NATIONAL ACADEMY OF SCIENCES OF BELARUS

INFORMATICS

Vol. 22, no. 4, October-December 2025

Published quarterly

Issued since January 2004

Founder and publisher – State Scientific Institution "The United Institute of Informatics Problems of the National Academy of Sciences of Belarus" (UIIP NASB)

Editor-in-Chief

Sergey V. Kruglikov, D. Sc. (Milit.), Ph. D. (Eng.), Assoc. Prof., UIIP NASB (Minsk, Belarus)

Deputy Editor-in-Chief

Alexander V. Tuzikov, D. Sc. (Phys.-Math.), Prof., Corr. Member of NASB,
UIIP NASB (Minsk, Belarus)

Editorial Board

Sergey V. Ablameyko, D. Sc. (Eng.), Prof., Academician of NASB, BSU (Minsk, Belarus)

Uladimir V. Anishchanka, Ph. D. (Eng.), Assoc. Prof., SoftClub Ltd. (Minsk, Belarus)

Petr N. Bibilo, D. Sc. (Eng.), Prof., UIIP NASB (Minsk, Belarus)

Mikhail N. Bobov, D. Sc. (Eng.), Prof., BSUIR (Minsk, Belarus)

Alexandre B. Dolgui, D. Sc. (Eng.), Prof., IMT Atlantique (Nantes, France)

Alexander N. Dudin, D. Sc. (Phys.-Math.), Prof., BSU (Minsk, Belarus)

Alexey A. Karpov, D. Sc. (Eng.), Assoc. Prof., SPII RAS (Saint Petersburg, Russia)

Mikhail Y. Kovalyov, D. Sc. (Phys.-Math.), Prof., Corr. Member of NASB, UIIP NASB (Minsk, Belarus)

Sergey Ya. Kilin, D. Sc. (Phys.-Math.), Prof., Academician of NASB, Center of Quantum Optics and Quantum Information of B. I. Stepanov Institute of Physics NASB (Minsk, Belarus)

Viktor V. Krasnoproschin, D. Sc. (Eng.), Prof., BSU (Minsk, Belarus)

Alexander M. Krot, D. Sc. (Eng.), Prof., UIIP NASB (Minsk, Belarus)

Nikolai A. Likhoded, D. Sc. (Phys.-Math.), Prof., BSU (Minsk, Belarus)

Petr P. Matus, D. Sc. (Phys.-Math.), Prof., Institute of Mathematics of NASB (Minsk, Belarus)

Valery A. Sklyarov, D. Sc. (Eng.), Prof., University of Aveiro (Aveiro, Portugal)

Yuri N. Sotskov, D. Sc. (Phys.-Math.), Prof., UIIP NASB (Minsk, Belarus)

Yuriy S. Kharin, D. Sc. (Phys.-Math.), Prof., Academician of NASB, RI APMI BSU (Minsk, Belarus)

Ljudmila D. Cheremisinova, D. Sc. (Eng.), Prof., UIIP NASB (Minsk, Belarus)

Alexander F. Cherniavsky, D. Sc. (Eng.), Prof., Academician of NASB, A. N. Sevchenko IAPP BSU (Minsk, Belarus)

Vyacheslav N. Yarmolik, D. Sc. (Eng.), Prof., BSUIR (Minsk, Belarus)

Editorial Council

Dmitry V. Efanov, Russian University of Transport (Moscow Institute of Transport Engineers) (Moscow, Russia)
Madhu Kumari, University Center for Research & Development, Chandigarh University (Mohali, Punjab, India)
Alexander A. Lazarev, V. A. Trapeznikov Institute of Control Sciences of the RAS (Moscow, Russia)
Tsung-Chyan Lai, Asia University at Taichung (The People's Republic of China, Taiwan)
Ninoslav Marina, St. Paul the Apostle University of Information Sciences and Technology (Ohrid, Macedonia)
Vazgen Sh. Melikyan, National Polytechnic University of Armenia (Yerevan, Armenia)
Erwin Pesch, University of Siegen (Siegen, Germany)
Tajinder Singh, Sant Longowal Institute of Engineering & Technology (Longowal, Punjab, India)
Maxim L. Khodachenko, Space Research Institute, Austrian Academy of Sciences (Graz, Austria)
Carlo Ciulla, Epoka University (Tirana, Albania)
Boris Steinberg, Institute of Mathematics, Mechanics and Computer Science Southern Federal University (Rostov-on-Don, Russia)

INFORMATICS

Vol. 22, no. 4, October-December 2025

Issue Head *Sviatlana S. Maiseichyk*

Editor *Halina B. Hancharenka*

Computer Imposition *Volha B. Butsevich*

Sent for press 16.11.2025. Output 18.12.2025. Format 60×84 1/8. Offset paper. Headset Times. Riesography. Printed sheets 11,4. Publisher's signatures 11,2. Circulation 40 copies. Order 15.

State Scientific Institution "The United Institute of Informatics Problems of the National Academy of Sciences of Belarus".

Certificate on the state registration of the publisher, manufacturer, distributor of printing editions no. 1/274 dated 04.04.2014. License for the press no. 38200000016516 dated 18.12.13.
6, Surganova Str., 220012, Minsk, Belarus.

ISSN 1816-0301 (Print)
ISSN 2617-6963 (Online)

СОДЕРЖАНИЕ

МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ

Гущинский Н. Н., Ковалев М. Я., Розин Б. М. О совместной оптимизации размещения зарядных станций, маршрутизации электробусов и расписаний зарядки.....	7
Osipova Yu. A. Latent factor analysis of regional growth in Belarus using structural equation modeling	24

ОБРАБОТКА СИГНАЛОВ, ИЗОБРАЖЕНИЙ, РЕЧИ, ТЕКСТА И РАСПОЗНАВАНИЕ ОБРАЗОВ

Кривальцевич Е. А., Вашкевич М. И. Нейронные сети на основе обучаемого двумерного разделимого преобразования для классификации изображений: теория и аппаратная реализация на FPGA	36
Курочка К. С., Ёвженко Ю. Д. Технология трансляции естественно-языковых правил мерчандайзинга в цифровые планограммы.....	55

ЗАЩИТА ИНФОРМАЦИИ И НАДЕЖНОСТЬ СИСТЕМ

Иванюк А. А., Бурко Л. А. Генерирование детерминированных идентификаторов и случайных чисел на основе схемы конфигурируемого кольцевого осциллятора	65
---	----

ЛОГИЧЕСКОЕ ПРОЕКТИРОВАНИЕ

Черемисинова Л. Д. Распознавание подсхем трехстабильных элементов в КМОП СБИС	82
---	----

УЧЕНЫЕ БЕЛАРУСИ

Крючков Александр Николаевич (к 75-летию со дня рождения)	94
---	----

ISSN 1816-0301 (Print)
ISSN 2617-6963 (Online)

CONTENTS

MATHEMATICAL MODELING

Guschinsky N. N., Kovalyov M. Y., Rozin B. M. On the joint optimization of charging stations location, electric bus routing and charging scheduling 7
Osipova Yu. A. Latent factor analysis of regional growth in Belarus using structural equation modeling 24

SIGNAL, IMAGE, SPEECH, TEXT PROCESSING AND PATTERN RECOGNITION

Krivalcevich E. A., Vashkevich M. I. Neural networks based on a learnable two-dimensional separable transform for image classification: theory and hardware implementation on FPGA 36
Kurochka K. S., Youzhanka Y. D. The technology of translating natural language merchandising rules into digital planograms 55

INFORMATION PROTECTION AND SYSTEM RELIABILITY

Ivaniuk A. A., Burko L. A. Generation of deterministic identifiers and random numbers using a configurable ring oscillator circuit 65

LOGICAL DESIGN

Cheremisinova L. D. Recognition of subcircuits of tristable elements in CMOS VLSI 82

SCIENTISTS OF BELARUS

Kruchkov Alexander Nikolaevich (on the 75th anniversary of his birth) 94

МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ MATHEMATICAL MODELING



УДК 519.8; 621.3:007; 621.3:001.891.57

DOI: 10.37661/1816-0301-2025-22-4-7-23

Оригинальная статья

Original Article

О совместной оптимизации размещения зарядных станций, маршрутизации электробусов и расписаний зарядки

Н. Н. Гущинский, М. Я. Ковалев, Б. М. Розин[✉]

Объединенный институт проблем информатики

Национальной академии наук Беларусь,

ул. Сурганова, 6, Минск, 220012, Беларусь

[✉]E-mail: rozin@newman.bas-net.by

Аннотация

Цели. При замене парка дизельных автобусов парком электробусов актуальной является задача планирования состава парка, зарядной инфраструктуры, ориентированной на быструю подзарядку на терминалах маршрутов, и расписаний зарядки электробусов. Цель исследования заключается в создании моделей и методов выработки экономически эффективных решений по выбору парка электробусов, зарядной инфраструктуры такого типа с учетом ряда специфических условий. Функционирование парка и зарядной инфраструктуры моделируется для терминалов маршрутов в представительном интервале дня, характеризующемся наибольшей интенсивностью пассажиропотока и максимальным расходом электроэнергии.

Методы. Используются методы теории множеств, теории графов, математического программирования. **Результаты.** Разработана математическая модель задачи оптимизации однородного парка электробусов, назначения электробусам сервисных рейсов, размещения и количества зарядных станций быстрой зарядки на терминалах маршрутов и расписания зарядки на них электробусов. В качестве целевой функции выбрана суммарная дневная стоимость электробусов, зарядных станций и потребленной электроэнергии. Предложена двухуровневая декомпозиционная схема решения задачи, на верхнем уровне которой осуществляется выбор назначения электробусов на сервисные рейсы из заданного набора, а на нижнем определяется инфраструктура однородных зарядных станций терминалов и расписания зарядки на них электробусов при фиксированном назначении электробусам сервисных рейсов. Для решения подзадачи верхнего уровня предложен эвристический рандомизированный алгоритм, для подзадачи нижнего уровня разработана модель смешанного целочисленного линейного программирования.

Заключение. Решение подзадачи верхнего уровня заключается в выборе назначения электробусов на сервисные рейсы из заданного набора. Для решения сформулированной подзадачи нижнего уровня могут использоваться стандартные решатели типа IBM ILOG CPLEX, Gurobi Optimizer.

Ключевые слова: электробус, аккумуляторная батарея, станция зарядки, оптимизация, смешанное целочисленное линейное программирование

Благодарности. Исследование выполнялось в рамках проекта «Модели и методы для систем поддержки принятия решений при проектировании, планировании и управлении высокотехнологичными, роботизированными производствами и электротранспортом» программы ГПНИ «Цифровые и космические технологии, безопасность человека, общества и государства».

Для цитирования. Гущинский, Н. Н. О совместной оптимизации размещения зарядных станций, маршрутизации электробусов и расписаний зарядки / Н. Н. Гущинский, М. Я. Ковалев, Б. М. Розин // Информатика. – 2025. – Т. 22, № 4. – С. 7–23. – DOI: 10.37661/1816-0301-2025-22-4-7-23.

Конфликт интересов. Авторы заявляют об отсутствии конфликта интересов.

Поступила в редакцию | Received 22.08.2025

Подписана в печать | Accepted 12.09.2025

Опубликована | Published 30.12.2025

On the joint optimization of charging stations location, electric bus routing and charging scheduling

Nikolai N. Guschinsky, Mikhail Y. Kovalyov, Boris M. Rozin[✉]

*The United Institute of Informatics Problems
of the National Academy of Sciences of Belarus,
st. Surganova, 6, Minsk, 220012, Belarus*

[✉]E-mail: rozin@newman.bas-net.by

Abstract

Objectives. When replacing a fleet of diesel buses with a fleet of electric buses, the problem of planning the fleet composition, charging infrastructure oriented towards fast recharging at route terminals, and electric bus charging schedules is relevant. The purpose of the study is to develop models and methods for elaboration of cost-effective solutions for selecting a fleet of electric buses, charging infrastructure of this type, taking into account a number of specific conditions. The functioning of the fleet and charging infrastructure is modeled for route terminals during the most representative period of the day, characterized by the highest passenger flow intensity and maximum energy consumption.

Methods. Methods of set theory, graph theory and mathematical programming are used.

Results. A mathematical model has been developed for the optimization problem of a homogeneous electric bus fleet, assignment of service trips to electric buses, location and quantity of fast charging stations at route terminals, and a schedule for charging electric buses on them. The total daily cost of electric buses, charging stations, and consumed electricity has been selected as the objective function. A two-level decomposition scheme for solving the problem has been proposed, at the upper level of which the assignment of electric buses to service trips from a given set is selected, and at the lower level, the infrastructure of homogeneous charging stations of terminals and the charging schedules of electric buses on them are determined with a fixed assignment of service trips to electric buses. A heuristic randomized algorithm has been proposed to solve the upper-level subproblem, and a mixed integer linear programming model has been developed for the lower-level subproblem.

Conclusion. The solution to the upper-level subproblem consists of selecting the assignment of electric buses to service trips from a given set. Standard solvers such as IBM ILOG CPLEX, Gurobi Optimizer can be used to solve the formulated lower-level subproblem.

Keywords: electric bus, battery, charging station, optimization, mixed integer linear programming

Acknowledgements. The study was carried out within the Project "Models and methods for decision support systems in the design, planning and management of high-tech, robotic production and electric transport" in frames of the State research program "Digital and space technologies, human, society and state safety".

For citation. Guschinsky N. N., Kovalyov M. Y., Rozin B. M. *On the joint optimization of charging stations location, electric bus routing and charging scheduling*. Informatika [Informatics], 2025, vol. 22, no. 4, pp. 7–23 (In Russ.). DOI: 10.37661/1816-0301-2025-22-4-7-23.

Conflict of interest. The authors declare of no conflict of interest.

Введение. В последние годы значительное внимание в литературе уделяется различным аспектам планирования парка и зарядной инфраструктуры для городского электротранспорта, в частности электробусов [1–5]. Применяемые технологии зарядки электробусов варьируются в широком диапазоне (URL: <https://zeeus.eu/uploads/publications/documents/zeeus-report2017-2018-final.pdf>): электробусы могут заряжаться только в депо, либо в депо и на конечных остановках маршрутов, либо в депо, на конечных и промежуточных автобусных остановках. На практике чаще всего применяются технологии быстрой зарядки аккумуляторных батарей (далее *батарея*) электробусов на остановках маршрутов и медленной зарядки в депо, как правило, в ночном периоде. Для зарядки на маршруте обычно требуются зарядные устройства большой мощности (до 400–500 кВт), в то время как для зарядки в депо могут использоваться зарядные устройства малой мощности (80–100 кВт) [2, 3]. Наиболее распространена в последнее время технология быстрой зарядки электробусов на остановках маршрутов [4, 5].

При выборе парка электробусов и их зарядной инфраструктуры требуется учитывать множество факторов, в том числе расписание выполняемых сервисных рейсов перевозки пассажиров по маршрутам, расход электроэнергии в зависимости от характеристик используемых электробусов, маршрутов, пассажиропотока и особенностей эксплуатации, размещение и пропускную способность зарядных станций, стоимостные характеристики электробусов, зарядных станций, тарифов и т. д. Также по мере роста парков электробусов, обслуживающих городские маршруты, возрастаёт нагрузка на соответствующие энергосистемы, что требует принятия нетривиальных решений по сглаживанию пиковых нагрузок.

Электробусы обычно отправляются в путь из депо и возвращаются в то же или другое депо в конце дня. Для восстановления затраченной при выполнении рейсов электроэнергии используются технологии медленной зарядки в ночное время в депо и быстрой подзарядки в течение дня между рейсами. После ночной зарядки в депо электробусы на следующий день отправляются с полностью заряженной батареей. При быстрой зарядке в течение дня уровень заряда батареи электробуса может быть восстановлен частично в любом месте размещения зарядных станций до уровня, достаточного для выполнения следующего рейса.

В работе исследуется комплексная задача оптимального выбора назначения электробусов парка на сервисные рейсы из заданного набора; мест размещения из множества возможных вариантов и количества идентичных зарядных станций, устанавливаемых в выбранных местах; расписаний зарядки электробусов на зарядных станциях в условиях их ограниченной пропускной способности.

Предлагаемый подход к решению комплексной задачи заключается в двухуровневой декомпозиции исходной задачи на подзадачу нижнего уровня выбора оптимальных зарядной инфраструктуры и расписаний зарядки электробусов при фиксированном наборе выполняемых каждым электробусом парка сервисных рейсов и подзадачу верхнего уровня – выбора назначения электробусам парка сервисных рейсов.

1. Обзор литературы. Многие исследования в области планирования электротранспортных систем в последнее время посвящены комплексным задачам совместной оптимизации парка электробусов, инфраструктуры зарядных станций, назначения электробусов на сервисные рейсы согласно заданному расписанию и расписаний зарядки электробусов.

В частности, в работе [6] для системы маршрутов с многими терминалами исследуется задача совместной оптимизации мест зарядки для однородного парка электробусов, диспетчеризации заданного набора рейсов электробусов и расписаний их быстрой подзарядки на зарядных станциях терминалов. Учитываются динамические тарифы на электроэнергию, возможности переключения электробусов между маршрутами и возможность управления скоростью зарядки. Переключение в течение дня электробуса с одного обслуживаемого маршрута на другой позволяет в ряде случаев сократить количество используемых электробусов за счет применения дополнительных порожних (без пассажиров) рейсов между терминалами. Последовательность переключений электробуса в течение дня определяет последовательность выполняемых им сервисных и порожних рейсов, каждый из которых определяется начальным и конечным терминалами. Любой электробус может заряжаться на последовательности временных интервалов в промежутке между сервисными рейсами с различными кусочно-постоянными скоростями

зарядки. Зарядка в депо не рассматривается. В качестве цели выбрана минимизация капитальных затрат на электробусы и зарядные станции, а также эксплуатационных расходов на порожние рейсы, потраченную электроэнергию и затраты на контрактную мощность сети для зарядных станций. В работе не учитывается ограниченность поставляемой энергосетью мощности для каждого места размещения зарядной инфраструктуры, как и предельное количество зарядных станций. Соответственно, исключены из рассмотрения вопросы конкуренции электробусов за ресурсы времени на зарядных станциях, что неоправданно упрощает задачу. Предложена формулировка смешанного целочисленного линейного программирования (СЦЛП) для этой задачи. Для подзадачи поиска мест размещения зарядных станций и назначения электробусов на поездки разработана процедура локального поиска.

В статье [7] рассматривается задача выбора расписания и назначения электробусов на сервисные рейсы для системы маршрутов с многими терминалами, а также выбора мест установки зарядных станций с целью минимизации количества используемых однородных электробусов при выполнении ограничений, касающихся маршрутизации и зарядки. Заданное множество сервисных рейсов, как и в работе [6], разделено по маршрутным линиям, определяемым начальным и конечным терминалами. Каждая маршрутная линия между терминалами характеризуется интервалом между рейсами и числом рейсов. Каждый сервисный рейс характеризуется длительностью, расстоянием и затратами энергии. Интервал между соседними рейсами является параметром, который может принимать фиксированное значение либо выбираться из заданного диапазона значений. Предложена формулировка задачи в форме СЦЛП. В этой формулировке имеется ограничение на количество используемых зарядных станций, но при этом не учитывается возможность конфликтов электробусов различных линий за временные ресурсы зарядных станций, что может приводить к некорректным решениям.

В статье [8] исследуется задача диспетчеризации электробусов и дизельных автобусов, а также планирования зарядки электробусов на одном терминале, оборудованном заданными зарядными устройствами. Для ее решения использовалась двухуровневая декомпозиционная схема, где в координирующей подзадаче верхнего уровня на базе метода типа Бендерса отыскивалось расписание рейсов электробусов с учетом ограничений на мощность зарядных станций, а во вспомогательной подзадаче нижнего уровня определялось расписание операций зарядки для заданных расписаний рейсов электробусов. Вопрос выбора количества зарядных станций либо их пропускной способности не рассматривался.

В ряде исследований подобных задач, например в [6], отмечается высокая вычислительная трудоемкость решения такой комплексной задачи. Поэтому многие авторы, чтобы иметь возможность получать приемлемое по качеству решение в условиях ограниченных ресурсов времени и вычислительных мощностей, рассматривают решение отдельных компонентов задачи либо их комбинаций.

Анализ известных результатов для комплексных моделей оптимального планирования электротранспорта, включающих выбор парка, мест размещения зарядной инфраструктуры и расписаний зарядки, выявил недостаток внимания к необходимости сокращения количества используемых зарядных станций в условиях ограниченной пропускной способности зарядной инфраструктуры и конкуренции электробусов за ресурсы времени зарядки.

2. Постановка задачи. Для выполнения заданного набора $I = \{1, 2, \dots, n\}$ сервисных рейсов используются идентичные электробусы из множества $V = \{v_1, v_2, \dots, v_n\}$, оснащенные идентичными батареями энергетической емкости E (кВт·ч), с одинаковым запасом хода, которые могут выполнять любые рейсы из I . Для каждого сервисного рейса $i \in I$ заданы его стартовый a_i и конечный b_i терминалы, моменты времени α_i начала и β_i окончания рейса и величина g_i расходуемого относительного заряда батареи электробуса (далее *заряд*), измеряемого в долях ее энергетической емкости. Моменты начала и окончания каждого сервисного рейса детерминированы. В дальнейшем (если не определено иное) моменты времени и длительности измеряются в часах.

Допускаются переключения обслуживания электробуса с одного маршрута на другой, т. е. электробус может обслуживать запланированные рейсы по нескольким маршрутам. Пара сервисных рейсов (i, j) , $i, j \in I$ при $a_j = b_i$ (стартовый терминал рейса j является конечным термина-

лом рейса i), совместимы, если $\beta_i \leq \alpha_j$. Паре сервисных рейсов (i, j) , $i, j \in I$, может соответствовать порожний рейс из терминала b_i в терминал a_j при $a_j \neq b_i$, если $\beta_i + \gamma_{i,j} \leq \alpha_j$, где $\gamma_{i,j} > 0$ и $g_{i,j} > 0$ – величины затрачиваемых электробусом в порожнем рейсе времени и заряда соответственно. При этом условии сервисные рейсы i, j также считаются совместимыми. На любом из терминалов b_i либо a_j (при $a_j \neq b_i$) или на обоих этих терминалах электробус, выполняющий рейсы i, j , может восстанавливать потраченный заряд батареи, если на соответствующем терминале установлены зарядные станции и интервал пребывания электробуса на терминале позволяет выполнить операцию зарядки. Уровень заряда в процессе функционирования электробуса должен оставаться в рамках допустимого диапазона $[u_{\min}, u_{\max}]$, где $0 < u_{\min}, u_{\max} \leq 1$, u_{\max} – уровень заряда полностью заряженной батареи. Любой терминал $\theta \in \Theta = \{a_i, b_i \mid i \in I\}$ может быть укомплектован идентичными зарядными станциями из набора $C_\theta = \{c_1, c_2, \dots, c_{|C_\theta|}\}$ с предельным их количеством $|C_\theta|$ в зависимости от θ . Любой электробус может заряжаться при каждом посещении любого терминала не более одного раза не более чем на одной зарядной станции. В каждый момент времени на одной зарядной станции может заряжаться не более одного электробуса. Каждой операции зарядки электробуса предшествует операция настройки длительности Δ . Длительность любой операции зарядки должна быть не меньше минимальной величины $t_{C_{\min}}$.

Будем говорить, что терминалу b_i (либо a_j при $a_j = b_i$) соответствует единственная действительная опция зарядки электробуса, последовательно выполняющего рейсы i, j , если выполняется условие $\beta_i + \Delta + t_{C_{\min}} \leq \alpha_j$, иначе такая опция отсутствует (будем называть ее *пустой*). Для случая, когда $a_j \neq b_i$, электробусу, выполняющему совместимые рейсы $i, j \in I$, в интервале между этими рейсами соответствует единственная действительная опция зарядки (либо на терминале b_i , либо на терминале a_j) при выполнении условий $\beta_i + \gamma_{i,j} + \Delta + t_{C_{\min}} \leq \alpha_j < \beta_i + \gamma_{i,j} + 2(\Delta + t_{C_{\min}})$. Если же выполняется условие $\beta_i + \gamma_{i,j} + 2(\Delta + t_{C_{\min}}) \leq \alpha_j$, то и терминалу b_i , и терминалу a_j соответствует по одной действительной опции зарядки. Другими словами, в интервале между рейсами i, j возможно выполнение двух операций зарядки (одна на зарядной станции терминала b_i , другая на зарядной станции терминала a_j). Таким образом, для единственного терминала, посещаемого электробусом между двумя сервисными рейсами, может существовать единственная действительная опция зарядки. При посещении электробусом двух терминалов между сервисными рейсами этим терминалам могут соответствовать две пустые опции; может соответствовать одна действительная опция зарядки (на одном из терминалов), а вторая – пустая либо могут соответствовать по одной действительной опции зарядки на каждом из терминалов.

Определим взаимно-однозначное отображение z подмножества $V(z) = \{v_1, v_2, \dots, v_{\eta(z)}\} \subseteq V$, где $\eta(z) \leq n$, на множество I , сопоставляющее электробусам $v \in V(z)$ упорядоченные по возрастанию времени выполнения подмножества $I_v(z) = \{j_{v,1}, j_{v,2}, \dots, j_{v,l_v}\}$ совместимых сервисных рейсов $j_{v,l} \in I$, $(\beta_{j_{v,l}} + \gamma_{j_{v,l}, j_{v,l+1}} \leq \alpha_{j_{v,l+1}})$, $l=1,2,\dots, l_v-1$, такие, что $I_v(z) \cap I_{v'}(z) = \emptyset$ для $v \neq v'$, $v, v' \in V(z)$ и $\bigcup_{v \in V(z)} I_v(z) = I$.

Пусть Z – множество таких взаимно-однозначных отображений.

При фиксированном отображении z каждое подмножество $I_v(z)$ однозначно определяет последовательность опций зарядки электробуса $v \in V(z)$. Эта последовательность опций зарядки электробуса v включает также опцию зарядки на начальном терминале $a_{j_{v,1}}$ первого сервисного рейса $j_{v,1}$ для восстановления заряда, потраченного в порожнем рейсе из депо.

Обозначим $Q_\theta(z)$ множество всех опций зарядки (включая пустые) всех электробусов $v \in V$ в течение дня на терминале θ при фиксированном z , $\theta \in \Theta$. Восстановление уровня заряда на любой зарядной станции $c \in C_\theta$, $\theta \in \Theta$, в допустимом диапазоне $[u_{\min}, u_{\max}]$ осуществляется на постоянной мощности $P = E \cdot f$ (кВт) со скоростью f единиц заряда в час (е.з./ч). Таким образом, максимальная длительность зарядки батареи для любого электробуса составляет

$$\tau_{\max} = \frac{u_{\max} - u_{\min}}{f}.$$

Первоначально каждый электробус размещается в одном депо из заданного множества $D = \{d_1, d_2, \dots, d_m\}$, в котором он восстанавливает заряд батареи в ночной период до максимального уровня u_{\max} . Процесс зарядки электробусов парка в депо не рассматривается. В течение суток электробусы посещают депо однократно. Электробус не обязан возвращаться в свое отправное депо в конце дня. Однако, поскольку дневные расписания обслуживания сервисных рейсов являются периодическими, требуется, чтобы начальное и конечное распределения количеств электробусов по депо совпадали. Первый рейс электробуса из депо $d \in D$ в начальный терминал сервисного рейса $i \in I$, как и последний рейс из конечного терминала последнего сервисного рейса $j \in I$ в депо $d' \in D$, являются порожними и характеризуются длительностями $\gamma_{d, i}$, $\gamma_{j, d'}$ и величинами $g_{d, i}$, $g_{j, d'}$ расхода заряда соответственно.

Предполагается, что все сервисные рейсы и все возможные опции зарядки электробуса в течение дня выполняются в рамках заданного диапазона $[\underline{\Gamma}, \bar{\Gamma}]$. Считаем выполненным достаточное для этого условие $0 \leq \underline{\Gamma} \leq \alpha_j - \tau_{\max} < \alpha_j < \beta_j < \beta_j + \tau_{\max} \leq \bar{\Gamma} \leq 24, j \in I$.

Суточная стоимость электробуса c^{bus} и зарядной станции c^{ch} заданы. Дневной тариф c^{el} на электроэнергию в течение дня не изменяется.

Дневной (повторяющийся) период $[\underline{\Gamma}, \bar{\Gamma}]$ работы транспорта разбивается на несколько интервалов, каждый из которых характеризуется постоянной частотой выхода электробусов на маршрут, отличной от других интервалов. *Представительный интервал* дня, относящийся к одному из интервалов часов пик, характеризуется максимальным пассажиропотоком и интенсивностью его обслуживания с наибольшей частотой рейсов для любой маршрутной линии. *Представительный день* года (для Европы, как правило, рабочий день февраля) характеризуется наибольшим расходом электробусами электроэнергии вследствие неблагоприятных погодных условий и дорожной обстановки. Пропускная способность зарядной инфраструктуры, обеспечивающая функционирование парка электробусов в представительном интервале такого дня, позволяет гарантировать его функционирование в любом другом интервале с меньшими частотами рейсов любого дня года. Для сокращения сложности задачи рассматривается только часть дневного расписания событий зарядки электробусов парка в выделенном представительном интервале из периода $[\underline{\Gamma}, \bar{\Gamma}]$. Полученное в результате решения задачи расписание зарядки электробусов парка для представительного интервала позволяет оценить расходы на потребленную парком электроэнергию других интервалов (пиковых и непиковых) периода $[\underline{\Gamma}, \bar{\Gamma}]$ за счет использования поправочных коэффициентов, учитывающих различные частоты выхода электробусов на маршруты.

Исследуемая задача заключается в назначении выбранным электробусам упорядоченных подмножеств сервисных рейсов, выборе мест размещения и количеств зарядных станций, а также расписаний зарядки на них электробусов, минимизирующих суточные капитальные и операционные затраты на выбранные электробусы, зарядные станции и потребленную парком электроэнергию в периоде $[\underline{\Gamma}, \bar{\Gamma}]$. Назначение электробусам подмножеств сервисных рейсов должно обеспечивать однократное выполнение всех запланированных в периоде $[\underline{\Gamma}, \bar{\Gamma}]$ сервисных рейсов. Размещение зарядных станций на терминалах должно обеспечивать возможность зарядки электробусов для выполнения назначенных им рейсов, а расписания их зарядки должны учитывать приведенные выше условия и ограничения по времени и уровню заряда при приходе и выходе из терминалов, совместимость выполнения операций зарядки электробусов на каждой зарядной станции каждого терминала.

3. Декомпозиционная схема решения задачи. Предлагается следующая двухуровневая декомпозиционная схема решения задачи. Подзадача нижнего уровня заключается в оптимизации размещения и количеств зарядных станций терминалов, определении стартовых и финишных депо для всех используемых электробусов и расписаний их зарядки на зарядных станциях терминалов при фиксированном наборе упорядоченных подмножеств сервисных рейсов. В каче-

стве целевой функции подзадачи нижнего уровня используются общие дневные затраты на зарядные станции и потребленную на зарядных станциях терминалов электроэнергию.

На верхнем уровне решается подзадача выбора для электробусов парка набора последовательностей совместимых сервисных рейсов. При выборе набора последовательностей таких рейсов обеспечиваются достаточные интервалы пребывания электробусов на терминалах для восстановления уровня их заряда в промежутках между рейсами. Возможная конкуренция за ресурсы зарядных станций терминалов не учитывается. Целевая функция для подзадачи верхнего уровня выражается суммой оптимального значения целевой функции подзадачи нижнего уровня и суточной стоимости всех использованных электробусов парка.

Для решения подзадач нижнего уровня предлагается новая модель СЦЛП, описание которой приводится в подразд. 3.1. Для решения подзадачи верхнего уровня используется и эвристический алгоритм, описываемый в подразд. 3.2.

3.1. Подзадача нижнего уровня. Модель смешанного целочисленного линейного программирования выбора расположения и количеств зарядных станций, расписания зарядки электробусов и распределения электробусов по депо при фиксированном наборе последовательностей сервисных рейсов. Фиксированному z соответствует набор из $\eta(z)$ упорядоченных подмножеств (последовательностей) $I_v(z) = \{j_{v,1}, j_{v,2}, \dots, j_{v,l}, \dots, j_{v,\bar{l}_v}\}$, $l = 1, 2, \dots, \bar{l}_v$, сервисных рейсов, выполняемых электробусами $v \in V(z) = \{v_1, v_2, \dots, \eta(z)\}$. Последовательность $I_v(z)$ однозначно определяет для каждой пары $j_{v,l}, j_{v,l+1}$ смежных сервисных рейсов наличие порожнего рейса (при $a_{j_{v,l+1}} \neq b_{j_{v,l}}$) либо его отсутствие (при $a_{j_{v,l+1}} = b_{j_{v,l}}$), $l \in \{1, 2, \dots, \bar{l}_v - 1\}$. В дальнейшем z считается фиксированным и опущено для краткости изложения.

Сформируем общую последовательность $\pi_v = (\pi_{v,1}, \pi_{v,2}, \dots, \pi_{v,k}, \dots, \pi_{v,\hat{k}_v})$ сервисных и порожних рейсов, выполняемых электробусами $v \in V$, $k = 1, 2, \dots, \bar{k}_v$. Напомним, что первый рейс $\pi_{v,1}$ последовательности π_v является порожним из некоторого стартового депо $d_v^s \in D$, а последний рейс π_{v,\hat{k}_v} – также порожним в некоторое финишное депо $d_v^f \in D$.

Каждый рейс $\pi_{v,k}$ последовательности π_v для электробусов $v \in V$, $k = 1, \dots, \bar{k}_v - 1$ предшествует терминалу $\theta_{v,k}$ с опцией (возможно, пустой) зарядки $(v, k) \in K_v = \{(v, k) | k = 1, \dots, \bar{k}_v - 1\}$. Обозначим $tt(k) \in I \cup DH$ тип рейса $\pi_{v,k}$ в последовательности π_v . Если $tt(k) \in I$, то рейс $\pi_{v,k}$ является сервисным $j_{v,l(k)}$ с номером $l(k)$ в последовательности I_v , предшествующим терминалу $\theta_{v,k} = b_{j_{v,l(k)}}$. Если $tt(k) \in DH$, то $\pi_{v,k}$ является порожним, связывающим конечный терминал $b_{j_{v,l(k-1)}}$ предшествующего сервисного рейса $\pi_{v,k-1} = j_{v,l(k-1)}$ с номером $l(k-1)$ в последовательности I_v со стартовым терминалом $a_{j_{v,l(k-1)+1}}$ очередного сервисного рейса $\pi_{v,k+1} = j_{v,l(k-1)+1}$. Порожнему рейсу $\pi_{v,k}$ соответствуют его длительность $\gamma_{j_{v,l(k-1)}, j_{v,l(k-1)+1}}$ и величина $g_{j_{v,l(k-1)}, j_{v,l(k-1)+1}}$ заряда для его выполнения.

В промежутках между соседними сервисными рейсами $j_{v,l}, j_{v,l+1}$, $l = 1, 2, \dots, \bar{l}_v - 1$, выполняемыми электробусом $v \in V$, а также перед первым и после последнего сервисных рейсов может выполняться зарядка его батареи на зарядных станциях терминалов в диапазоне опции зарядки. Диапазоны $[\hat{A}_{v,k}, \hat{B}_{v,k}]$ опций зарядки (v, k) в терминале $\theta_{v,k}$, $k = 1, \dots, \bar{k}_v - 1$, $\theta_{v,k} \in \Theta$, определяются в зависимости от типов рейсов $k, k+1$ последовательности π_v :

- 1) если $tt(k), tt(k+1) \in I$ (порожний рейс отсутствует), то $[\hat{A}_{v,k}, \hat{B}_{v,k}] = [\beta_{j_{v,l(k)}}, \alpha_{j_{v,l(k)+1}}]$;
- 2) если $tt(k) \in I$, $tt(k+1) \in DH$, то $[\hat{A}_{v,k}, \hat{B}_{v,k}] = [\beta_{j_{v,l(k)}}, \alpha_{j_{v,l(k)+1}} - \gamma_{j_{l(k)}, j_{l(k)+1}}]$, $[\hat{A}_{v,k+1}, \hat{B}_{v,k+1}] = [\beta_{j_{v,l(k)}} + \gamma_{j_{l(k)}, j_{l(k)+1}}, \alpha_{j_{v,l(k)+1}}]$;
- 3) $[\hat{A}_{v,1}, \hat{B}_{v,1}] = [\alpha_{j_{v,1}} - \tau_{\max}, \alpha_{j_{v,1}}]$, $[\hat{A}_{v,\bar{k}-1}, \hat{B}_{v,\bar{k}-1}] = [\beta_{j_{v,\bar{k}}}, \beta_{j_{v,\bar{k}}} + \tau_{\max}]$.

Управляемые переменные:

$S_{c,v,k} \geq 0$ – момент начала события зарядки на зарядной станции $c \in C_{\theta_{v,k}}$ для $(v, k) \in K_v$, $v \in V$;

$F_{c,v,k} \geq 0$ – момент завершения события зарядки на зарядной станции $c \in C_{\theta_{v,k}}$, $v \in V$ для $(v, k) \in K_v$, $v \in V$;

$t_{c,v,k} \geq 0$ – длительность события зарядки на зарядной станции $c \in C_{\theta_{v,k}}$ для $(v, w) \in K_v$, $v \in V$;

$y_{c,v,k} \in \{0,1\}$ – булева переменная, принимающая значение 1 тогда и только тогда, когда $t_{c,v,k} > 0$, $c \in C_{\theta_{v,w}}$ для $(v, k) \in K_v$, $v \in V$;

$u_{v,k}$ – уровень заряда электробуса v перед k -й опцией (v,k) зарядки для π_v , $k=1, \dots, \bar{k}_v$ (u_{v,\bar{k}_v} – при приходе в депо d_v^f), $v \in V$;

$\bar{u}_{v,k}$ – уровень заряда электробуса v после k -й опции (v,w) зарядки для π_v , $k=1, \dots, \bar{k}_v-1$;

$x_{v,d,j_{v,1}} \in \{0,1\}$ – булева переменная, принимающая значение 1, если электробус $v \in V$ выполняет первый порожний рейс своего дневного задания из стартового депо $d \in D$ в начальный терминал $a_{j_{v,1}}$ первого сервисного рейса $j_{v,1}$;

$x_{v,j_{v,\bar{j}_v},d} \in \{0,1\}$ – булева переменная, принимающая значение 1, если электробус $v \in V$ выполняет последний порожний рейс своего дневного задания из конечного терминала $b_{j_{v,\bar{j}_v}}$ последнего сервисного рейса j_{v,\bar{j}_v} в депо $d \in D$;

$\sigma_{\theta,c} \in \{0,1\}$ – булева переменная, принимающая значение 1, если $\sum_{(v,k) \in Q_\theta} y_{c,v,k} > 0$ для $c \in C_\theta$, иначе 0;

μ_θ – целое число зарядных станций терминала θ , $\theta \in \Theta$.

Обозначим через $\mu(z)$, $t(z)$, $y(z)$, $S(z)$, $F(z)$, $u(z)$, $x(z)$ векторы значений переменных μ , t , y , S , F , u , x при фиксированном наборе z подмножеств сервисных рейсов, $z \in Z$.

Подзадача нижнего уровня заключается в отыскании векторов $\mu(z)$ количеств зарядных станций в терминалах $\theta \in \Theta$, длительностей $t(z)$ зарядки всех электробусов v для всех опций (v, k) зарядки, $k \in \{1,2, \dots, \bar{k}_v-1\}$, $v \in V$ в соответствующих терминалах $\theta_{v,k}$, моментов $S(z)$ старта и моментов $F(z)$ завершения этих событий зарядки, уровней $u(z)$ заряда электробуса $v \in V$ перед и после каждой опции зарядки, а также стартового d_v^s и финишного d_v^f депо для каждого $v \in V$, минимизирующих целевую функцию $\Phi(\mu(z), t(z), y(z), S(z), F(z), u(z), x(z))$ при выполнении системы ограничений. Описание целевой функции и ограничений приведено ниже.

Целевая функция подзадачи нижнего уровня. Целевая функция (1) представляет общие суточные затраты на зарядные станции и потраченную электроэнергию при фиксированном z . Суточные инвестиционные и операционные затраты на зарядные станции определяются суммой произведений суточных затрат c^{ch} на одну станцию на количество $\mu_\theta(z)$ зарядных станций каждого из терминалов $\theta \in \Theta$. Суточные затраты на электроэнергию электробусов парка выражаются произведением дневного тарифа c^{el} на величину $\sum_{\theta \in \Theta} \sum_{(v,k) \in Q_{\theta,v,k}(z)} \sum_{c \in C_{\theta,v,k}} P \cdot t_{c,v,k}(z)$ электроэнергии, потребленной электробусами $v \in V(z)$ для опций событий зарядки из множеств $Q_\theta(z)$ на зарядных станциях $c \in C_\theta$ терминалов $\theta \in \Theta$:

$$\min \Phi(\mu(z), t(z), y(z), S(z), F(z), u(z), x(z)) = \sum_{\theta \in \Theta} c^{ch} \mu_\theta(z) + c^{el} \sum_{\theta \in \Theta} \sum_{(v,k) \in Q_{\theta,v,k}(z)} \sum_{c \in C_{\theta,v,k}} P \cdot t_{c,v,k}(z). \quad (1)$$

Ограничения (параметр z для краткости опущен). Условия для инфраструктуры зарядных станций терминала определяются формулами

$$\mu_0 = \sum_{c \in C_0} \sigma_{\theta, c}, \theta \in \Theta, \quad (2)$$

$$\sigma_{\theta, c} \leq \sum_{(v, k) \in Q_0} y_{c, v, k}, c \in C_0, \theta \in \Theta, \quad (3)$$

$$0 \leq |Q_0| \cdot \sigma_{\theta, c} - \sum_{(v, k) \in Q_0} y_{c, v, k}, c \in C_0, \theta \in \Theta, \quad (4)$$

$$\sigma_{\theta, c+1} \leq \sigma_{\theta, c}, c, c+1 \in C_0, \theta \in \Theta, \quad (5)$$

$$\mu_0 \leq |C_0|, \theta \in \Theta. \quad (6)$$

Условия (2)–(6) обеспечивают равенство значения параметра μ_0 числу используемых зарядных станций терминала θ , не превосходящему предельно допустимого $|C_0|$, и исключают наличие неиспользуемых зарядных станций. Ограничения (3), (4) обеспечивают значения булевой переменной $\sigma_{\theta, c}$ равными 1, только если зарядная станция с номером c используется для выполнения опций зарядки электробуса. Из ограничения (5) следует, что исключается учет неиспользуемых зарядных станций. Согласно ограничению (6) число используемых зарядных станций не может превышать предельно допустимого числа $|C_0|$.

Условия баланса уровней заряда определяются следующим образом:

$$\underline{u}_{v, k+1} = \bar{u}_{v, k} - g_{j_{v, l(k+1)}}, tt(k+1) \in I, k \in \{1, 2, \dots, \bar{k}_v - 2\}, v \in V, \quad (7)$$

$$\underline{u}_{v, k+1} = \bar{u}_{v, k} - g_{j_{v, l(k)}, j_{v, l(k)+1}}, tt(k+1) \in DH, k \in \{2, \dots, \bar{k}_v - 3\}, v \in V, \quad (8)$$

$$\underline{u}_{v, 1} = \bar{u}_{v, 0} - \sum_{d \in D} x_{v, d, j_{v, 1}} g_{d, j_{v, 1}}, tt(1) \in DH, v \in V, \quad (9)$$

$$\underline{u}_{v, \bar{k}_v} = \bar{u}_{v, \bar{k}_v - 1} - \sum_{d \in D} x_{v, j_{v, \bar{k}_v}, d} g_{j_{v, \bar{k}_v}, d}, tt(\bar{k}_v) \in DH, v \in V, \quad (10)$$

$$\bar{u}_{v, 0} = u_{max}, v \in V, \quad (11)$$

$$\bar{u}_{v, k} = \underline{u}_{v, k} + \sum_{c \in C_{0_{v, k}}} f \cdot t_{c, v, k}, tt(k) \in I \cup DH, k \in \{1, 2, \dots, \bar{k}_v - 1\}, v \in V. \quad (12)$$

Условия (7)–(12) обеспечивают баланс значений уровня заряда электробуса при приходе в терминал (либо финишное депо) и соответствующих значений при выходе из предыдущего терминала (либо из начального депо) с учетом потерь заряда в сервисных либо порожних рейсах и восстановления заряда при зарядках на следующих за рейсами терминалах. При этом уравнения (7) определяют уровень заряда при приходе в конечный терминал любого сервисного рейса как разность между уровнем заряда при выходе из начального терминала и затратами заряда на выполнение рейса, а уравнения (8) – при приходе в начальный терминал очередного сервисного рейса после текущего порожнего рейса за исключением начального ($k = 1$) и последнего ($k = \bar{k}_v$) порожних рейсов. Условия (9) устанавливают уровень заряда при приходе в начальный терминал первого сервисного рейса после выполнения первого порожнего рейса из стартового депо, а условие (10) – уровень заряда при приходе в финишное депо из конечного терминала последнего сервисного рейса. Условия (11) обеспечивают максимальный уровень заряда электробуса при выходе из стартового депо в начале выполнения его дневного задания. Условия (12) устанавливают уровень заряда при выходе из любого терминала равным уровню заряда при приходе в этот терминал, увеличенному на величину заряда, полученного при зарядке на этом терминале.

Ограничения допустимого диапазона уровней заряда:

$$\bar{u}_{v,k} \leq u_{max}, k \in \{1, 2, \dots, \bar{k}_v - 1\}, v \in V, \quad (13)$$

$$u_{min} \leq \underline{u}_{v,k}, k \in \{1, 2, \dots, \bar{k}_v\}, v \in V. \quad (14)$$

Ограничения (13) препятствуют превышению максимально допустимого значения уровня заряда электробуса после любой опции зарядки. Ограничения (14) препятствуют снижению уровня относительного заряда ниже минимально допустимой величины после выполнения любого рейса дневного задания.

Условия для моментов начала, завершения и длительностей опции зарядки определяются формулами

$$t_{Cmin} y_{c,v,k} \leq t_{c,v,k} \leq y_{c,v,k} \cdot \tau_{max}, k \in \{1, 2, \dots, \bar{k}_v - 1\}, c \in C_{\theta_{v,k}}, v \in V, \quad (15)$$

$$\hat{A}_{v,k} y_{c,v,k} \leq S_{c,v,k}, tt(k) \in I \cup DH, k \in \{1, 2, \dots, \bar{k}_v - 1\}, c \in C_{\theta_{v,k}}, v \in V, \quad (16)$$

$$\sum_{c' \in C_{\theta_{v,k}}} F_{c',v,k} + \gamma_{j_{l(k)}, j_{l(k)+1}} \leq \sum_{c'' \in C_{\theta_{v,k+1}}} S_{c'',v,k+1} + 24(1 - \sum_{c'' \in C_{\theta_{v,k+1}}} y_{c'',v,k+1}), tt(k+1) \in DH, k \in \{2, \dots, \bar{k}_v - 3\}, v \in V, \quad (17)$$

$$F_{c,v,k} = S_{c,v,k} + t_{c,v,k} + \Delta y_{c,v,k}, tt(k) \in I \cup DH, k \in \{1, 2, \dots, \bar{k}_v - 1\}, c \in C_{\theta_{v,k}}, v \in V, \quad (18)$$

$$F_{c,v,k} \leq \hat{B}_{v,k} y_{c,v,k}, tt(k) \in I \cup DH, k \in \{1, 2, \dots, \bar{k}_v - 1\}, c \in C_{\theta_{v,k}}, v \in V, \quad (19)$$

$$(F_{c,v,k} \leq S_{c,v',k'}) \vee (F_{c,v',k'} \leq S_{c,v,k}), [\hat{A}_{v,k}, \hat{B}_{v,k}] \cap [\hat{A}_{v',k'}, \hat{B}_{v',k'}] \neq \emptyset, c \in C_{\theta}, (v, k), (v', k') \in Q_{\theta}, (v, k) \neq (v', k'), \theta \in \Theta, \quad (20)$$

$$\sum_{c \in C_{\theta}} y_{c,v,k} \leq 1, (v, k) \in Q_{\theta}, \theta \in \Theta. \quad (21)$$

Ограничения (15) обеспечивают значение длительности зарядки электробуса для любой опции зарядки равным либо 0, если оно меньше минимально допустимого, либо величине из интервала между минимально допустимой и максимальной длительностями зарядки. Ограничения (16) обеспечивают начало зарядки электробуса на любом терминале после любого рейса не ранее момента прибытия в терминал. Ограничения (17) обеспечивают начало зарядки после порожнего рейса, следующего за сервисным рейсом, не ранее момента завершения зарядки на конечном терминале предыдущего сервисного рейса, увеличенного на величину длительности порожнего рейса. Таким образом, ограничения (16), (17) гарантируют допустимость начала зарядки после порожнего рейса и препятствуют пересечению событий зарядки одного электробуса на двух зарядных станциях различных терминалов. Условия (18) устанавливают равенство момента завершения события зарядки после любого рейса моменту начала зарядки, увеличенному на суммарную длительность зарядки и операции настройки. Ограничения (19) обеспечивают завершение зарядки не позднее верхней границы допустимого интервала опции зарядки. При отсутствии зарядки после сервисного рейса значения моментов начала и окончания зарядки равны 0. При отсутствии зарядки после порожнего рейса значения моментов начала и окончания зарядки совпадают и превышают момент завершения зарядки предыдущей опции. Ограничения (20) обеспечивают отсутствие пересечения по времени различных опций зарядки различных электробусов на одной зарядной станции терминала для пересекающихся допустимых временных интервалов для этих опций. Ограничения (21) устанавливают, что любая опция зарядки может выполняться не более чем на одной станции зарядки.

Балансовые условия для количеств электробусов в различных депо:

$$\sum_{d \in D} x_{v,j_v, \bar{j}_v, d} = 1, v \in V, \quad (22)$$

$$\sum_{d \in D} x_{v,d,j_{v,1}} = 1, v \in V, \quad (23)$$

$$\sum_{v \in V} x_{v,d,j_{v,1}} - \sum_{v \in V} x_{v,j_{v,\bar{l}_v},d} = 0, d \in D. \quad (24)$$

Уравнения (22)–(24) устанавливают распределение электробусов по депо в начале дня перед выполнением ими дневных заданий и после возвращения в депо в конце дня. Уравнения (22) обеспечивают возвращение каждого электробуса в одно депо, уравнения (23) – старт каждого электробуса из одного депо в начале дня, уравнения (24) – баланс количества вышедших утром и количества вернувшихся после выполнения задания для каждого депо. При этом стартовое и финишное депо одного и того же электробуса могут не совпадать.

Ограничения (25)–(36) задают область значений переменных задачи:

$$0 \leq t_{c,v,k}, c \in C_{\theta}, (v, k) \in Q_{\theta}, \theta \in \Theta, v \in V, \quad (25)$$

$$S_{c,v,k} \in [0, \hat{B}_{v,k} - \Delta - tc_{\min}], c \in C_{\theta_{v,k}}, tt(k) \in I, k \in \{2, \dots, \bar{k}_v - 1\}, v \in V, \quad (26)$$

$$S_{c,v,k+1} \in [0, \hat{B}_{v,k+1} - \gamma_{j_{l(k)}, j_{l(k)+1}} - \Delta - tc_{\min}], c \in C_{\theta_{v,k+1}}, tt(k+1) \in DH, k \in \{3, \dots, \bar{k}_v - 3\}, v \in V, \quad (27)$$

$$F_{c,v,k} \in [0, \hat{B}_{v,k}], c \in C_{\theta_{v,k}}, tt(k) \in I, k \in \{2, \dots, \bar{k}_v - 1\}, v \in V, \quad (28)$$

$$F_{c,v,k+1} \in [0, \hat{B}_{v,k+1}], c \in C_{\theta_{v,k+1}}, tt(k+1) \in DH, k \in \{3, \dots, \bar{k}_v - 3\}, v \in V, \quad (29)$$

$$F_{c,v,1} \in [0, \alpha_{j_{v,1}}], c \in C_{\theta_{v,1}}, tt(1) \in DH, v \in V, \quad (30)$$

$$F_{c,v,\bar{k}_v-1} \in [0, \beta_{j_{v,\bar{l}_v}} + \tau_{\max}], c \in C_{\theta_{v,\bar{l}_v}}, tt(\bar{k}_v - 1) \in I, v \in V, \quad (31)$$

$$y_{c,v,k} \in \{0, 1\}, c \in C_{\theta}, (v, k) \in Q_{\theta}, \theta \in \Theta, v \in V, \quad (32)$$

$$x_{v,d,i} \in \{0, 1\}, i \in I_v, d \in D, v \in V, \quad (33)$$

$$x_{v,j,d} \in \{0, 1\}, j \in I_v, d \in D, v \in V, \quad (34)$$

$$\sigma_{\theta,c} \in \{0, 1\}, c \in C_{\theta}, \theta \in \Theta, \quad (35)$$

$$\mu_0 - \text{целое число}, \theta \in \Theta. \quad (36)$$

Обозначим $\Phi^*(z)$ минимальное значение целевой функции подзадачи нижнего уровня при фиксированном наборе $z \in Z$. Положим $\Phi^*(z)=\infty$, если подзадача нижнего уровня для z неразрешима.

3.2. Подзадача верхнего уровня. Эвристический алгоритм выбора набора подмножеств сервисных рейсов. Общие суточные инвестиционные и операционные затраты на электробусы, зарядную инфраструктуру и электроэнергию при фиксированном z имеют вид $\Omega(z) = c^{bus}\eta(z) + \Phi^*(z)$, где $\Phi^*(z)$ – минимальное значение целевой функции подзадачи нижнего уровня при фиксированном z . Задача верхнего уровня заключается в выборе набора $z^* \in Z$, минимизирующего суточные затраты $\Omega(z)$.

Предлагаются следующие варианты процедуры выбора подмножества сервисных рейсов для электробуса. Посредством каждой из описанных ниже процедур для каждого электробуса $v \in V$ в порядке возрастания индекса v формируется максимальное по включению упорядоченное подмножество $I_v = \{j_{v,1}, j_{v,2}, \dots, j_{v,\bar{l}_v}\}$ совместимых сервисных рейсов $j_{v,l} \in I$, $\beta_{j_{v,l}} + \gamma_{j_{v,l}, j_{v,l+1}} \leq \alpha_{j_{v,l+1}}$, $l = 1, 2, \dots, \bar{l}_v - 1$, назначаемых этому электробусу.

Каждый из вариантов 1–4 описываемой ниже процедуры отличается способом выбора для рассматриваемого электробуса очередного сервисного рейса. Процедура 1 ориентирована на минимизацию суммарной длительности интервалов между сервисными рейсами электробуса

с целью увеличения количества выполняемых им сервисных рейсов. Процедура 2 минимизирует взвешенную комбинацию дневных затрат на электробус и потраченную им в порожних рейсах электроэнергию. Процедуры 3 и 4 являются рандомизированными аналогами в заданной окрестности минимального значения для процедур 1 и 2 соответственно.

Используемые обозначения:

I^c – текущее подмножество нераспределенных сервисных рейсов;

i^c – текущий сервисный рейс, распределяемый электробусу;

I_v^c – текущее подмножество I_v ;

R_θ^- – доступный ресурс заряда в терминале $\theta \in \{b_i, a_j | i, j \in I^c\}$ перед возможной зарядкой в θ ;

R_θ^+ – доступный ресурс заряда в терминале $\theta \in \{b_i, a_j | i, j \in I^c\}$ после возможной зарядки в θ ;

$D_{i^c}^-$ – подмножество депо из D . Для каждого $d \in D_{i^c}^-$ заряд g_{d,i^c} , потраченный в порожнем рейсе из депо d в начальный терминал a_{i^c} , не может быть восстановлен на этом терминале вследствие условия $g_{d,i^c} < f \cdot tc_{min}$;

E_θ – доступная величина заряда при возможной зарядке в θ , $E_\theta \geq 0$;

g_i – заряд, требуемый для выполнения сервисного рейса i ;

γ_{ij} – длительность порожнего рейса из терминала b_i после рейса i в терминал a_j рейса j ;

g_{ij} – заряд, требуемый для выполнения порожнего рейса из терминала b_i после рейса i в терминал a_j рейса j ;

$H(x) = 0$ при $x < 0$ и 1 при $x \geq 0$ – ступенчатая функция Хэвисайда;

$\text{sgn}(x) = -1$ при $x < 0$, 0 при $x = 0$ и 1 при $x \geq 0$.

Процедура 1. Выбор очередного сервисного рейса осуществляется по принципу минимизации допустимого момента $\alpha^* = \min\{a_j | j \in I^c\}$ времени начала этого рейса. При наличии нескольких сервисных рейсов с минимальным значением момента начала выбор среди них осуществляется равновероятно.

Процедура 2. Выбор очередного сервисного рейса осуществляется по принципу минимизации величины $\delta_{ij}^* = \min\{a_j + \lambda\gamma_{ij} | j \in I^c\}$ (суммы момента времени a_j начала следующего сервисного рейса j и длительности γ_{ij} порожнего рейса с весовым коэффициентом λ , соединяющего его с предыдущим сервисным рейсом i).

Процедура 3. Выбор очередного сервисного рейса осуществляется равновероятно из окрестности $[\alpha^*, \alpha^* + \varepsilon]$ в множестве I^c минимального значения α^* момента начала сервисных рейсов, где $\varepsilon > 0$ – заданный параметр.

Процедура 4. Выбор очередного сервисного рейса осуществляется равновероятно из окрестности $[\delta_{ij}^*, \delta_{ij}^* + \varepsilon]$ величины δ_{ij}^* в множестве I^c , где $\varepsilon > 0$ – заданный параметр.

Задано текущее множество нераспределенных сервисных рейсов $I^c \neq \emptyset$ ($I^c = I$ для $v = v_1$).

Эвристический алгоритм определения подмножества I_v сервисных рейсов для $v \in V$:

Шаг 0. Полагаем $I_v^c = \emptyset$.

Шаг 1. Выбор первого сервисного рейса для электробуса v .

Если $I^c = \emptyset$, то конец (I_v^c – решение). Иначе, если $I^c \neq \emptyset$, выбираем первый сервисный рейс $i^* = j_{v,1}$ согласно одной из процедур 1–4. Переход к шагу 2.

Шаг 2. Проверка допустимости первого сервисного рейса.

Если $R_{a_{i^*}}^+ = u_{\max} - u_{\min} - \max_{d \in D_{i^*}^-} g_{d,i^*} - g_{i^*} \geq 0$, то полагаем $i^c = i^*$, $I_v^c = \{i^c\}$, $R_{a_{i^c}}^+ = u_{\max} - u_{\min} - \max_{d \in D_{i^c}^-} g_{d,i^c}$.

Определяем $R_{b_{i^c}}^- = R_{a_{i^c}}^+ - g_{i^c} \geq 0$. Полагаем $I^c = I^c \setminus \{i^c\}$. Переход к шагу 3.

Иначе, если $R_{a_{i^*}}^+ = u_{\max} - u_{\min} - \max_{d \in D_{i^*}^-} g_{d,i^*} - g_{i^*} < 0$, то i^* – недопустимый рейс. Полагаем $I^c = I^c \setminus \{i^*\}$.

Переход к шагу 1.

Шаг 3. Выбор очередного сервисного рейса $j_v, l \in I^c, l = 2, \dots, \bar{l}_v$ для текущего v .

Если $I^c = \emptyset$, то конец (I_v^c – решение). Иначе для i^c ищем очередной сервисный рейс $j^* \in I^c$ согласно одной из процедур 1–4 при выполнении условия допустимости

$$R_{b_{i^c}}^- + H(\alpha_j - \beta_{i^c} - \gamma_{i^c, j} - \Delta - tc_{min}) \cdot f \cdot (\alpha_j - \beta_{i^c} - \gamma_{i^c, j} - \Delta) \geq g_{i^c, j} + g_j. \quad (37)$$

Шаг 3.1. Если j^* найден, то полагаем $E_{b_{i^c}} = \min \{u_{\max} - u_{\min} - R_{b_{i^c}}^-, H(\alpha_{j^*} - \beta_{i^c} - \gamma_{i^c, j^*} - \Delta - tc_{min}) \cdot f \cdot (\alpha_{j^*} - \beta_{i^c} - \gamma_{i^c, j^*} - \Delta)\}$, $R_{b_{i^c}}^+ = R_{b_{i^c}}^- + E_{b_{i^c}}$.

Определяем $R_{a_{j^*}}^- = R_{b_{i^c}}^+ - g_{i^c, j^*}$, $E_{a_{j^*}} = \min \{u_{\max} - u_{\min} - R_{a_{j^*}}^-, H(\alpha_{j^*} - \beta_{i^c} - \gamma_{i^c, j^*} - \frac{E_{b_{i^c}}}{f} - \Delta \cdot \text{sgn}(E_{b_{i^c}}) - \Delta - tc_{min}) \cdot f \cdot (\alpha_{j^*} - \beta_{i^c} - \gamma_{i^c, j^*} - \frac{E_{b_{i^c}}}{f} - \Delta \cdot \text{sgn}(E_{b_{i^c}}) - \Delta) - \Delta\}$, $R_{a_{j^*}}^+ = R_{a_{j^*}}^- + E_{a_{j^*}}$, $R_{b_{j^*}}^- = R_{a_{j^*}}^+ - g_{j^*}$.

Полагаем $I^c = I^c \setminus \{j^*\}$, $I_v^c = I_v^c \cup \{j^*\}$, $i^c = j^*$. Переход к шагу 3.

Шаг 3.2. Иначе (если j^* не найден) полагаем $I^c = I^c \setminus \{j^*\}$. Переход к шагу 3.

Замечание. Если в результате выполнения алгоритма $I_v^c = \emptyset$, то электробус v не используется. В этом случае $I_o^c = \emptyset$ для всех $v \leq o \leq \eta(z)$.

Если при завершении работы алгоритма $I^c \neq \emptyset$, то допустимое решение задачи не найдено.

4. Результаты численных экспериментов для реального примера г. Минска.

4.1. Реальный пример г. Минска. Рассматривается набор из 12 маршрутных линий г. Минска перевозки пассажиров в прямом и обратном направлениях, начинающихся и заканчивающихся в одном из девяти терминалов: ДС Дружная (Др), ДС Кунцевщина (Кунц), ДС Серова (Сер), ДС Курасовщина (Кур), ж.-д. станция Минск-Южный (М.-Ю.), ДС Малиновка-4 (Мал-4), ДС Масюковщина (Масюк), ж.-д. станция Брилевичи (Брил), Колядичи (Коляд). Используются также два депо: Транспортный парк 2 (ТП2) и Автобусный парк 5 (АП5) для размещения электробусов и их подзарядки в ночное время (Справочник маршрутов Минска, URL: <https://wikiroutes.info/minsk/catalog>).

Парк электробусов состоит из идентичных электробусов Е420 производства ОАО «Белкоммунмаш», Минск (БКМ Холдинг, URL: <https://holdingbkm.com>), оборудованных аккумуляторной батареей емкостью $E = 48 \text{ кВт}\cdot\text{ч}$. Допустимый диапазон значений уровня заряда составляет $[0,3, 1]$. Терминалы оборудуются идентичными зарядными станциями быстрой зарядки СЗА-1 номинальной мощностью $P = 288 \text{ кВт}$. Кривая функции уровня заряда от времени зарядки аппроксимируется в допустимом диапазоне линейной функцией изменения заряда с постоянной скоростью $f = 6 \text{ е.з./ч}$ ($0,1 \text{ е.з./мин}$), из чего следует, что для зарядки от уровня заряда 0,3 до 1 требуется $\tau^{\max} = 0,1167 \text{ ч}$ (7 мин). Время Δ выполнения настройки электробуса на зарядку, как и минимальная длительность tc_{min} зарядки, составляет 0,0167 ч (1 мин). Удельный расход электроэнергии электробусов парка в сервисных рейсах составляет 1,6 $\text{кВт}\cdot\text{ч}/\text{км}$ (или в единицах относительного заряда 0,0333 е.з./км), в порожних рейсах 1,2 $\text{кВт}\cdot\text{ч}/\text{км}$ (или 0,025 е.з./км) соответственно. Средняя скорость движения электробуса в сервисном рейсе меняется в диапазоне 16–25 км/ч в зависимости от характеристик маршрута, средняя скорость движения электробуса в порожнем рейсе составляет 32 км/ч. Годовая стоимость СЗА-1 составляет 37 060 евро (далее все стоимости и значения целевой функции измеряются в евро). Общие капитальные и операционные годовые затраты на электробусы Е420 [5] составляют 53 000. Дневной тариф на электроэнергию составляет 0,195 (евро/кВт·ч).

Дневной период работы транспорта г. Минска разбивается на две категории: часы пик (с 7:00 до 10:00 и с 16:00 до 19:00) и непиковые часы (с 5:00 до 7:00, с 10:00 до 16:00 и с 19:00 до 24:00), характеризующиеся различными частотами выхода электробуса на маршрут (частота для часов пик превышает более чем в два раза частоту для непиковых часов). Поправочный коэффициент для оценки потребленной электроэнергии электробусами парка в течение дня по

отношению к интервалу с 7:00 до 10:00 с учетом различия частот прихода электробусов в терминалы принят равным 4,17.

В табл. 1 приведены характеристики набора из 112 сервисных рейсов в представительном интервале «часы пик» с 7:00 до 10:00 (государственное предприятие «Минсктранс», URL: <https://www.minsktrans.by>; Справочник маршрутов Минска). В табл. 2–4 приведены характеристики порожних рейсов между терминалами маршрутов, между депо и стартовыми терминалами сервисных рейсов, а также между конечными терминалами сервисных рейсов и депо, включая расстояние (км), длительность (мин) и количество единиц относительного заряда (е.з.), необходимых для выполнения рейсов.

Таблица 1
Характеристики сервисных рейсов, обслуживаемых электробусами

Table 1
Characteristics of service trips, served by electric buses

Терминал 1—терминал 2 Term.1—Term.2	Расстояние, км Distance, km	Длительность, мин Duration, min	Расход заряда, е.з. Charge consumption, c.u.	Период, ч:мин Period, h:min	Интервал, мин Interval, min
Др—М.-Ю.	4,7	18	0,16	7:00–10:00	30
М.-Ю. —Др	5,1	22	0,17	7:30–10:30	30
Др—Мал-4	13	38	0,43	7:00–10:00	20
Мал-4—Др	13	42	0,43	7:20–10:20	20
Др—Брил	8,5	30	0,27	7:00–9:55	25
Брил—Др	8,2	29	0,28	7:00–9:55	25
Кунц—Кур	15,18	32	0,51	7:00–10:00	15
Кур—Кунц	15,52	33	0,52	7:00–10:00	15
Кур—Коляд	8,7	21	0,29	7:00–10:00	20
Коляд—Кур	9,67	24	0,32	7:00–10:00	20
Сер—Масюк	20,53	69	0,68	7:00–9:55	25
Масюк—Сер	20,81	70	0,69	7:00–9:55	25

Таблица 2
Характеристики порожних рейсов между терминалами

Table 2
Characteristics of deadhead trips between terminals

Терм. Term.	Др	Кунц	Сер	Кур	М.-Ю.	Мал-4	Масюк	Брил	Коляд
Др	—	10,4/19/0,26	5,5/10/0,14	7,7/14/0,19	4,7/9/0,12	9,3/17/0,23	11,6/21/0,29	8,5/16/0,21	11,4/21/0,29
Кунц	9,8/18/0,25	—	14,4/27/0,36	16,4/31/0,41	14/26/0,35	8,4/16/0,21	5,1/10/0,13	11,7/22/0,29	20,6/39/0,52
Сер	5,4/10/0,14	14,2/27/0,36	—	5,5/10/0,14	4/8/0,1	13/24/0,32	12,7/24/0,32	9,7/18/0,24	11,1/21/0,28
Кур	7,8/14/0,2	16,1/30/0,4	3,2/7/0,08	—	6,7/12/0,17	8,6/16/0,22	12,7/24/0,32	5,2/10/0,13	8,7/16/0,22
М.-Ю.	5,1/9/0,13	14,5/27/0,36	5,7/10/0,14	8,3/16/0,21	—	13,1/25/0,33	12,9/24/0,32	12,4/23/0,31	8/15/0,2
Мал-4	10,6/20/0,26	14,4/27/0,36	13,6/25/0,34	8,8/16/0,22	12,5/23/0,31	—	11,8/22/0,3	5,8/11/0,15	15,7/29/0,39
Масюк	8,1/15/0,2	4/8/0,1	12,3/23/0,3	12,4/23/0,31	11,2/210,28	11,4/21/0,29	—	13,2/25/0,33	19,3/360,48
Брил	8,2/15/0,2	11,8/22/0,3	7/13/0,18	4/8/0,1	10,7/20/0,27	4,9/9/0,12	11,5/21/0,29	—	10,9/20/0,27
Коляд	11,5/22/0,29	14,2/27/0,36	5,3/10/0,13	2,5/4,7/0,06	9,1/17/0,23	7,7/14/0,19	18,8/35/0,47	11,7/22/0,29	—

Таблица 3
Характеристики порожних рейсов между депо и терминалами

Table 3
Characteristics of deadhead trips between depots and terminals

	Др	Кунц	Сер	Кур	М.-Ю.	Мал-4	Масюк	Брил	Коляд
ТП2	4,7/9/0,12	13,6/25/0,34	4,7/9/0,12	7,6/14/0,19	2,86/5/0,07	14,5/27/0,36	11,1/21/0,28	11,6/22/0,29	9,9/19/0,25
АП5	5,8/11/0,14	7,3/14/0,18	8,4/16/0,21	8,6/16/0,21	7,4/14/0,18	5,3/10/0,13	5,6/10/0,14	6,6/12/0,17	14,9/28/0,37

Таблица 4

Характеристики порожних рейсов из терминалов в депо

Table 4
Characteristics of deadhead trips from terminals to depots

	Др	Кунц	Сер	Кур	М.-Ю.	Мал-4	Масюк	Брил	Коляд
ПП2	4,8/9/0,12	14,4/27/0,36	4,5/9/0,12	7/13/0,18	3,2/8/0,08	12,8/24/0,32	11,4/21/0,28	11,2/21/0,28	9,4/17/0,24
ПП5	5,4/10/0,13	8,9/17/0,22	8,4/16/0,21	8,6/16/0,21	8,8/17/0,22	5,6/13/0,14	6,1/11/0,15	6,3/12/0,16	15,4/29/0,39

4.2. Результаты численных экспериментов. Для оценки работоспособности предложенной двухуровневой декомпозиционной схемы решения задачи выполнены две серии численных экспериментов для описанного выше примера г. Минска. В первой серии для решения подзадачи верхнего уровня использовались эвристические рандомизированные процедуры 1–4 (для процедуры 2 весовой коэффициент $\lambda=0,25$) назначения электробусам сервисных рейсов при предельном числе итераций 10 000, максимальном числе итераций без улучшения целевой функции 1000 и максимальном времени решения задачи 600 с. Для решения подзадач нижнего уровня использовалась некоммерческая версия универсального решателя задач СЦЛП Gurobi 9.5.1. Эксперименты выполнялись на стандартном ПК с процессором Intel Core i7 (1,8 ГГц) и 16 Гб оперативной памяти. В табл. 5 приведены результаты этих экспериментов.

В табл. 5 и 6 использованы следующие сокращения: *ObjOpt* – оптимальное значение целевой функции, *#Bus* – количество использованных электробусов, *#Char* – общее количество использованных зарядных станций на терминалах, *Char_Time* – общая длительность зарядки электробусов в представительном интервале, *Bus_Cost* – общая суточная стоимость электробусов, *Char_Cost* – общая суточная стоимость зарядных станций, *En_Cost* – общая суточная стоимость потраченной электроэнергии, *CPU* – время решения задачи (с), *#Iter* – общее число выполненных итераций, *It_opt* – номер итерации, на которой было получено решение.

Таблица 5

Результаты решения задачи с использованием эвристических процедур 1–4
назначения электробусов на сервисные рейсы

Table 5
Results of solving the problem using heuristic procedures 1–4 for assigning electric
buses to service trips

Характеристика Characteristic	Процедура 1 Procedure 1	Процедура 2 Procedure 2	Процедура 3 Procedure 3	Процедура 4 Procedure 4
<i>ObjOpt</i>	6863,4	6742,41	7087,36	7055,37
<i>#Bus</i>	31	31	34	33
<i>#Char</i>	8	8	6	7
<i>Char_Time</i>	6,6195	6,1028	6,58 258	6,63 258
<i>Bus_Cost</i>	4501,2	4501,2	4936,8	4791,6
<i>Char_Cost</i>	812	812	609	710,5
<i>En_Cost</i>	1550,2	1429,21	1541,56	1553,27
<i>CPU</i>	600,147	354,66	388,958	600,343
<i>#Iter</i>	2263	1580	1001	1391
<i>It_opt</i>	1557	580	1	1224

Во второй серии экспериментов при применении эвристических рандомизированных процедур 1–4 для решения подзадачи верхнего уровня исследовалась зависимость минимального количества использованных электробусов парка при ограничении на доступное суммарное количество зарядных станций на терминалах маршрутов. В каждом эксперименте предельное общее число итераций и число итераций без улучшения целевой функции составляло 3000, максимальное время решения экземпляра задачи – 600 с. Результаты этой серии экспериментов приведены в табл. 6, где символ – означает, что решение задачи не найдено.

Таблица 6

Зависимость минимального количества электробусов от количества доступных зарядных станций

Table 6

Dependence of the minimum number of electric buses on the number of available charging stations

Процедура 1 Procedure 1				Процедура 2 Procedure 2				Процедура 3 Procedure 3				Процедура 4 Procedure 4			
#Char	#Bus	CPU	It_opt	#Bus	CPU	It_opt	#Bus	CPU	It_opt	#Bus	CPU	It_opt	#Bus	CPU	It_opt
8	32	129,9	1709	32	139,2	63	34	158,5	1	34	133,2	1387			
7	33	116,0	461	32	124,3	142	34	148,7	1	34	200,7	737			
6	33	104,9	1752	33	109,5	1791	–	118,9	0	36	128,6	891			
5	–	100,3	–	–	110,6	–	–	–	–	–	–	101,4	–		

Результаты экспериментов показали, что эвристическая процедура 2 превосходит процедуры 1, 3, 4 как по значению целевой функции, так и по времени получения решения. Минимальное число электробусов, необходимых для выполнения заданного множества сервисных рейсов, возрастает при сокращении суммарного количества доступных зарядных станций на терминалах.

Заключение. Разработаны математическая модель и декомпозиционная двухуровневая схема решения задачи оптимизации состава однородного парка электробусов, назначения электробусов на сервисные рейсы из заданного множества, мест размещения и количеств размещаемых на терминалах маршрутов идентичных зарядных станций, а также расписания зарядки на них в течение представительного интервала дня. Проведенные вычислительные эксперименты на примере транспортной сети г. Минска показали хорошую работоспособность предложенной модели для задачи средней размерности и выявили предпочтительную по эффективности эвристическую рандомизированную процедуру выбора назначения электробусов на сервисные рейсы. В дальнейшем целесообразно провести более масштабный эксперимент по исследованию эффективности предложенного подхода к решению исследуемой задачи большей размерности и сравнению его с другими подходами к ее решению. Интерес представляет расширение предложенного подхода на нелинейные зависимости уровня заряда от времени зарядки, более точно отражающие реальный процесс, а также рассмотрение парка разнотипных электробусов и зарядной инфраструктуры на базе разнотипных зарядных станций различной мощности.

Вклад авторов. *М. Я. Ковалев* предложил постановку задачи выбора парка электробусов, назначения наборов сервисных рейсов электробусам парка, их зарядной инфраструктуры и расписаний зарядки. Им выполнена верификация предложенных математической модели, декомпозиционной схемы и методов решения подзадач, а также редактирование статьи. *Н. Н. Гуцинский* предложил схему декомпозиции рассматриваемой задачи и эвристический рандомизированный алгоритм решения подзадачи назначения наборов сервисных рейсов электробусам парка. Им уточнена модель смешанного целочисленного линейного программирования для подзадачи оптимизации зарядной инфраструктуры и расписаний зарядки электробусов, разработана программа решения задачи с использованием универсального решателя задач целочисленного линейного программирования и выполнены численные эксперименты. *Б. М. Розин* предложил модель смешанного целочисленного линейного программирования для подзадачи оптимизации зарядной инфраструктуры и расписаний зарядки электробусов при фиксированных наборах сервисных рейсов для электробусов парка. Им разработана эвристическая процедура назначения электробусам рейсов с минимальным временем старта, подготовлены исходные данные для примера электротранспортной сети г. Минска и составлен первоначальный текст статьи.

References

1. Dirks N., Schiffer M., Walther G. On the integration of battery electric buses into urban bus networks. *Transportation Research Part C: Emerging Technologies*, 2022, vol. 139, p. 103628. DOI: 10.1016/j.trc.2022.103628.
2. Gao Z., Lin Z., LaClair T. J., Liu C., Li J.-M., ..., Ward J. Battery capacity and recharging needs for electric buses in city transit service. *Energy*, 2017, vol. 122, pp. 588–600.

3. Olsson O., Grauers A., Pettersson S. Method to analyze cost effectiveness of different electric bus systems. *EVS29 International Battery, Hybrid and Fuel Cell Electric Vehicle Symposium, Montreal, Quebec, Canada, June 2016*, pp. 1–12.
4. Guschinsky N., Kovalyov M. Y., Rozin B., Brauner N. Fleet and charging infrastructure decisions for fast-charging city electric bus service. *Computers and Operations Research*, 2021, vol. 135, p. 105449. DOI: 10.1016/j.cor.2021.105449.
5. Battaia O., Dolgui A., Guschinsky N., Kovalev M. Y. Designing fast-charge urban electric bus services: An Integer Linear Programming model. *Transportation Research Part E: Logistics and Transportation Review*, 2023, vol. 171, p. 103065. DOI: 10.1016/j.tre.2023.103065.
6. Nath R. B., Rambha T., Schiffer M. On the impact of co-optimizing station locations, trip assignment, and charging schedules for electric buses. *Transportation Research Part C: Emerging Technologies*, 2024, vol. 167, p. 104839.
7. Quttineh N.-H., Häll C. H., Ekström E. J., Ceder A. Integrated solution for electric bus timetabling and vehicle scheduling combined with choices of charging locations. *Journal of Public Transportation*, 2023, vol. 25, p. 100055.
8. Alvo M., Angulo G., Klapp M. A. An exact solution approach for an electric bus dispatch problem. *Transportation Research Part E: Logistics and Transportation Review*, 2021, vol. 156, p. 102528.

Информация об авторах

Гуцинский Николай Николаевич, кандидат физико-математических наук, доцент, ведущий научный сотрудник, Объединенный институт проблем информатики Национальной академии наук Беларусь.
E-mail: gyshin@newman.bas-net.by
<http://orcid.org/0000-0003-1166-2387>

Ковалев Михаил Яковлевич, доктор физико-математических наук, профессор, главный научный сотрудник, Объединенный институт проблем информатики Национальной академии наук Беларусь.
E-mail: kovalyov_my@newman.bas-net.by
<http://orcid.org/0000-0003-0832-0829>

Розин Борис Матвеевич, кандидат технических наук, доцент, ведущий научный сотрудник, Объединенный институт проблем информатики Национальной академии наук Беларусь.
E-mail: rozin@newman.bas-net.by
<http://orcid.org/0000-0003-3814-6569>

Information about the authors

Nikolai N. Guschinsky, Ph. D. (Phys.-Math.), Assoc. Prof., Leading Researcher, The United Institute of Informatics Problems of the National Academy of Sciences of Belarus.
E-mail: gyshin@newman.bas-net.by
<http://orcid.org/0000-0003-1166-2387>

Mikhail Y. Kovalyov, D. Sc. (Phys.-Math.), Prof., Principal Researcher, The United Institute of Informatics Problems of the National Academy of Sciences of Belarus.
E-mail: kovalyov_my@newman.bas-net.by
<http://orcid.org/0000-0003-0832-0829>

Boris M. Rozin, Ph. D. (Eng.), Assoc. Prof., Leading Researcher, The United Institute of Informatics Problems of the National Academy of Sciences of Belarus.
E-mail: rozin@newman.bas-net.by
<http://orcid.org/0000-0003-3814-6569>



UDC: 330.43, 330.46, 519.237
DOI: 10.37661/1816-0301-2025-22-4-24-35

Original Article
Оригинальная статья

Latent factor analysis of regional growth in Belarus using structural equation modeling

Yuliya A. Osipova

*The Economy Research Institute of the Ministry of Economy of the Republic of Belarus,
st. Slavinskogo, 1/1, Minsk, 200086, Belarus
E-mail: osipova@niei.by*

Abstract

Objectives. The study provides a quantitative assessment of the impact of latent variables on regional economic activity and examines the interactions among various economic sectors and their contributions to regional growth. The analysis focuses on identifying latent factors of socio-economic development across Belarusian regions during the period 2016–2024, using factor analysis techniques and structural equation modeling (SEM).

Methods. To reduce data dimensionality and identify latent factors, exploratory factor analysis (EFA) was applied using the Principal Component Analysis (PCA) extraction method. A structural equation model was constructed using the SEMOPY library in Python to estimate relationships among the identified factors. To assess the quality of the model, standard fit indices were used: CFI – comparative fit index; TLI – Tucker-Lewis index; RMSEA – root mean square error of approximation. The values of these indices allow evaluating the degree of consistency between the model and the empirical data. The model is based on a system of 23 socio-economic indicators across 128 administrative districts and cities of regional subordination.

Results. The resulting SEM demonstrates high internal consistency and statistical reliability (CFI = 0,98; TLI = 0,97; RMSEA = 0,04), revealing significant causal linkages between latent factors. It was established that the financial sector is a key driver of investment activity, while growth in the housing stock directly stimulates consumer demand. Negative relationships were identified between agricultural potential and financial stability, as well as between industrial development and financial sustainability.

Conclusion. The developed model is an effective analytical tool for formulating evidence-based regional policy, optimizing resource allocation, and strategic planning. Promising directions for future research include incorporating time lags, adding indicators of innovation and human potential, and applying spatial econometrics methods.

Keywords: structural equation modeling, factor analysis, PCA, latent variables, Varimax, DWLS-estimation

For citation. Osipova Yu. A. *Latent factor analysis of regional growth in Belarus using structural equation modeling*. Informatika [Informatics], 2025, vol. 22, no. 4, pp. 24–35 (In Russ.). DOI: 10.37661/1816-0301-2025-22-4-24-35.

Conflict of interest. The author declares of no conflict of interest.

Received | Поступила в редакцию 09.09.2025

Accepted | Подписана в печать 23.10.2025

Published | Опубликована 30.12.2025

Анализ латентных факторов регионального роста в Беларуси с использованием структурного моделирования

Ю. А. Осипова

Научно-исследовательский экономический институт
Министерства экономики Республики Беларусь,
ул. Славинского, 1/1, Минск, 200086, Беларусь
E-mail: osipova@niei.by

Аннотация

Цели. Целью исследования являются количественная оценка влияния скрытых переменных на экономическую активность регионов Беларуси и анализ взаимодействия между различными секторами экономики путем выявления латентных факторов социально-экономического развития регионов за 2016–2024 гг. с использованием методов факторного анализа и моделирования структурными уравнениями (structural equation modeling, SEM).

Методы. Для сокращения размерности данных и выявления латентных факторов был применен метод объясняющего факторного анализа (EFA) с использованием процедуры извлечения факторов методом главных компонент (PCA). Для оценки связей между выявленными факторами была построена модель структурных уравнений с использованием библиотеки SEMOPY в Python. Для проверки качества модели использовались стандартные показатели соответствия: CFI – сравнительный индекс соответствия; TLI – индекс Текера-Льюиса; RMSEA – среднеквадратичная ошибка аппроксимации. Значения этих индексов позволяют оценить степень согласованности модели с эмпирическими данными. Модель основана на системе из 23 социально-экономических показателей по 128 административным районам и городам областного подчинения за девять лет.

Результаты. Построенная модель структурных уравнений демонстрирует высокую внутреннюю согласованность и статистическую надежность ($CFI = 0,98$; $TLI = 0,97$; $RMSEA = 0,045$) и выявляет значимые причинно-следственные связи между факторами. Финансовый сектор выступает ключевым драйвером инвестиционной активности, а рост жилищного фонда прямо усиливает потребительский спрос. Одновременно обнаружены отрицательные взаимосвязи между аграрным потенциалом и финансовой стабильностью, а также между промышленным развитием и финансовой устойчивостью.

Заключение. Представленная модель – это эффективный аналитический инструмент для разработки научно обоснованной региональной политики, оптимизации распределения ресурсов и стратегического планирования. Перспективные направления дальнейших исследований включают учет временных лагов, добавление показателей инновационного и человеческого потенциала, применение методов пространственной эконометрики.

Ключевые слова: моделирование структурными уравнениями, факторный анализ, PCA, латентные переменные, Varimax-ротация, DWLS-оценка

Для цитирования: Осипова, Ю. А. Анализ латентных факторов регионального роста в Беларуси с использованием структурного моделирования / Ю. А. Осипова // Информатика. – 2025. – Т. 22, № 4. – С. 24–35. – DOI: 10.37661/1816-0301-2025-22-4-24-35.

Конфликт интересов: Автор заявляет об отсутствии конфликта интересов.

Introduction. The socio-economic development of Belarusian regions is shaped by a complex interaction of factors, including agricultural potential, investment activity, and infrastructure provision. This paper serves as a continuation of prior research, in which the author applied machine learning techniques to regional datasets – specifically, factor and cluster analysis – to identify key indicators reflecting common development patterns across Belarus.

In the present study, structural equation modeling (SEM) is employed to quantitatively assess the influence of latent factors on regional socio-economic activity.

The purpose of this work is to uncover latent interdependencies among socio-economic indicators for the period 2016-2024 through factor analysis, and to identify their defining characteristics by constructing a structural equation model.

To examine the spatial dynamics of regional development, the author addresses the following research tasks:

1. Identify the patterns of socio-economic development based on indicators that reflect spatial-temporal dimensions.
2. Assess the quantitative impact of generalized latent patterns on regional development.
3. Construct a structural equation model that incorporates linear relationships between observed and latent variables.

The database was assembled using statistical indicators published by Belstat for 2016-2024. 23 indicators characterizing the socio-economic development of Belarusian administrative-territorial entities (ATEs) were selected for analysis, and their values were calculated for each district. Because some of these indicators reflect agricultural activity, the sample comprises 118 districts and 10 cities of regional subordination, with the agricultural-related indicators set to zero for those 10 cities. The full data matrix is not included in the article owing to its high dimensionality.

The research algorithm includes four stages.

I. Dataset Overview and Indicator Selection. A portion of the indicators was calculated per 1000 residents or per capita. Each region over time is represented as a vector of observations: $X_s(t) = (x_{s1}(t), x_{s2}(t), \dots, x_{sp}(t))$, where $s = 1, 2, \dots, m$, and in our case $m = 128$. Given the nature of the data, the temporal parameter is modeled as a discrete sequence of evenly spaced intervals, defined as $t = 1, 2, \dots, T$. In statistical yearbooks, the duration of one interval typically corresponds to one year. Accordingly, our analysis spans the period 2016-2024, with $T = 9^1$.

The resulting dataset includes 23 indicators reflecting socio-economic characteristics for each of ATEs, yielding $n = 128 \times 9 = 1152$ complete records across nine years.

II. Data Standardization. Standardization was performed using the `StandardScaler` method from the `scikit-learn` library². This technique adjusts feature values so that each has a mean of zero and a standard deviation of one. The purpose of this preprocessing is to enhance algorithmic convergence and prevent features with large magnitudes from dominating the model. In this study, scaling parameters were computed and applied directly to the raw values of each indicator.

III. Factor Analysis: Principal Components. To extract latent patterns, we employed exploratory factor analysis (EFA). The mathematical model is written as

$$x_i = \lambda_{i1}F_1 + \lambda_{i2}F_2 + \dots + \lambda_{im}F_m + \varepsilon_i,$$

where x_i is an observed variable, F_j is a latent factor, λ_{ij} is a factor loading, and ε_i is a unique error. The objective of exploratory factor analysis (EFA) is to estimate the loading matrix Λ and to determine the number of factors m that explain a significant portion of the total variance.

Factor extraction method: factors were extracted using principal component analysis (PCA) based on common variances, employed to reduce data dimensionality while preserving the maximum possible variance. The model for principal components is represented as a linear combination of the original variables:

$$z_j = a_{j1}x_1 + a_{j2}x_2 + \dots + a_{jp}x_p,$$

where z_j is the j -th principal component, a_{jk} are the coefficients (elements of the eigenvectors), and x_k are the original variables. The eigenvalues λ_j of the covariance matrix determine the share of variance explained by each component. According to Kaiser's criterion, components with $\lambda_j \geq 1$ are retained.

¹National Statistical Committee of the Republic of Belarus. Available at: <https://www.belstat.gov.by/> (accessed 03.09.2025).

²scikit-learn developers. Preprocessing data – scikit-learn 1.7.0 documentation. Available at: <https://scikit-learn.org/stable/modules/preprocessing.html> (accessed 03.09.2025).

The algorithm comprises several steps:

- calculation of pairwise correlation coefficients among the original indicators;
- construction and rotation of the factor loading matrix;
- extraction of principal components;
- selection of factors based on eigenvalue criteria $\lambda_j, j = 1, 2, \dots, S; S < q$, for $\lambda_j \geq 1$, where $S < n$ and $q = 23$.

To ensure interpretability, components are retained such that their cumulative explained variance exceeds 70 %. Full details of the factorization procedure are described in [1].

IV. Transition to Structural Equation Modeling. Following factor extraction, SEM is applied to model the relationships among latent and observed variables³.

To identify latent factors in the development of Belarusian ATEs, the following 23 socio-economic indicators for 2016-2024 were selected and are shown in table 1.

Table 1
Statistical indicator designation

Variable	The variable names in the program	Statistical indicator
x_1	agriculturalland	Agricultural land per 1000 inhabitants, ha
x_2	cattlestockpercapita	Number of cattle per capita, thousands of head
x_3	grainpercapita	Gross grain and legume production per capita, thousand tons
x_4	housingprovisionpercap	Housing provision per capita, m^2
x_5	milk	Milk production per capita, tons
x_6	overduereceivables	Overdue receivable, % of total receivables
x_7	overduecredible	Overdue credible, % of total payables
x_8	overdueloans	Overdue loans, % of total payables
x_9	profitability	Profitability of sales, %
x_{10}	contractedworks	Growth rate of contracted works, %
x_{11}	investment	Growth rate of capital investment, %
x_{12}	constructionwork	Growth rate of construction and installation work, %
x_{13}	residential area	Growth rate of commissioning of residential area, %
x_{14}	housingstock	Growth rate of housing stock, %
x_{15}	smallentreprises	Number of small and medium-sized enterprises per 1000 inhabitants
x_{16}	industrialpercapita	Industrial output per capita, thousand BYN
x_{17}	revenuepercapita	Revenue per capita, thousand BYN
x_{18}	ratehireddismissed	Ratio of hired to dismissed employees
x_{19}	employed	Growth rate of employment, %,
x_{20}	density	Population density, residents per km^2
x_{21}	indexagri	Growth rate of agricultural output, %
x_{22}	publicltrade	Growth rate of public trade turnover, %
x_{23}	retail	Retail turnover per capita, rubles

³Contreras, C. Analysis and visualisation of SEM using Python SEMOPY (Structural Equation Modeling). *Medium*. Available at: <https://cr-c.medium.com/analysis-and-visualisation-of-sem-using-python-semopy-structural-equation-modeling-bd6c3f463e33> (accessed 03.09.2025).

Factorization procedure and dimensional reduction. The factorization procedure was performed using the Python programming language in the Colab environment. Factor component analysis (FCA) was applied to determine latent factors. To validate the computations, additional calculations were conducted using SPSS Statistics.

The dataset includes 128 administrative districts across the years 2016-2024. For further processing, the data were compiled into a matrix comprising $128 \times 9 = 1152$ complete observations over a 9-year period covering all 128 administrative districts of Belarus across 23 variables. Eigenvalues of the correlation coefficient matrix – representing the proportion of total variance in the original indicator system explained by each principal component following axis rotation – are provided in table 2.

As a result of the factor analysis, the initial 23 indicators were transformed into 7 principal components, which together preserve 70,4 % of the system's total variance. Based on data from 2016-2024, the first two components explain approximately 17,3 % of the original variance. In this case, the first four components retain 47,8 % of the variance, while the cumulative variance retained by the first six components reaches approximately 63,2 %. Therefore, the structural relationships among indicators were analyzed using 7 principal components (fig. 1).

Determination of the number of components based on eigenvalue distribution. Model adequacy was assessed using the Kaiser–Meyer–Olkin (KMO) measure of sampling suitability. The KMO value, which reflects the degree to which the data are appropriate for factor analysis, was approximately 0,74, indicating a satisfactory level of adequacy (table 3). An additional measure of inter-variable association is Bartlett's test of sphericity. This test evaluates the null hypothesis that the variables in the correlation matrix are mutually uncorrelated. The observed significance level was 0.00, which is sufficiently low to reject the null hypothesis. Consequently, the relationships among variables are statistically significant, and the extraction of latent factors is justified.

Table 2
Total explained variance

Factor	Extraction (sum of squared loadings)			Rotation (sum of squared loadings)		
	Eigenvalue	Variance, %	Cumulative variance, %	Eigenvalue	Variance, %	Cumulative variance, %
F1	5,74	24,94	24,94	3,98	17,31	17,31
F2	3,02	13,14	38,08	2,73	11,88	29,19
F3	2,00	8,68	46,76	2,43	10,58	39,77
F4	1,79	7,79	54,54	1,84	7,99	47,76
F5	1,30	5,66	60,21	1,78	7,74	55,50
F6	1,22	5,29	65,50	1,77	7,69	63,19
F7	1,12	4,86	70,36	1,65	7,17	70,36

KMO: 0.739 – right
Bartlett's test: $\chi^2 = 99310.0$, $p = 0.0000$

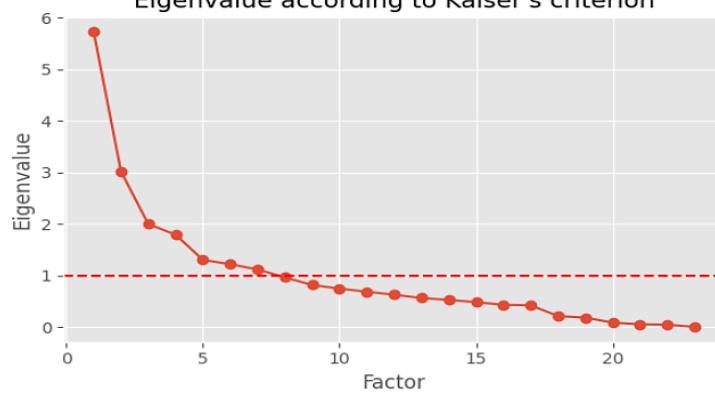


Fig. 1. KMO criteria

Table 3
Kaiser–Meyer–Olkin measure of sampling adequacy
and Bartlett's test of sphericity

Statistic	Value
Kaiser–Meyer–Olkin measure	0,739
Bartlett's test of sphericity:	
approximate χ^2 :	99310
degrees of freedom;	136
significance (p < 0,001)	0,000

At the next stage, each of the seven extracted components is interpreted based on the variables that load onto it. For further analysis, exactly seven principal components were retained. The factor analysis was performed on a standardized 1152×23 data matrix, and after Varimax rotation with Kaiser normalization we obtained a 23×7 loading matrix. The rotated loading matrix is visualized in fig. 2 as a heatmap generated in the Colab environment.

In the figure, «warm» colors correspond to loadings exceeding 0,6–0,7, while «cool» colors represent coefficients below this threshold, yet still meaningful in magnitude. Each of the six components incorporates a distinct subset of the 20 selected indicators.

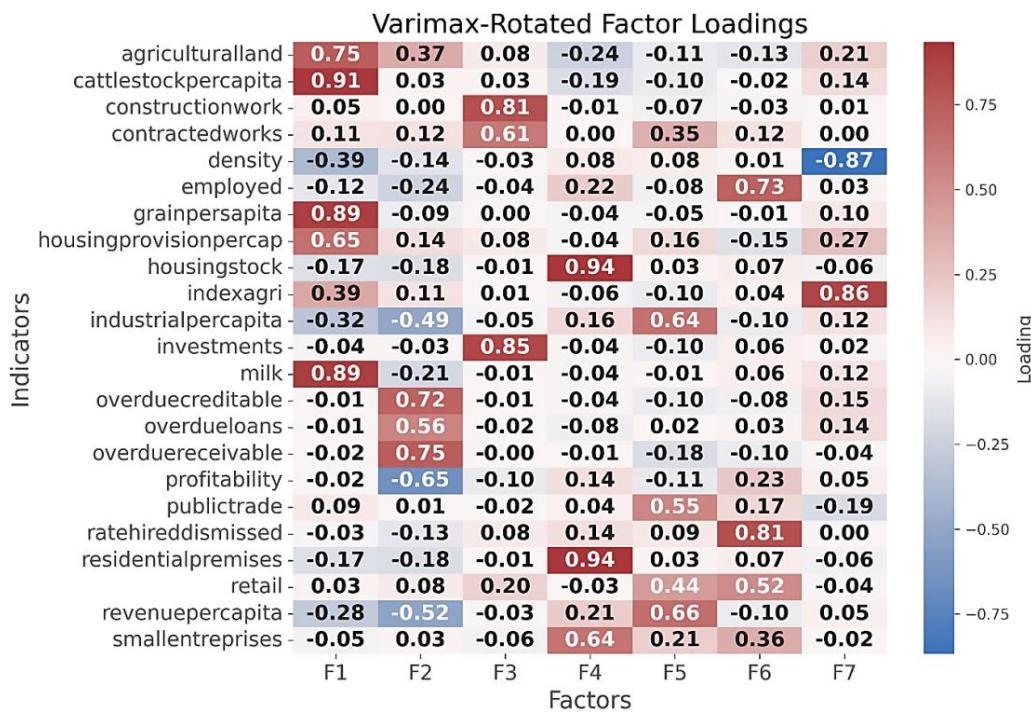


Fig. 2. Heatmap visualization of the rotated factor loading matrix

The rotated factor loading matrix reveals the following generalized structure: components F1–F7 are formed under the influence of specific observed indicators. Table 4 presents each principal component, its economic interpretation, and the variables associated with each of the 7 factors. Based on table 4, and to make the factor labels in the measurement and structural models as clear as possible, we introduce the following labels for factors in the Python program. These labels map factors F1–F7 to names that will be used in the measurement and structural models, tables, figures, and model outputs. In the code and the measurement and structural models F1 is displayed as «Agriculture», F2 – as «Finance», F3 – as «Investments», F4 – as «HousingFund», F5 – as «Industry», F6 – as «Employment», F7 – as «DemographicAgriculturalBalance».

Table 4
Economic interpretation of the factors

Factor	Economic interpretation	Variables forming the factor
F1	Level of agricultural development of the region	x_1 – agricultural land area per 1000 inhabitants, ha; x_2 – number of cattle per capita, thousands of heads; x_3 – grain per capita, tons; x_4 – housing provision per capita, m^2 ; x_5 – milk production per capita, tons
F2	Regional financial stability	x_6 – overdue receivable, % of total receivables; x_7 – overdue credible, % of total payables; x_9 – profitability of sales, %
F3	Investment potential level	x_{10} – growth rate of contracted works, %; x_{11} – growth rate of capital investment, %; x_{12} – growth rate of construction and installation work, %
F4	Level of business and construction activity	x_{13} – growth rate of commissioning of residential area, %; x_{14} – growth rate of housing stock, %; x_{15} – number of small and medium-sized enterprises per 1000 inhabitants
F5	Industrial potential	x_{16} – industrial output per employee, thousand BYN; x_{17} – revenue per capita, thousand BYN
F6	Employment level	x_{18} – ratio of hired to dismissed employees; x_{19} – growth rate of employment, %
F7	Demographic-agrarian balance	x_{20} – population density, residents per km^2 ; x_{21} – growth rate of agricultural output, % Note: negative loading on agricultural output index reflects urbanization and workforce concentration in developed urban territories

The set of variables influencing the first principal component reflects the level of agricultural development and specialization of the territories. Component F1 is economically interpreted as the «*Level of agricultural development of the region*». High loadings on agricultural land area, gross grain harvest, cattle population, and milk production per capita indicate a pronounced agrarian profile. This factor is important for assessing the resilience of the agricultural sector and the region's food-security potential, explaining approximately 17,3 % of the variance attributed to the first component.

Component F2, labeled «*Regional financial stability*», explains 11,9 % of the variance and captures key indicators of organizational risk and solvency. It loads positively on overdue receivables and payables and negatively on sales profitability, meaning that higher F2 scores correspond to regions where mounting debt burdens coincide with declining profit margins – signaling weak liquidity and elevated financial risk.

Component F3 «*Investment potential level*» reflects the region's capacity to attract and deploy investments with high returns. Two key variables exhibit strong loadings, and this component explains about 10,6 % of the total variance.

Component F4 clusters indicators describing housing infrastructure and entrepreneurship: total housing stock, number of dwelling units, and count of small enterprises. It accounts for roughly 8,0 % of the variance and is interpreted as the «*Level of business and construction activity*».

Component F5 «*Industrial potential*» characterizes regions with developed industrial sectors, high revenues. This factor, indicating macroeconomic resilience and an industrial orientation, explains 7,7 % of the variance.

Component F6 «*Employment level*» (7,7 % of the variance) captures labor market conditions, consumer activity, and social stability.

Component F7 «*Demographic-agrarian balance*» links population density with physical agricultural output. The negative loading on agricultural growth rates suggests urbanization trends: as economies mature, workforce and capital shift from agriculture to industrial and service sectors. This factor explains 7,2 % of the variance.

After preliminary factor analysis using Varimax rotation, the structural equations were formulated and depicted in a comprehensive path diagram (fig. 3). SEM belongs to second-generation multivariate

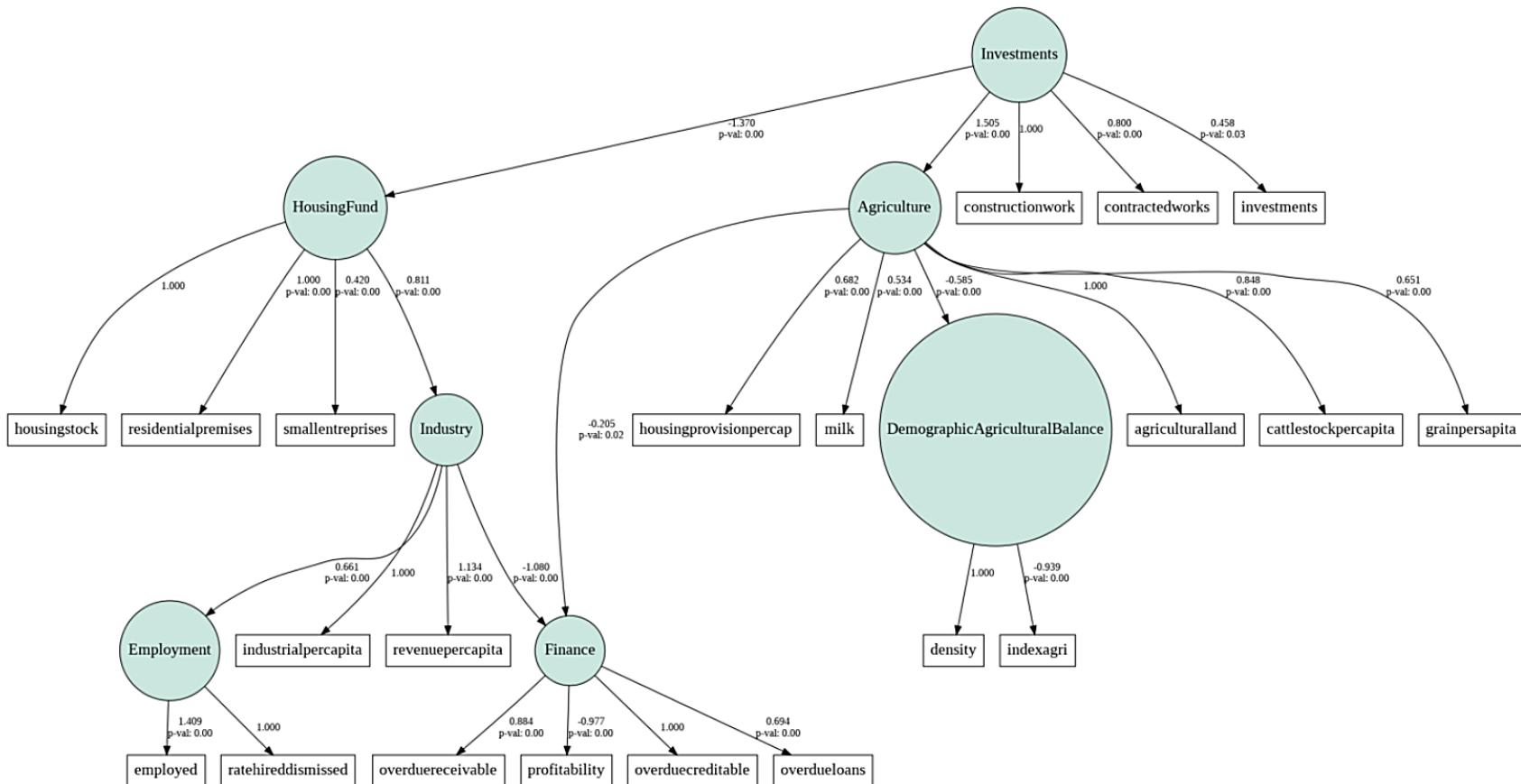


Fig. 3. Structural equation model of the socio-economic development of Belarusian regions for 2016-2024

Each of these measurement linkages is confirmed by high z-statistic values and extremely low p-values ($p < 0,001$ or thereabouts), indicating the statistical significance of the selected indicators in describing the latent variables (fig. 3). The path diagram in fig. 3 illustrates how latent (unobserved) variables and their observed indicators interrelate and influence one another. Latent variables are depicted as circles and observed indicators as rectangles. Arrows indicate the direction of influence, alongside standardized coefficients and p-values that attest to the statistical significance of each path. Path Diagram of relationships among 7 factors illustrated on fig. 4. On the fig. 4 * denotes ($p < 0,05$) – the probability of obtaining this or a more extreme result under the null hypothesis is less than 5 %, ** denotes ($p < 0,01$) – the probability of obtaining this effect by chance is less than 1 %.

These results underscore the reliability of the measurement model and the objectivity of the obtained estimates, allowing them to be used for further interpretation of structural relationships in the regional development model.

Structural model:

DemographicAgriculturalBalance ~ Agriculture

Agriculture ~ Investments

Finance ~ Agriculture + Industry

HousingFund ~ Investments

Industry ~ HousingFund

Employment ~ Industry

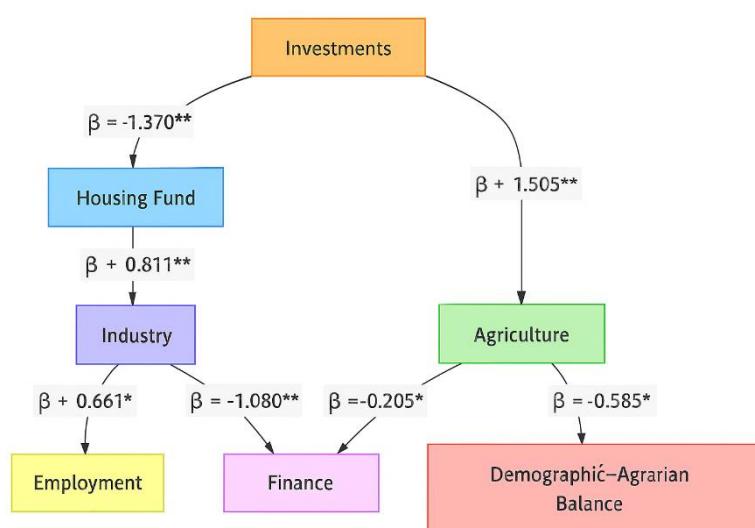


Fig. 4. Diagram of relationships among 7 factors

In fig. 3, adjacent to each path, the following metrics are shown:

Estimate – the standardized regression coefficient indicating the effect size;

Std. Err. – the standard error of the estimate; lower values indicate higher precision;

z-value – the Wald statistic testing how far the coefficient deviates from zero;

p-value – the significance level, with all paths deemed significant at $p < 0,01$.

Let's examine and analyze these metrics in greater detail. Each indicator loading onto the latent factor is significant ($p \leq 0,05$).

Main structural relationships and their interpretation. The structural part of the model comprises seven directed paths between latent factors:

1. *Agriculture → DemographicAgriculturalBalance* (Estimate = $-0,585$, $p < 0,05$). This indicates that in regions with higher demographic indicators and agricultural potential, the economy experiences a structural shift characterized by a relative decline in the role of agriculture. This pattern can be explained by processes of urbanization and the migration of labor resources into other sectors of the economy. Beyond this, such shifts often mirror the classic Lewis dual-sector model, where surplus rural

labor is absorbed into industrial and service sectors. Policy responses might include diversifying rural economies, investing in agro-processing industries, and bolstering social infrastructure to manage urban migration pressures.

2. *Agriculture → Finance* (Estimate = $-0,205$ (weak connection), $p < 0,05$). This may indicate that agrarian-oriented regions, on average, exhibit a less developed financial sector.

3. *Industry → Finance* (Estimate = $-1,080$, $p < 0,05$). Since the Finance factor shows a strong direct relationship with debt indicators, this suggests that industrial enterprises in the regions are highly indebted. A high level of industrial development accompanied by a high share of overdue debt (as indicated by the loadings on the Finance factor) can create macroeconomic imbalances that negatively affect the short-term stability of the financial system. This warrants further in-depth analysis.

4. *Investments → Agriculture* (Estimate = $+1,505$, $p < 0,05$). A coefficient exceeding 1 indicates a high sensitivity of the agricultural sector to investment volume: small changes in investment levels lead to more pronounced shifts in the agrarian sector.

5. *Investments → HousingFund* (Estimate = $-1,370$, $p < 0,05$). A negative relationship points to a «crowding-out effect»: as investments in the economy increase – for example, channelled into industrial infrastructure, major construction projects, or modernization – resources allocated to maintaining and expanding the housing stock (such as capital repairs, adding new residential units, or affordable-housing programs) may be reduced. In practical terms, this means that in regions experiencing intensified investment activity, housing-sector indicators tend to weaken simultaneously, highlighting the need for a more detailed analysis of how public and private investment funds are redistributed across sectors.

6. *HousingFund → Industry* (Estimate = $+0,811$, $p < 0,05$). Conceptually, stronger housing-sector performance (larger housing stock, more residential premises, and vibrant small-enterprise activity in construction) appears to stimulate industrial output. Improved housing infrastructure can boost labor availability, raise demand for building materials and durable goods, and spur related manufacturing, all of which drive up overall industrial activity.

7. *Industry → Employment* (Estimate = $+0,061$, $p < 0,05$). Although modest in size, this coefficient confirms that stronger industrial activity contributes to higher employment levels. The relatively small magnitude suggests that while industry does drive job creation, other sectors or macro-factors (e.g., services, government policy, automation) also play important roles in shaping overall employment.

All paths are statistically significant ($p < 0,05$), underscoring the robustness and interpretability of the proposed structural scheme.

Model fit assessment. The model fit indices are summarized in table 5.

Table 5
Model fit indices

Metric	DoF	DoF Baseline	chi2	chi2 p-value	chi2 Baseline	CFI	GFI	AGFI	NFI	TLI	RMSEA	AIC	BIC	LogLik
Value	182	210	229,3	0,01	2448,8	0,98	0,91	0,89	0,91	0,98	0,045	75,5	215,2	11,3

For the assessment of model-data fit, traditional fit indices were used, based on a comparison of the observed and implied covariance structures.

Interpretation of model fit indices:

DoF (Degrees of Freedom). The number of independent constraints in the model. In the table: **DoF = 182** for the tested model and **DoF Baseline = 210** for the baseline model. Higher DoF indicates a more «free» model; very low DoF may signal over-identification.

Chi-square (χ^2) and p-value. A small χ^2 and a high p-value ($>0,05$) indicate good fit. In this case **p = 0,01**, so the model differs statistically from a perfect fit; however, with large samples such deviations are common and should be interpreted together with other indices.

Comparative Fit Index (CFI). **CFI = 0,98**, exceeding the conventional threshold ($\geq 0,95$), which indicates excellent model fit.

Goodness-of-Fit Index (GFI). **GFI = 0,91**, above the usual cutoff ($\geq 0,90$), indicating good fit.

Adjusted Goodness-of-Fit Index (AGFI). **AGFI = 0,89**, slightly below the conventional threshold ($\sim 0,90$). This suggests a modest degree of over-parameterization that could be simplified; AGFI adjusts GFI for model complexity.

Normed Fit Index (NFI). **NFI = 0,91**, above the common threshold ($\geq 0,90$), which is considered acceptable. NFI compares the χ^2 of the tested model to that of the baseline model.

Tucker–Lewis Index (TLI). **TLI = 0,98**, exceeding the recommended cutoff ($\geq 0,95$). TLI penalizes model complexity and rewards parsimony.

Root Mean Square Error of Approximation (RMSEA). **RMSEA = 0,0452**, below the conventional threshold ($\leq 0,05$), indicating a close approximate fit of the model to the data.

Information criteria and log-likelihood. Relatively low values of **AIC = 75,47** and **BIC = 215,22**, together with a higher log-likelihood (**LL = 11,26**), support the conclusion that this specification achieves a favorable balance between fit and parsimony compared with alternative models.

Taken together, these indices demonstrate that the model effectively captures the relationships among latent and observed variables while maintaining statistical stability and interpretability. Further refinement could involve accounting for temporal lags or conducting multi-group analyses across different types of regions. Overall, the proposed SEM can serve as an analytical tool for identifying priority areas in territorial development, evaluating regional policy effectiveness, and informing evidence-based economic planning [3].

Future research directions include:

- incorporating time-lag effects to assess the long-term impact of investments;
- adding supplementary indicators (environmental quality, innovation capacity, human capital);
- employing of updated regional statistical data to increase the accuracy of estimates and to test the robustness of the identified latent factors;
- applying spatial econometric models to explore interregional interactions.

This study introduces a structural equation model to explore the nuanced interdependencies among regional investment activity, infrastructure provision, agricultural productivity, industrial output, financial performance, population density, and employment.

The framework combines a measurement model – linking observed indicators to their underlying latent constructs – and a structural model – estimating the directional effects between those constructs – thereby allowing us to assess both construct reliability and the causal interplay among key economic drivers.

Thus, the model identifies the most critical drivers of sustainable development and competitiveness in regional economies. SEM has proven effective for analyzing intricate socio-economic systems, and the findings can underpin targeted management of regional development in Belarus.

The practical significance of this research lies in its potential to optimize regional policy by accounting for territorial specifics, guiding the targeted allocation of investments for balanced growth, and monitoring the effectiveness of government development programs.

References

1. Chitai G. O. Analysis of the dynamics of integral economic indicators of the regions of the Republic of Belarus. *Economics, Modeling, Forecasting*, 2018, no. 12, pp. 43–51 (In Russ.).
2. Igolkina A. G., Meshcheryakov G. Semopy: A python package for structural equation modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, 2020, vol. 27, iss. 6, pp. 952–963. DOI: 10.1080/10705511.2019.1704289.
3. Kline R. B. *Principles and Practice of Structural Equation Modeling*. New York, Guilford Press, 2023, 494 p.

Information about the author

Yuliya A. Osipova, Scientist, M. Sc. (Phys.-Math.), The Economy Research Institute of the Ministry of Economy of the Republic of Belarus.
E-mail: osipova@niei.by
<https://orcid.org/0009-0005-3109-6775>

Информация об авторе

Осипова Юлия Александровна, научный сотрудник, магистр физико-математических наук, Научно-исследовательский экономический институт Министерства экономики Республики Беларусь.
E-mail: osipova@niei.by
<https://orcid.org/0009-0005-3109-6775>

ОБРАБОТКА СИГНАЛОВ, ИЗОБРАЖЕНИЙ, РЕЧИ, ТЕКСТА И РАСПОЗНАВАНИЕ ОБРАЗОВ

SIGNAL, IMAGE, SPEECH, TEXT PROCESSING AND PATTERN RECOGNITION



УДК 004.21
DOI: 10.37661/1816-0301-2025-22-4-36-54

Оригинальная статья
Original Article

Нейронные сети на основе обучаемого двумерного разделимого преобразования для классификации изображений: теория и аппаратная реализация на FPGA

Е. А. Кривальцевич, М. И. Вашкевич[✉]

Белорусский государственный университет
информатики и радиоэлектроники,
ул. П. Бровки, 6, Минск, 220013, Беларусь
[✉]E-mail: vashkevich@bsuir.by

Аннотация

Цели. Целями исследования являются разработка методов построения компактных и эффективных нейронных сетей для задач распознавания изображений, а также их аппаратная реализация на базе программируемых логических интегральных схем (ПЛИС) типа FPGA.

Методы. Предложена концепция обучаемого двумерного разделимого преобразования (ОДРП) для построения нейронных сетей прямого распространения для задач распознавания изображений. Особенностью ОДРП является последовательная обработка строк изображения полносвязанным слоем, после чего полученное представление обрабатывается по строкам вторым полносвязанным слоем. В предлагаемой архитектуре нейронной сети прямого распространения ОДРП рассматривается как способ извлечения вектора признаков из исходного изображения. Аппаратная реализация нейронной сети на базе ОДРП основана на концепции вычисления «на месте» (общая память для хранения исходных и промежуточных данных), а также на использовании единого набора вычислительных ядер для расчета всех слоев нейронной сети.

Результаты. Предложено семейство компактных нейросетевых архитектур LST-1, различающихся размерностью векторного представления изображения. Эксперименты по классификации рукописных цифр базы MNIST показали высокую эффективность данных моделей: сеть LST-1-28 достигает точности 98,37 % при 9,5 тыс. параметров, а более компактная LST-1-8 показывает 96,53 % точности при 1,1 тыс. параметров. Тестирование аппаратной реализации LST-1-28 подтверждает устойчивость архитектуры к ошибкам квантования параметров.

Заключение. Предложенная концепция ОДРП обеспечивает проектирование компактных и эффективных нейросетевых архитектур, характеризующихся малым числом обучаемых параметров, высокой точностью распознавания и регулярной структурой алгоритма, что позволяет получать их эффективные реализации на базе ПЛИС.

Ключевые слова: обучаемое двумерное разделимое преобразование, нейронные сети, программируемые логические интегральные схемы, распознавание изображений, база данных MNIST

Благодарности. Авторы выражают благодарность резиденту ПВТ компании «Инженерный Центр Ядро», которая является одним из центров разработки YADRO, за предоставленное оборудование для проведения экспериментов в рамках работы совместной учебной лаборатории с Белорусским государственным университетом информатики и радиоэлектроники.

Для цитирования. Кривальцевич, Е. А. Нейронные сети на основе обучаемого двумерного разделимого преобразования для классификации изображений: теория и аппаратная реализация на FPGA / Е. А. Кривальцевич, М. И. Вашкевич // Информатика. – 2025. – Т. 22, № 4. – С. 36–54. – DOI: 10.37661/1816-0301-2025-22-4-36-54.

Конфликт интересов. Авторы заявляют об отсутствии конфликта интересов.

Поступила в редакцию | Received 16.09.2025

Подписана в печать | Accepted 07.10.2025

Опубликована | Published 30.12.2025

Neural networks based on a learnable two-dimensional separable transform for image classification: theory and hardware implementation on FPGA

Egor A. Krivalcevich, Maxim I. Vashkevich[✉]

Belarusian State University
of Informatics and Radioelectronics,
st. Brovki, 6, Minsk, 220013, Belarus
[✉]E-mail: vashkevich@bsuir.by

Abstract.

Objectives. Development of methods for design compact and efficient neural networks for image recognition tasks, as well as their hardware implementation based on FPGA.

Methods. The paper proposes the concept of a learnable two-dimensional separable transformation (LST) for designing feedforward neural networks for image recognition tasks. A feature of the LST is the sequential processing of image rows by a fully connected layer, after which the resulting representation is processed by columns using second fully connected layer. In the proposed architecture of a feedforward neural network, the LST is considered as a feature extractor. The hardware implementation of LST-based neural network is based on the concept of in-place computing (shared memory for storing source and intermediate data), as well as using a single set of computing cores to calculate all layers of the neural network.

Results. A family of compact neural network architectures LST-1 is proposed, differing in the image embedding size. Experiments on the classification of MNIST handwritten digits have shown the high efficiency of these models: the LST-1-28 network achieves 98.37 % accuracy with 9.5 K parameters, and the more compact LST-1-8 shows 96.53 % accuracy with 1.1 K parameters. Testing of the LST-1-28 hardware implementation confirms the architecture's resistance to parameter quantization errors.

Conclusion. The proposed concept of a learnable two-dimensional separable transformation provides the design of compact and efficient neural network architectures characterized by: a small number of learnable parameters, high recognition accuracy, and the regular structure of the algorithm, which makes it possible to obtain their effective implementations based on FPGAs.

Keywords: learnable two-dimensional separable transform, neural networks, FPGA, image recognition, MNIST dataset

Acknowledgements. The authors express their gratitude to the HTP resident company "Engineering Center Yadro", which is one of the YADRO development centers, for providing equipment for conducting experiments as part of the joint educational laboratory with the Belarusian State University of Informatics and Radioelectronics.

For citation. Krivalcevich E. A., Vashkevich M. I. *Neural networks based on a learnable two-dimensional separable transform for image classification: theory and hardware implementation on FPGA*. Informatika [Informatics], 2025, vol. 22, no. 4, pp. 36–54 (In Russ.). DOI: 10.37661/1816-0301-2025-22-4-36-54.

Conflict of interest. The authors declare of no conflict of interest.

Введение. Глубокие нейронные сети (НС) являются ключевыми компонентами многих систем компьютерного зрения и обработки изображений благодаря их высокой эффективности и способности моделировать сложные зависимости. Однако реализация НС на аппаратных платформах с ограниченными вычислительными ресурсами, таких как ПЛИС, сталкивается с рядом проблем, главной из которых является высокая вычислительная нагрузка, обусловленная большим числом параметров НС. Для повышения эффективности необходимо разрабатывать НС с уменьшенным числом параметров [1–5]. Такой подход позволяет снизить требования к объему памяти и вычислительным ресурсам. Однако большинство существующих архитектур с малым числом параметров основаны на классическом многослойном персептроне [6, 7], что зачастую приводит к снижению точности распознавания.

Таким образом, актуальной научной задачей является разработка эффективных архитектур НС, способных достигать компромисса между числом параметров и уровнем точности распознавания. Данный факт подтверждается и проводимыми в научной среде соревнованиями по разработке высокоскоростных и малопотребляющих реализаций НС на базе ПЛИС. В частности, на Международной конференции по обработке изображений в 2025 г. проводилось соревнование «Digit Recognition Low Power and Speed Challenge», где предлагалось разработать архитектуру ускорителя на базе ПЛИС для распознавания рукописных цифр из базы MNIST. К соревнованиям допускались только проекты, которые на тестовой выборке демонстрировали точность не менее 97,5 %.

Одним из перспективных подходов к построению компактных и эффективных НС для распознавания изображений является использование ОДРП [8]. В настоящей статье рассматриваются принципы работы ОДРП и практические аспекты реализации НС с его применением на базе ПЛИС. Экспериментальные исследования показывают, что использование ОДРП в качестве средства получения векторного представления (вложения) изображения позволяет получить компактную архитектуру НС, демонстрирующую высокие показатели точности в задаче классификации изображений. Таким образом, данная работа направлена на развитие методов построения компактных и эффективных НС для задач распознавания изображений с реализацией на ПЛИС.

Обучаемое двумерное разделимое преобразование. С вычислительной точки зрения ОДРП (англ. *LST, learned 2D separable transform*) преобразует двумерное изображение \mathbf{X} размером $d_{in} \times d_{in}$ в изображение \mathbf{Y} размером $d_{out} \times d_{out}$. Преобразование называется разделимым, поскольку вначале выполняется обработка строк изображения \mathbf{X} , после чего образуется промежуточное представление изображения, имеющее размерность $d_{in} \times d_{out}$. Далее полученное промежуточное представление обрабатывается по столбцам, в результате чего получается выходное изображение \mathbf{Y} размером $d_{out} \times d_{out}$. Таким образом, разделимое преобразование над изображением можно представить как композицию двух преобразований, работающих с одномерными данными (рис. 1).

Особенностью ОДРП, представленного в работе [8], является то, что для обработки строк и столбов изображения предложено использовать однослойные полносвязные НС. Математически ОДРП можно записать в виде

$$\mathbf{Y} = LST(\mathbf{X}) = \sigma(\mathbf{W}_2 \sigma(\mathbf{W}_1 \mathbf{X}^T + \mathbf{b}_1)^T + \mathbf{b}_2), \quad (1)$$

где $\sigma(\cdot)$ – функция активации; $\mathbf{W}_1, \mathbf{W}_2$ – матрицы $d_{out} \times d_{in}$ линейных преобразований, описывающих полносвязные слои НС для обработки строк и столбцов соответственно; $\mathbf{b}_1, \mathbf{b}_2$ – векторы смещений размером $d_{out} \times 1$.

На рис. 2 показано графическое представление математических действий, выполняемых при реализации ОДРП.

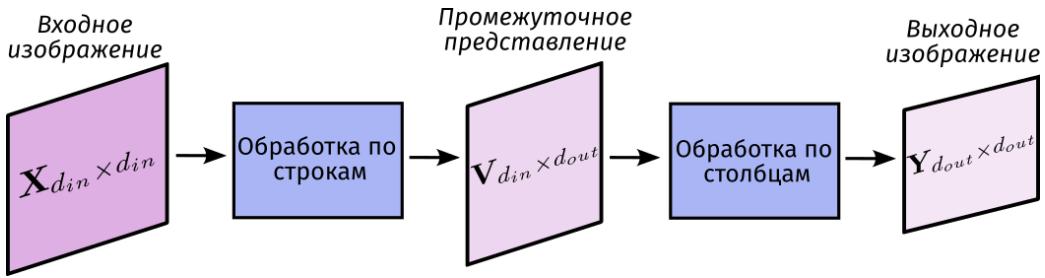


Рис. 1. Принцип работы разделимого преобразования
 Fig. 1. The principle of operation of a separable transform

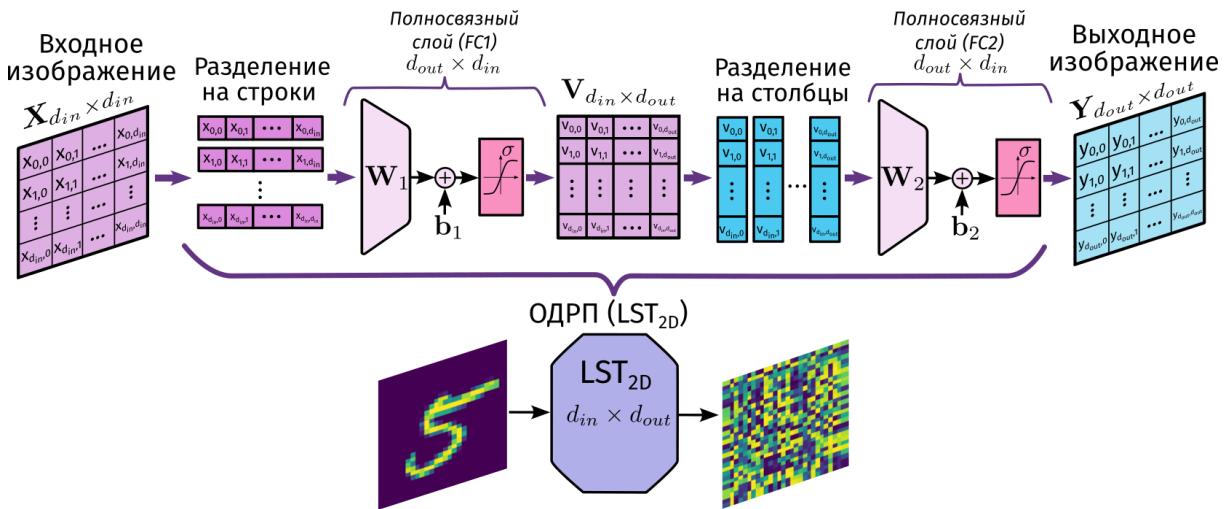


Рис. 2. Обработка изображения обучаемым двумерным разделимым преобразованием
 Fig. 2. Image processing by a learnable two-dimensional separable transform

Можно заметить, что ОДРП состоит из двух этапов обработки. На первом этапе входное изображение разделяется на строки и обрабатывается полносвязным слоем FC1, к выходу которого применяется нелинейная функция активации. После этого результирующее представление обрабатывается по столбцам с использованием слоя FC2. Выход \mathbf{Y} ОДРП будем называть вложением (англ. *embedding*). Размер полученного вложения определяется параметром d_{out} . Общее число настраиваемых параметров ОДРП находится по формуле

$$N_{LST} = 2 \cdot (d_{in} + 1) \cdot d_{out}. \quad (2)$$

НС LST-1 для распознавания изображений на основе ОДРП. ОДРП можно использовать в качестве базового блока при конструировании НС для распознавания изображений. В настоящей работе предлагается рассмотреть НС, состоящую из одного блока ОДРП и классифицирующего полносвязного слоя с активационной функцией softmax (рис. 3). Внутри блока ОДРП в качестве активационной функции используется гиперболический тангенс.

НС, изложенная на рис. 3, работает следующим образом: на вход подается изображение размером 28×28 , которое при помощи блока ОДРП преобразуется во вложение размером $d_{out} \times d_{out}$. Слой Flatten выполняет преобразование вложения в одномерный вектор размером d_{out}^2 . Далее полносвязный слой, параметрами которого являются матрица весов \mathbf{W}_o размером

$d_{out}^2 \times n_{classes}$ и вектор смещений \mathbf{b}_o размерности $n_{classes}$, с активационной функцией softmax выполняет классификацию изображения, вычисляя вероятности отнесения изображения к каждому из $n_{classes}$ классов. В качестве результата классификации выбирается класс, который имеет наибольшую вероятность.

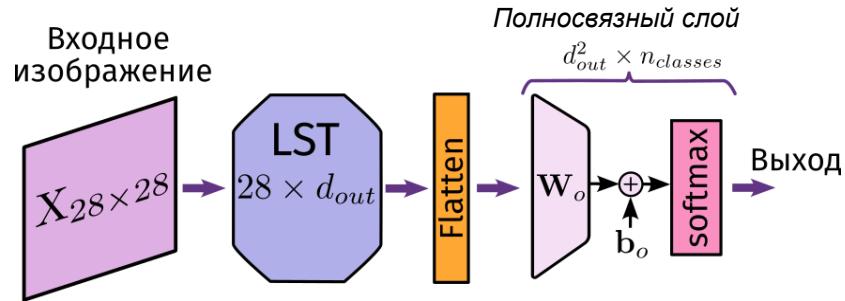


Рис. 3. Архитектура нейронной сети LST-1- d_{out}

Fig. 3. Architecture of neural network LST-1- d_{out}

Представленная модель в дальнейшем будет обозначаться как LST-1- d_{out} , где d_{out} – число, определяющее размерность получаемого внутри НС вложения, от которого зависит общее число настраиваемых параметров модели. Более строго общее число параметров модели LST-1- d_{out} определяется как сумма числа параметров ОДРП и числа параметров полносвязного выходного слоя:

$$N_{\text{LST-1}} = N_{\text{LST}} + N_{\text{FC}} = 2 \cdot (d_{in} + 1) \cdot d_{out} + n_{classes} \cdot (d_{out}^2 + 1), \quad (3)$$

где $n_{classes}$ – число распознаваемых классов.

Значения d_{in} и $n_{classes}$, как правило, определяются условиями задачи и являются фиксированными, в то время как d_{out} может определяться на этапе разработки модели. Заметим, что от параметра d_{out} зависит емкость (англ. *capacity*) модели и ее обобщающая способность. Однако, как следует из выражения (3), его увеличение вызывает квадратичный рост числа настраиваемых параметров.

На рис. 4 показано, как происходит обработка изображения в обученной сети LST-1-28.

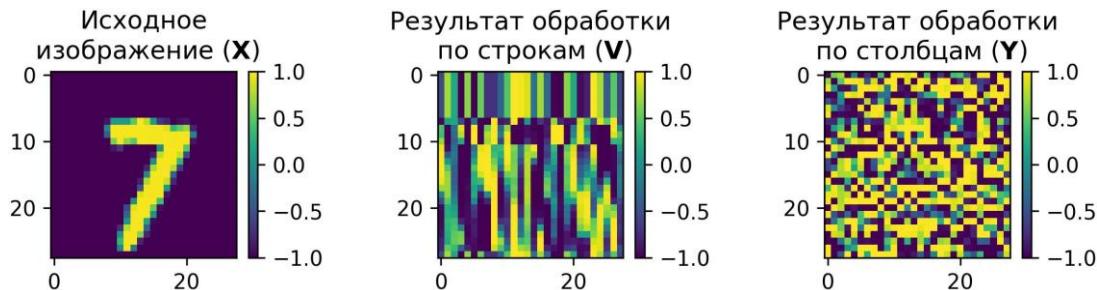


Рис. 4. Обработка изображения в нейронной сети LST-1-28

Fig. 4. Image processing in the LST-1-28 neural network

На первом этапе исходное изображение \mathbf{X} обрабатывается по строкам, в результате чего образуется промежуточное представление \mathbf{V} . Можно заметить, что первые семь строк \mathbf{V} имеют одинаковые значения. Это связано с тем, что на исходном изображении первые семь

строк также имеют одинаковые значения, равные -1 . Результат обработки по столбцам \mathbf{Y} (или выход ОДРП) представляет собой вложение, которое после «разворачивания» его в вектор подается на вход классифицирующего слоя. Можно отметить, что \mathbf{Y} является рандомизированным представлением исходного изображения, соседние пиксели \mathbf{Y} не имеют или имеют очень слабую корреляционную связь между собой.

Реализация НС LST-1 на ПЛИС. В данном разделе описывается функциональная схема IP-блока для аппаратной реализации модели LST-1 на ПЛИС. Интерфейс разработанного аппаратного модуля показан на рис. 5.

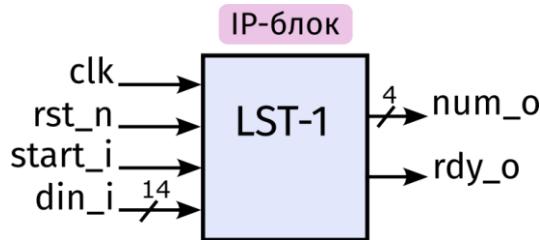


Рис. 5. Интерфейс IP-блока нейронной сети

Fig. 5. Interface of the neural network IP block

IP-блок имеет входы для управляющих сигналов сброса (rst_n) и запуска ($start_i$), а также информационный вход din_i , использующийся для подачи пикселей изображения. После подачи на вход всех пикселей изображения IP-блок автоматически переходит в режим вычисления выходной метки. По окончании вычислительного процесса на выходе rdy_o устанавливается значение логической единицы, а метка распознанного класса поступает на вход num_o .

Общий подход, применяемый в реализации данного IP-блока, заключается в переиспользовании аппаратных ресурсов ПЛИС. IP-блок имеет ОЗУ, в которое в начальный момент времени помещается исходное изображение. Затем выполняется обработка изображения по строкам и запись результата в то же ОЗУ. Далее происходит обработка полученного представления изображения по столбцам с записью результата в ОЗУ. На заключительном этапе производятся вычисления, реализующие полносвязанный классифицирующий слой, данные для которого также берутся из общего ОЗУ. Используя возможности ПЛИС по реализации параллельных вычислений, предлагается параллельно вычислять элементы строк и столбцов ОДРП, для чего в структуре IP-блока применяются d_{out} вычислительных элементов (англ. *PE, processing element*). Каждый вычислительный элемент (ВЭ) состоит из MAC-ядра (англ. *Multiply and ACcumulate*) и набора ПЗУ, которые хранят параметры модели LST-1.

Ниже приведен общий алгоритм работы IP-блока, реализующего модель LST-1- d_{out} :

1. $cnt_o = 0$
2. **if** ($start_i == 1$) перейти к шагу 3, **else** перейти к шагу 2
3. ОЗУ[cnt_o] = din_i
4. $cnt_o = cnt_o + 1$
5. **if** ($cnt_o == d_{out}^2 - 1$) перейти к шагу 6, **else** перейти к шагу 2
6. **for** $cnt_r = 0, 1, \dots, d_{in} - 1$
7. **for** $i = 0, 1, \dots, d_{in} - 1$
8. $ACC[i] = b_1[i]$
9. **end for**
10. **for** $cnt_c = 0, 1, \dots, d_{in} - 1$
11. **for** $i = 0, 1, \dots, d_{in} - 1$
12. $ACC[i] = ACC[i] + w_1[i, cnt_c] * OZU[cnt_r * d_{in} + cnt_c]$

```

13. end for
14. end for
15. for cnt_c = 0, 1, ...,  $d_{in} - 1$ 
16.   OЗУ[cnt_r *  $d_{in}$  + cnt_c] = Tanh(ACC[cnt_c])
17. end for
18. end for
19. for cnt_c = 0, 1, ...,  $d_{in} - 1$ 
20.   for i = 0, 1, ...,  $d_{in} - 1$ 
21.     ACC[i] = b2[i]
22.   end for
23.   for cnt_r = 0, 1, ...,  $d_{in} - 1$ 
24.     for i = 0, 1, ...,  $d_{in} - 1$ 
25.       ACC[i] = ACC[i] + W2[i, cnt_r] * OЗУ[cnt_c *  $d_{in}$  + cnt_r]
26.     end for
27.   end for
28.   for cnt_r = 0, 1, ...,  $d_{in} - 1$ 
29.     OЗУ[cnt_c *  $d_{in}$  + cnt_r] = Tanh(ACC[cnt_r])
30.   end for
31. end for
32. for j = 0, 1, ...,  $n_{classes} - 1$ 
33.   ACC[i] = b[i]
34. end for
35. for cnt_o = 0, 1, ...,  $d_{out} - 1$ 
36.   for j = 0, 1, ...,  $n_{classes} - 1$ 
37.     ACC[j] = ACC[j] + W[j, cnt_o] * OЗУ[cnt_o]
38.   end for
39. end for
40. num_o = argmax(ACC[0, 1, ...,  $n_{classes} - 1$ ])
41. rdy_o = 1

```

В представленном алгоритме шаги 1–5 нужны для начальной записи изображения в ОЗУ. Шаги 6–18 описывают обработку изображения по строкам, а шаги 19–31 – обработку по столбцам. Наконец, шаги 32–41 описывают вычисления, связанные с классифицирующим слоем. Общая структура IP-блока, реализующего данный алгоритм вычисления для случая $d_{out} = 28$, изображена на рис. 6.

Для корректной выборки данных в системе используются три счетчика: два из них отвечают за определение строки и столбца соответствующих данных, а третий служит для сквозной адресации ОЗУ. Вычисление функции гиперболического тангенса (tanh) осуществляется последовательно, что позволяет за 56 системных тактов выполнить обработку одной строки изображения (28 тактов тратится на расчет суммы произведений и 28 тактов – на запись результата в ОЗУ). Для определения выходного значения применяется упрощенная по сравнению с softmax функция argmax, которая не использует вычисление экспоненциальных функций, что способствует уменьшению вычислительной сложности.

Все вычисления, относящиеся к ОДРП и к выходному полносвязному слою, осуществляются с использованием $d_{out} = 28$ ВЭ. При реализации одного физического вычислительного слоя возникает необходимость разделения ВЭ на два типа (рис. 7): rco – row-column-output и rc – row-column.

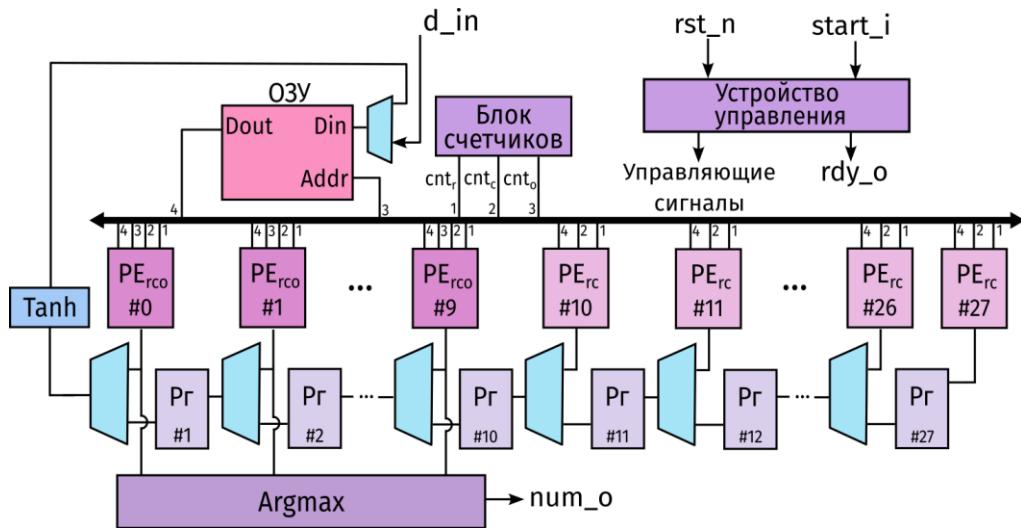


Рис. 6. Структура IP-блока нейронной сети LST-1-28
 Fig. 6. The structure of the LST-1-28 neural network IP block

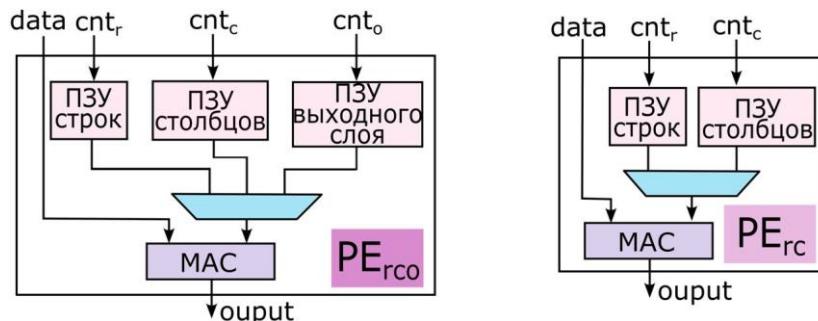


Рис. 7. Архитектура вычислительных блоков нейронной сети
 Fig. 7. Architecture of neural network processing elements

В IP-блоке LST-1 используются 10 ВЭ типа гс0, которые многократно применяются как для вычислений ОДРП, так и для формирования значений выходного полно связного слоя НС. Для поддержки двойной функциональности ВЭ требуются хранилища для коэффициентов выходного слоя и дополнительный адресный вход, обеспечивающий управление доступом к памяти. Оставшиеся 18 ВЭ типа гс предназначены исключительно для обработки строк и столбцов в ОДРП. Переключение режимов вычислений осуществляется устройством управления.

Блок счетчиков (рис. 6) служит для генерации адресов доступа как к памяти весов, так и к ОЗУ, где хранятся промежуточные результаты на всех этапах вычислений.

Слой softmax (см. рис. 3) отвечает за формирование распределения вероятностей по 10 выходным классам (цифры от 0 до 9). Однако нет необходимости использовать функцию активации softmax при аппаратной реализации НС. Вместо этого можно использовать функцию argmax для сравнения всех 10 выходных значений классифицирующего слоя и выбора класса с наивысшим значением (англ. *score*). Кроме того, аппаратная реализация функции argmax задействует значительно меньше ресурсов ПЛИС по сравнению с softmax.

Устройство управления реализовано как конечный автомат с комбинационной логикой для генерации следующего состояния и соответствующих управляющих сигналов для других модулей.

Нелинейная функция активации \tanh аппроксимируется с помощью аппаратно-ориентированной кусочной функции, поскольку прямая реализация потребовала бы больших вычислительных затрат. Аппроксимация выполняется по формуле [2]

$$\tanh(x) \approx F(x) = \begin{cases} \text{sign}(x), |x| > 2, \\ (1 + \frac{x}{4}) \cdot x, -2 < x < 0, \\ (1 - \frac{x}{4}) \cdot x, 0 < x < 2. \end{cases} \quad (4)$$

На рис. 8 показаны исходная функция гиперболического тангенса и ее аппроксимация.

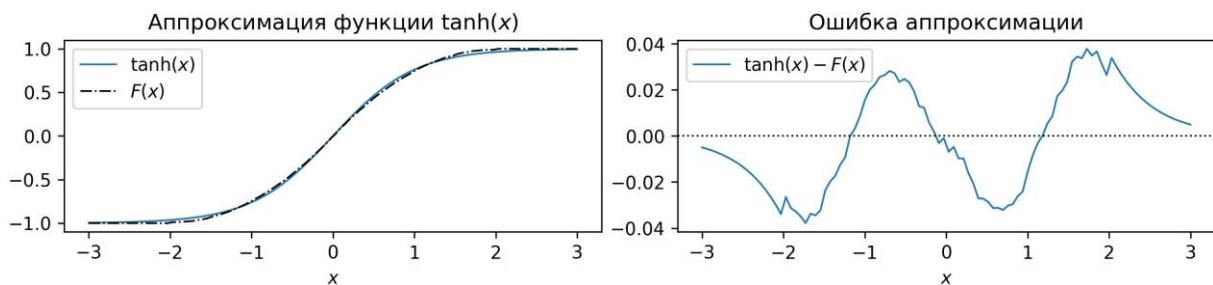


Рис. 8. Сравнение функции \tanh с ее аппроксимацией
Fig. 8. Comparison of the \tanh function with its approximation

Исходя из рис. 8 можно сделать вывод, что аппроксимация позволяет корректно отразить поведение функции активации \tanh . Это говорит о релевантности ее использования с целью снижения аппаратных затрат при реализации на ПЛИС.

Процесс обучения НС LST-1. Обучение НС LST-1 осуществлялось с использованием базы изображений рукописных цифр MNIST. База MNIST состоит из 70 тыс. изображений размером 28×28 пикселей в градациях серого и разбита на две части – тренировочный набор (60 тыс. изображений) и тестовый (10 тыс. изображений). На этапе обучения из тренировочного набора случайным образом выделялась 1 тыс. изображений для использования в качестве валидационного набора, оставшиеся 59 тыс. применялись для обучения модели. Для обучения модели использовались язык Python и библиотека PyTorch.

Перед подачей в НС выполнялась предварительная нормализация изображений. Изначально пиксели изображений представлены числами в диапазоне от 0 до 255. Для упрощения процесса обучения выполняется нормализация данных таким образом, чтобы значение каждого пикселя было в диапазоне от -1 до 1, среднее значение равнялось нулю, а среднеквадратическое отклонение составляло 0,5. Нормализация изображений – распространенный практический прием [9], повышающий устойчивость градиентного спуска и ускоряющий сходимость модели НС. В качестве функции потерь использовался отрицательный логарифм функции правдоподобия (англ. *negative log-likelihood loss, NLLLoss*).

Для оптимизации параметров НС LST-1 применялся метод Adam (англ. *ADaptive Moment estimation*) – один из самых эффективных алгоритмов градиентного спуска с адаптивным шагом. Основной параметр любого метода стохастического градиентного спуска – скорость обучения η , которая, по сути, является гиперпараметром, влияющим на результат обучения. Слишком большое значение η может привести к тому, что в процессе минимизации функции потерь модель не сможет стабилизироваться в точке минимума (глобального или хорошего локального). С другой стороны, слишком малое значение η , как правило, приводит к тому, что модель «застревает» в точке непроизводительного локального минимума. Чтобы избежать двух

описанных сценариев, в данной работе использовался планировщик скорости обучения – метод, который плавно изменяет η в процессе обучения. В частности, скорость обучения определялась функцией циклического косинусного отжига [10]:

$$\eta(t) = \eta_{\min} + \frac{1}{2}(\eta_{\max} - \eta_{\min}) \left(1 + \cos \left(\frac{\text{mod}(t, T_0)}{T_0} \pi \right) \right), \quad (5)$$

где t – номер текущей эпохи; η_{\min} – минимальная скорость обучения; η_{\max} – максимальная скорость; T_0 – количество эпох, в течение которых происходит спад косинуса, прежде чем скорость сбросится.

Таким образом, после каждой эпохи обучения параметр скорости обучения пересчитывался по формуле (5) и с помощью валидационного набора для текущей модели рассчитывалось значение функции потерь. В качестве итоговой выбиралась модель, которая за все время обучения имела наименьшее значение функции потерь на валидационном наборе.

Рассмотренные в настоящей работе модели обучались в течение 300 эпох, длительность одного цикла отжига (параметр T_0 в выражении (5)) выбиралась равной 100, начальная скорость обучения $\eta_{\max} = 0,001$, а минимальная скорость обучения $\eta_{\min} = 5 \cdot 10^{-6}$. На рис. 9 показан график изменения скорости обучения, использовавшийся при обучении моделей.



Рис. 9. График изменения скорости обучения с использованием метода косинусного циклического отжига

Fig. 9. Graph of the change in the learning rate using the cosine annealing with warm restart

Для регуляризации рассмотренных в работе моделей применялся метод дропаут (англ. *dropout*) [11], который заключается в случайном отключении заданной доли p_{drop} нейронов в полносвязном слое в процессе обучения. Это заставляет модель учиться более устойчивым представлениям и давать верные ответы даже при наличии неполных данных на входе. В полносвязных слоях ОДРП модели LST-1, отвечающих за обработку строк и столбцов, значение p_{drop} выбиралось равным 0,1.

Экспериментальная часть работы включает два основных этапа: первый – исследование производительности модели LST-1 при использовании внутренних вложений различной размерности, второй – валидация и тестирование аппаратной реализации модели LST-1 на базе ПЛИС.

Результаты обучения НС LST-1. В работе выполнен анализ производительности модели LST-1- d_{out} для различных значений параметра d_{out} . Как говорилось ранее, в предлагаемой модели LST-1 параметр d_{out} отвечает за размер внутреннего представления входного изображения, а также влияет на общее число параметров модели. Ниже приведены результаты обучения 11 различных моделей LST-1 с разными значениями параметра d_{out} . Для каждой модели проведены 10 независимых экспериментов обучения с различной начальной инициали-

зацией весов. Статистические показатели (среднее значение и стандартное отклонение) точности, рассчитанные по результатам этих экспериментов, представлены в табл. 1. Для регуляризации приведенных в табл. 1 моделей к классифицирующему слою в процессе обучения также применялся метод дропаут. Для моделей со значением параметра $d_{out} = 2, 4, \dots, 16$ для последнего слоя устанавливалось значение $p_{drop} = 0,1$. Для моделей со значениями параметра $d_{out} = 20, 24, 28, 32, 36$ и 40 использовались значения $p_{drop} = 0,15; 0,15; 0,18; 0,20$ и $0,22$ соответственно.

Таблица 1

Число параметров моделей LST-1- d_{out} и их точность на тестовой выборке MNIST

Table 1

The number of parameters of the LST-1- d_{out} models and their accuracy on the MNIST test set

Модель <i>Model</i>	Число параметров <i>Number of parameters</i>	Точность, % <i>Accuracy, %</i>
LST-1-2	166	$75,13 \pm 2,28$
LST-1-4	402	$92,93 \pm 0,66$
LST-1-8	1114	$96,28 \pm 0,19$
LST-1-12	2146	$97,14 \pm 0,17$
LST-1-16	3498	$97,27 \pm 0,13$
LST-1-20	5170	$97,61 \pm 0,06$
LST-1-24	7162	$97,83 \pm 0,14$
LST-1-28	9474	$98,03 \pm 0,14$
LST-1-32	12 106	$98,17 \pm 0,18$
LST-1-36	15 058	$98,23 \pm 0,13$
LST-1-40	18 330	$98,32 \pm 0,08$

Для примера на рис. 10 показаны кривые обучения модели LST-1-20 на тестовом и валидационном наборах. Кривые показывают, что модель LST-1 имеет стабильную сходимость, хотя процесс обучения значительно замедляется после 50-й эпохи. Следует отметить, что после 100-й и 200-й эпох видны возмущения функции потерь как на тренировочном, так и на валидационном наборах, которые объясняются применением метода косинусного отжига для управления процессом обучения модели. Если обратиться к графику на рис. 9, то можно заметить, что именно на 100-й и 200-й эпохах происходят скачкообразные переходы к начальной скорости обучения, которые и вызывают возмущения.



Рис. 10. Функция потерь на тренировочном и валидационном наборах для модели LST-1-20

Fig. 10. Loss function on the training and validation sets for the LST-1-20 model

Разрыв между значениями функции потерь на тренировочном и валидационном наборах обусловлен применением дропаут-регуляризации в процессе обучения. Данный метод, реализуемый через стохастическое зануление активаций скрытых нейронов, вносит дополнительный шум в данные, что приводит к смещению оценки потерь в сторону увеличения для тренировочного набора относительно валидационного.

Для полноты оценки и выявления преимуществ модели LST-1 проведен сравнительный анализ с архитектурами НС, которые чаще всего используются для аппаратной реализации изображений рукописных цифр на базе ПЛИС.

Чаще всего в работах, посвященных реализации НС на ПЛИС для распознавания изображения, в качестве базовой архитектуры рассматривается многослойный перцептрон (МСП) (англ. *MLP, multilayer perceptron*), состоящий из каскада полносвязных слоев [2, 4, 6, 7, 12, 13]. Такой выбор можно объяснить тем, что регулярная структура полносвязных слоев МСП хорошо отображается на архитектуру ПЛИС. Вычислительное ядро таких слоев – операции умножения с накоплением – эффективно реализуется с использованием встроенных DSP-блоков, имеющихся во многих современных ПЛИС. Наконец, в отличие от сверточных НС, МСП имеет простые и предсказуемые паттерны доступа к памяти данных и весовых коэффициентов, что существенно упрощает логику управления памятью.

Для обозначения топологии конкретного МСП используется специальная нотация. Например, МСП 784-13-10 обозначает, что НС имеет 784 нейрона на входном слое (соответствует числу пикселей изображения), 13 нейронов скрытого слоя и 10 нейронов выходного слоя, что соответствует количеству распознаваемых классов цифр (0-9).

В табл. 2 приведены результаты сравнения различных МСП с моделью LST-1 по числу параметров и достигаемой точности. Для моделей LST-1-8 и LST-1-28 указаны максимальные значения точности, полученные в серии из 10 независимых экспериментов обучения с различной начальной инициализацией весов (см. табл. 1).

Таблица 2
 Сравнение модели LST-1 с различными архитектурами МСП

Table 2
 Comparison of the LST-1 model with various MLP architectures

Авторы <i>Authors</i>	Модель <i>Model</i>	Число параметров <i>Number of parameters</i>	Точность, % <i>Accuracy, %</i>
[данная работа]	LST-1-8 (предлагаемая)	1114	96,53
Kwon, et al. [6]	МСП 196-14-10	2908	94,03
[данная работа]	LST-1-28 (предлагаемая)	9474	98,37
Westby, et al. [7]	МСП 784-12-10	9550	93,25
Huynh [13]	МСП 784-40-40-40-10	34 960	97,20
Huynh [13]	МСП 784-126-126-10	115 920	98,16
Medus, et al. [2]	МСП 784-600-600-10	891 610	98,63
Liang, et al. [12]	МСП 784-2048-2048-2048-10	10 100 000	98,32

Среди компактных архитектур для аппаратной реализации распознавания рукописных цифр выделяется МСП 196-14-10 [6]. Особенность данной архитектуры заключается в предварительном снижении размерности входного изображения 28×28 пикселей с помощью слоя субдискретизации (англ. *max-pooling*), который формирует уменьшенное изображение размером 14×14 . Данное представление затем преобразуется в одномерный вектор длиной 196, который подается на вход двухслойной сети. Такой трюк с понижением размерности на входе позволяет значительно сократить число обучаемых параметров (до 2,9 тыс.), сохранив при этом высокую точность 94 %. Для сравнения однослойный перцептрон, обрабатывающий исходное изображение без предварительного понижения размерности (784 входных пикселя) и формирующий 10 выходных вероятностей, содержит 7,9 тыс. параметров и достигает точности 92,4 % [14]. Однако представленная в настоящем исследовании модель LST-1-8 превосходит

архитектуру [6] по двум основным показателям: точности (превышение на 2,5 %) и компактности (количество параметров меньше в 2,5 раза). МСП, предложенные в работе [13], позволяют достичь точности от 97,2 до 98,16 % с помощью небольших скрытых полносвязных слоев. В работе [12] применяется подход с увеличением числа параметров скрытых слоев. Это дает возможность добиться высокой точности (98,32 %) и требует более 10 млн параметров, что превышает все рассмотренные архитектуры.

Среди компактных архитектур, обрабатывающих все 784 пикселя исходного изображения, следует выделить МСП 784-12-10 [7], который имеет 9550 параметров и достигает точности 93,25 %. Однако предлагаемая в настоящей работе модель LST-1-28 имеет практически такое же число параметров и точность, превосходящую на 5,12 %.

Предлагаемая модель LST-1-28 показывает точность, сопоставимую с более сложными архитектурами МСП 784-126-126-10 [13] и МСП 784-2048-2048-2048-10 [12], которые, однако, имеют одна в 12, а вторая в 1060 раз больше параметров, чем LST-1-28. МСП 784-600-600-10, представленный в работе [2], имеет точность, на 0,26 % превышающую точность модели LST-1-28, но при этом число его параметров в 94 раза больше.

Квантование параметров модели LST-1 и реализация с фиксированной запятой. При аппаратной реализации НС важно предотвратить переполнение разрядной сетки. Поскольку модель LST-1, по сути, состоит из трех полносвязных слоев, достаточно рассмотреть вопрос определения разрядности данных на примере одного полносвязного слоя. Определение разрядности выполнялось по принципу «наихудшего случая». Известно, что на входе каждого слоя модели LST-1 данные находятся в диапазоне $[-1, 1]$. Наибольшие значения на выходе полносвязного слоя будут получаться в том случае, если на вход поступит вектор, каждая компонента которого по модулю будет равна единице, а знаки будут совпадать со знаками строки матрицы весов \mathbf{W} . Более формально наибольшие возможные значения на выходе полносвязного слоя (до функции активации) можно оценить по формуле

$$\mathbf{Y}_{\max}(\mathbf{W}, \mathbf{b}) = \mathbf{W} \cdot \text{sign}(\mathbf{W}^T) + \mathbf{b}. \quad (6)$$

На рис. 11 приведены значения \mathbf{Y}_{\max} для трех полносвязных слоев обученной сети LST-1-28.

На выходе слоя, выполняющего обработку по строкам, максимальное значение равняется 15,02. Аналогичное значение для слоя, выполняющего обработку по столбцам, равно 11,9. Максимально возможное значение для выходного слоя равняется 126,3. Таким образом, можно сделать вывод, что для представления целой части данных потребуется восемь разрядов (один разряд на знак и оставшиеся семь для того, чтобы представить максимально возможное число – 126). Число дробных разрядов для представления данных выбиралось равным семи. Следует заметить, что восемь разрядов для представления целой части данных является завышенной оценкой. Дополнительные проверки путем симуляции модели LST-1 в арифметике с фиксированной запятой (ФЗ) показали, что уже при использовании семи разрядов для представления целой части переполнения разрядной сетки не наступает. Таким образом, для внутреннего представления данных в IP-блоке использовался формат Q7.7. При переводе значений параметров модели из формата с плавающей запятой (ПЗ) в формат с ФЗ использовался метод округления к ближайшему меньшему числу, который позволяет минимизировать ошибку округления.

Для верификации IP-блока LST-1 была разработана его эталонная модель на языке Python с использованием библиотеки `fixedpoint`. Данная библиотека дает возможность моделировать вычислительные процессы в арифметике с ФЗ. Модель LST-1 с ФЗ позволяет получать доступ ко всем промежуточным результатам вычислений, что существенно сократило время отладки IP-блока на этапе RTL-проектирования. Кроме того, проведенное моделирование выявило, что основной причиной расхождений между исходной версией LST-1 с ПЗ и реализацией с ФЗ является использование аппроксимации функции гиперболического тангенса.

На рис. 12 представлены результаты работы моделей с ФЗ и ПЗ на этапе вычисления выхода полносвязного слоя, выполняющего обработку строк изображения. Можно видеть, что до применения активационной функции обе модели дают очень близкий результат. Однако после применения активационной функции расхождение между моделями существенно увеличивается.

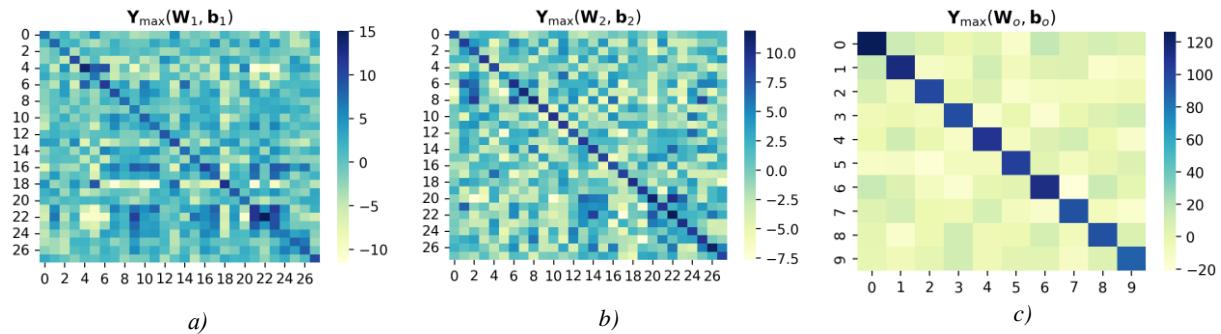


Рис. 11. Оценка максимального значения на выходе полносвязных слоев модели LST-1-28:
 а) слой обработки строк; б) слой обработки столбцов; в) классификационный слой

Fig. 11. Estimation of the maximum value at the output of fully connected layers of LST-1-28 model:
 a) row processing layer; b) column processing layer; c) classification layer

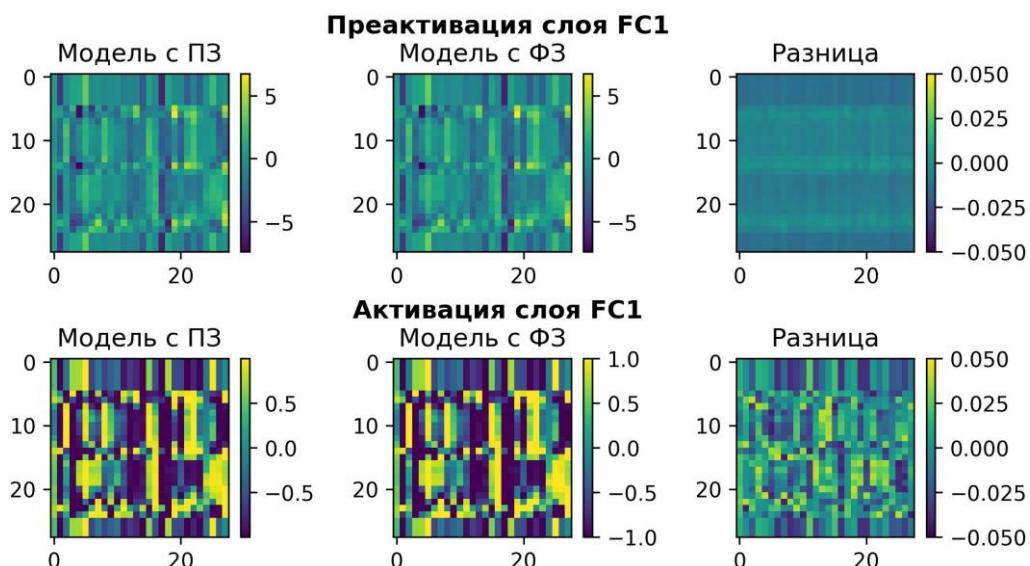


Рис. 12. Сравнение преактивации и активации в слоях FC1 моделей LST-1-28
 с плавающей и фиксированной запятой

Fig. 12. Comparison of preactivation and activation in FC1 layers of floating-point
 and fixed-point models LST-1-28

Таким образом, можно сделать вывод, что квантование параметров модели вносит меньший вклад в расхождение моделей с ФЗ и ПЗ, чем использование аппроксимированной функции активации.

Тестирование и анализ аппаратной реализации модели LST-1. Для реализации модели LST-1 была выбрана отладочная плата Zybo на базе ПЛИС фирмы Xilinx Zynq-7000. Платформа Zynq объединяет процессор ARM с программируемой логикой FPGA, обеспечивая гибкое и эффективное аппаратно-программное решение. Для упрощения разработки и тестирования на данной платформе используется дистрибутив Linux-PYNQ, который запускался на процессоре ARM. PYNQ позволяет взаимодействовать с аппаратными блоками ПЛИС, реализованными в виде IP-ядер, с помощью ноутбука Jupyter, что делает процесс разработки более удобным.

IP-блок HC LST-1 управляется через регистровый файл, подключенный по uP-интерфейсу (данный интерфейс разработан фирмой Analog Devices). При разработке IP-блоков для платформ Xilinx стандартным является использование AXI-интерфейса. Поэтому для подключения к процессорной системе Zynq внутри IP-блока LST-1 применяется преобразователь интерфейсов uP-AXI4-lite. Общая архитектура системы, использованной для тестирования аппаратной реализации HC LST-1, показана на рис. 13.

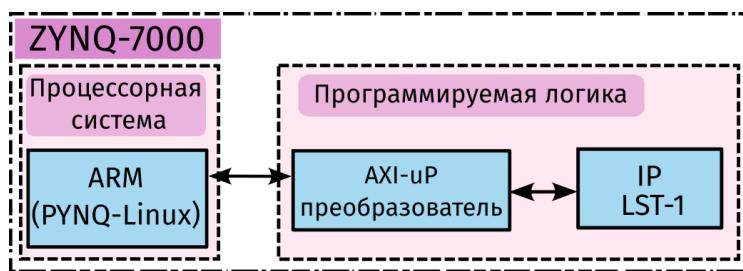


Рис. 13. Прототипирование нейронной сети LST-1 на ПЛИС

Fig. 13. Prototyping the LST-1 neural network on FPGA

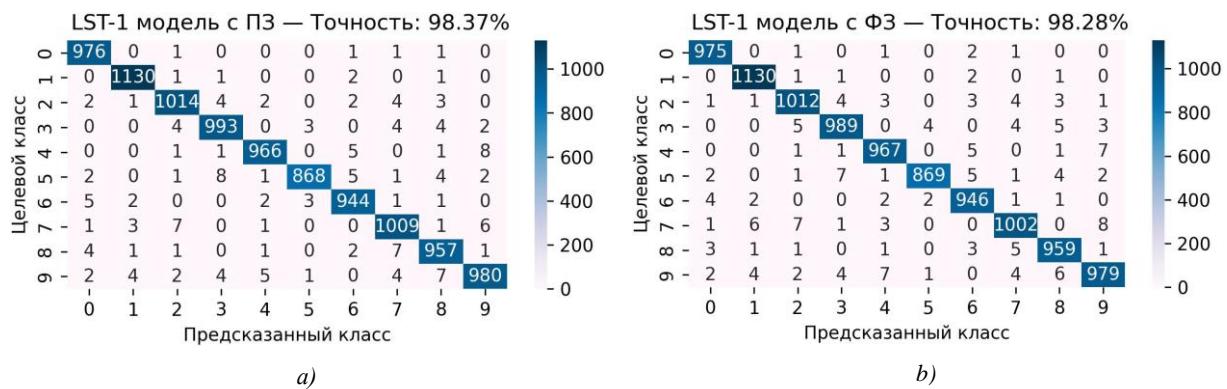


Рис. 14. Матрица ошибок нейронной сети LST-1-28: a) модель с плавающей запятой; b) модель с фиксированной запятой после квантования параметров

Fig. 14. Confusion matrix of the LST-1-28 : a) floating-point model; b) fixed-point model with quantized parameters

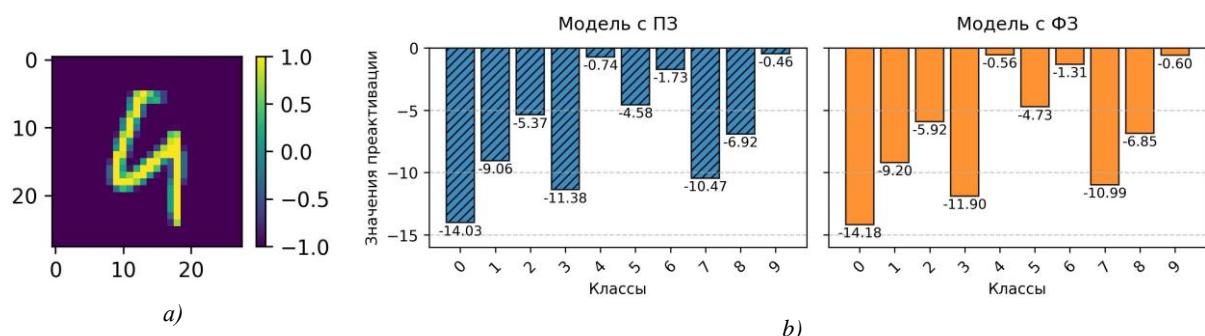


Рис. 15. Анализ работы модели LST-1-28: a) изображение цифры девять на входе нейронной сети; b) сравнение значений преактивации до применения функции softmax/argmax

Fig. 15. Analysis of the LST-1-28 model: a) image of a digit nine at the input of a neural network; b) comparison of preactivations before applying the softmax/argmax function

Для тестирования аппаратной реализации модели LST-1 на нее подавались 10 тыс. тестовых изображений базы MNIST. Для анализа полученных результатов выполнялось построение матрицы ошибок, которая количественно отображает распределение предсказаний модели относительно истинных классов объектов.

На рис. 14 показаны матрицы ошибок, полученные для модели с ПЗ, а также при реализации LST-1 на ПЛИС (данные матрица также совпадают с матрицей, полученной с помощью модели с ФЗ на основе Python). Общая точность модели LST-1 с весами, квантованными в формате Q7.7, составляет 98,28 %, что всего на 0,09 % отличается от точности модели с ПЗ. Следует заметить, что для некоторых классов точность предсказаний даже выросла. Так, исходная модель с ПЗ распознала правильно 944 изображения цифры шесть, в то время как модель с ФЗ правильно распознала 946 изображений цифры шесть.

Для того чтобы прояснить характер ошибок, допускаемых моделью с ФЗ, был проанализирован случай классификации тестового изображения 4823 базы MNIST (рис. 15, a), которое было правильно распознано моделью с ПЗ и ошибочно – моделью с ФЗ.

На рис. 15, b представлены значения преактивации выходного слоя сравниваемых моделей. В реализации с ПЗ максимальное значение (-0,46) соответствует верному классу 9 (истинная метка MNIST), однако близкое значение на классе 4 (-0,74) свидетельствует о высокой схожести цифр девять и четыре. В модели с ФЗ эффекты квантования параметров и аппроксимации гиперболического тангенса вызывают изменение выходных значений: минимальное значение (-0,56) соответствует ошибочному классу 4, тогда какциальному классу 9 соответствует следующее по величине значение (-0,60). Данный пример, с одной стороны, поясняет характер ошибок, возникающих в модели с ФЗ, а с другой – также позволяет понять причину, по которой в некоторых случаях модель с ФЗ (например, для класса 6) дала больше правильных прогнозов, чем модель с ПЗ. Имеется в виду, что ошибки квантования параметров модели вероятно скорректировали значения выходного слоя таким образом, что максимальные значения преактивации переместились на истинный класс.

Синтез IP-блока LST-1 в САПР Vivado 2024.2 показал, что для реализации НС требуется 1288 LUT-блоков и 1071 триггер. Общие аппаратные затраты представлены в табл. 3.

Таблица 3
 Аппаратные затраты на реализацию IP-блока LST-1

Table 3
Hardware resources for the LST-1 IP block implementation

Блок <i>Resource</i>	Использовано <i>Utilization</i>	Доступно <i>Available</i>	Соотношение, % <i>Utilization, %</i>
LUT	1288	17 600	7,32
LUTRAM	54	6000	0,90
FF	1071	35 200	3,04
BRAM	33,5	60	55,83
DSP	57	80	71,25

Разработанная архитектура НС требует использования лишь 55,83 % блочной памяти (BRAM) для хранения всех весовых коэффициентов и смещений, что указывает на пониженные требования к памяти модели LST-1. DSP-блоки, представленные в табл. 3, используются для реализации MAC-ядер внутри вычислительных элементов PE_{rc}/PE_{reco}. Максимальная тактовая частота IP-ядра 90 МГц. На обработку одного изображения IP-ядру требуется 3921 такт. Для сравнения – аппаратная реализация НС с архитектурой VGG16, представленная в работе [15] и имеющая точность на наборе MNIST 98,34 %, требует 108 тыс. тактов на обработку одного изображения.

Суммарная потребляемая мощность IP-ядра на чипе составляет 1,53 Вт. Температура зафиксирована на уровне 42,7 °C, что находится в допустимых пределах для стабильной работы устройства. Температурный зазор составляет 42,3 °C, что указывает на запас по тепловой стабильности и безопасное функционирование системы.

Заключение. В работе предложена концепция ОДРП для построения эффективных и компактных архитектур НС. Разработано семейство компактных моделей LST-1 с настраиваемой размерностью векторных представлений изображения (LST-1-8, LST-1-28 и др.). Достигнута высокая точность распознавания на наборе данных MNIST (98,37 % для LST-1-28 при 9,5 тыс. параметров и 96,53 % для LST-1-8 при 1,1 тыс. параметров), что дает более чем двухкратное превосходство по числу параметров относительно аналогов. Алгоритм вычисления модели LST-1 имеет регулярную структуру, что позволило получить ее эффективную реализацию на ПЛИС. Перспективным направлением дальнейших исследований является интеграция ОДРП в архитектуры НС с остаточными связями (англ. *residual networks, ResNet*), а также в архитектуры сверточных НС. В качестве еще одного направления исследований можно предложить возможность построения трехмерного обучаемого преобразования для обработки мультиспектральных изображений.

Вклад авторов. *Е. А. Кривальцевич* участвовал в разработке ОДРП, подготовке численных экспериментов, разработал и протестировал аппаратную реализацию модели LST-1, подготовил первоначальный вариант статьи; *М. И. Ващевич* предложил идею ОДРП, определил задачи исследования, разработал модель LST-1 с фиксированной запятой, корректировал текст статьи, принимал участие в подготовке графического материала, а также в интерпретации результатов экспериментов.

Список использованных источников

1. Park, J. FPGA based implementation of deep neural networks using on-chip memory only / J. Park, W. Sung // IEEE Intern. Conf. on Acoustics, Speech and Signal Processing (ICASSP), Shanghai, China, 20–25 March 2016. – Shanghai, 2016. – P. 1011–1015.
2. A novel systolic parallel hardware architecture for the FPGA acceleration of feedforward neural networks / L. D. Medus, T. Iakymchuk, J. V. Frances-Villora [et al.] // IEEE Access. – 2019. – Vol. 7. – P. 76084–76103.
3. Han, S. Deep compression: compressing deep neural networks with pruning, trained quantization and Huffman coding / S. Han, H. Mao, W. J. Dally // Intern. Conf. on Learning Representations (ICLR), San Juan, Puerto Rico, 2–4 May 2016. – San Juan, 2016. – P. 1–14.
4. Samragh, M. Customizing neural networks for efficient FPGA implementation / M. Samragh, M. Ghasemzadeh, F. Koushanfar // Annual Intern. Symp. on Field-Programmable Custom Computing Machines (FCCM), Napa, CA, USA, 30 Apr. – 2 May 2017. – Napa, 2017. – P. 85–92.
5. Usatyuk, V. Boosting DNN efficiency: replacing FC layers with graph embeddings for hardware acceleration / V. Usatyuk, S. Egorov // Intern. Conf. on Digital Signal Processing and its Applications (DSP), Moscow, Russia, 26–28 March 2025. – Moscow, 2025. – P. 1–6.
6. Kwon, J. Design of a low-area digit recognition accelerator using MNIST database / J. Kwon, S. Kim // JOIV: International Journal on Informatics Visualization. – 2022. – Vol. 6, no. 1. – P. 53–59.
7. FPGA acceleration on a multilayer perceptron neural network for digit recognition / I. Westby, X. Yang, T. Liu, H. Xu // The Journal of Supercomputing. – 2021. – Vol. 77, no. 12. – P. 14356–14373.
8. Vashkevich, M. Compact and efficient neural networks for image recognition based on learned 2D separable transform / M. Vashkevich, E. Krivalcevich // Intern. Conf. on Digital Signal Processing and its Applications (DSP), Moscow, Russia, 26–28 March 2025. – Moscow, 2025. – P. 1–5.
9. Старовойтов, В. В. Нормализация данных в машинном обучении / В. В. Старовойтов, Ю. И. Голуб // Информатика. – 2021. – Т. 18, № 3. – С. 83–96.
10. Loshchilov, I. SGDR: Stochastic Gradient Descent with Warm Restarts / I. Loshchilov, F. Hutter. – 2016. – URL: <https://arxiv.org/abs/1608.03983> (date of access: 01.08.2025).
11. Srivastava, N. Dropout: a simple way to prevent neural networks from overfitting / N. Srivastava, G. Hinton, A. Krizhevsky // The Journal of Machine Learning Research. – 2014. – Vol. 15, no. 1. – P. 1929–1958.
12. FP-BNN: Binarized neural network on FPGA / S. Liang, S. Yin, L. Liu [et al.] // Neurocomputing. – 2018. – Vol. 275. – P. 1072–1086.

13. Huynh, T. V. Deep neural network accelerator based on FPGA / T. V. Huynh // NAFOSTED Conf. on Information and Computer Science, Hanoi, Vietnam, 24–25 Nov. 2017. – Hanoi, 2017. – P. 254–257.
14. Кривальцевич, Е. А. Исследование аппаратной реализации нейронной сети прямого распространения для распознавания рукописных цифр на базе FPGA / Е. А. Кривальцевич, М. И. Ващекевич // Доклады БГУИР. – 2025. – Вып. 23, № 2. – С. 101–108.
15. Аппаратная реализация сверточной нейронной сети в ПЛИС на базе вычислений с фиксированной точкой / Р. А. Соловьев, А. Г. Кустов, В. С. Рухлов [и др.] // Известия Южного федерального университета. Технические науки. – 2017. – Т. 192, № 7. – С. 186–197.

References

1. Park J., Sung W. FPGA based implementation of deep neural networks using on-chip memory only. *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Shanghai, China, 20–25 March 2016*. Shanghai, 2016, pp. 1011–1015.
2. Medus L. D., Iakymchuk T., Frances-Villora J. V., Bataller-Mompean M., Rosado-Munoz A. A novel systolic parallel hardware architecture for the FPGA acceleration of feedforward neural networks. *IEEE Access*, 2019, vol. 7, pp. 76084–76103.
3. Han S., Mao H., Dally W. J. Deep compression: compressing deep neural networks with pruning, trained quantization and Huffman coding. *International Conference on Learning Representations (ICLR), San Juan, Puerto Rico, 2–4 May 2016*. San Juan, 2016, pp. 1–14.
4. Samragh M., Ghasemzadeh M., Koushanfar F. Customizing neural networks for efficient FPGA implementation. *Annual International Symposium on Field-Programmable Custom Computing Machines (FCCM), Napa, CA, USA, 30 April – 2 May 2017*. Napa, 2017, pp. 85–92.
5. Usatyuk V., Egorov S. Boosting DNN efficiency: replacing FC layers with graph embeddings for hardware acceleration. *International Conference on Digital Signal Processing and its Applications (DSP), Moscow, Russia, 26–28 March 2025*. Moscow, 2025, pp. 1–6.
6. Kwon J., Kim S. Design of a low-area digit recognition accelerator using MNIST database. *JOIV: International Journal on Informatics Visualization*, 2022, vol. 6, no. 1, pp. 53–59.
7. Westby I., Yang X., Liu T., Xu H. FPGA acceleration on a multilayer perceptron neural network for digit recognition. *The Journal of Supercomputing*, 2021, vol. 77, no. 12, pp. 14356–14373.
8. Vashkevich M., Krivalcevich E. Compact and efficient neural networks for image recognition based on learned 2D separable transform. *International Conference on Digital Signal Processing and its Applications (DSP), Moscow, Russia, 26–28 March 2025*. Moscow, 2025, pp. 1–5.
9. Starovoitov V. V., Golub Yu. I. *Data normalization in machine learning*. Informatika [Informatics], 2021, vol. 18, no. 3, pp. 83–96 (In Russ.).
10. Loshchilov I., Hutter F. *SGDR: Stochastic Gradient Descent with Warm Restarts*, 2016. Available at: <https://arxiv.org/abs/1608.03983> (accessed 01.08.2025).
11. Srivastava N., Hinton G., Krizhevsky A. Dropout: a simple way to prevent neural networks from overfitting. *The Journal of Machine Learning Research*, 2014, vol. 15, no. 1, pp. 1929–1958.
12. Liang S., Yin S., Liu L., Luk W., Wei S. FP-BNN: Binarized neural network on FPGA. *Neurocomputing*, 2018, vol. 275, pp. 1072–1086.
13. Huynh T. V. Deep neural network accelerator based on FPGA. *NAFOSTED Conference on Information and Computer Science, Hanoi, Vietnam, 24–25 November 2017*. Hanoi, 2017, pp. 254–257.
14. Krivalcevich E. A., Vashkevich M. I. *Investigation of hardware implementation of a feedforward neural network for handwritten digit recognition based on FPGA*. *Doklady BGU*, 2025, vol. 23, no. 2, pp. 101–108 (In Russ.).
15. Solov'yev R. A., Kustov A. G., Ruhlov V. S., Shchelokov A. N., Puzyrkov D. V. *Hardware implementation of a convolutional neural network in FPGA based on fixed point calculations*. *Izvestija Juzhnogo federal'nogo universiteta. Tehnicheskie nauki [Proceedings of Southern Federal University. Engineering Sciences]*, 2017, vol. 192, no. 7, pp. 186–197 (In Russ.).

Информация об авторах

Кривальцевич Егор Александрович, магистрант кафедры электронных вычислительных средств, Белорусский государственный университет информатики и радиоэлектроники.

Вашкевич Максим Иосифович, доктор технических наук, профессор кафедры электронных вычислительных средств, Белорусский государственный университет информатики и радиоэлектроники.
E-mail: vashkevich@bsuir.by

Information about the authors

Egor A. Krivalcevich, Undergraduate of Computer Engineering Department, Belarusian State University of Informatics and Radioelectronics.

Maxim I. Vashkevich, D. Sc. (Eng.), Prof. of Computer Engineering Department, Belarusian State University of Informatics and Radioelectronics.
E-mail: vashkevich@bsuir.by



УДК 004.89
DOI: 10.37661/1816-0301-2025-22-4-55-64

Оригинальная статья
Original Article

Технология трансляции естественно-языковых правил мерчандайзинга в цифровые планограммы

К. С. Курочка, Ю. Д. Ёвженко[✉]

Гомельский государственный технический
университет им. П. О. Сухого,
пр. Октября, 48, Гомель, 246029, Беларусь
✉E-mail: yuevzhenko@gmail.com

Аннотация

Цели. Целями исследования являются разработка и апробация подхода для автоматического преобразования правил выкладки товаров, сформулированных на естественном языке, в формализованные машино-читаемые инструкции для преодоления разрыва между бизнес-требованиями и их технической реализацией.

Методы. Предложен гибридный подход, в котором большая языковая модель выполняет функцию семантического транслятора, преобразуя пользовательский запрос в команду на специализированном предметно-ориентированном языке. Полученная команда затем обрабатывается детерминированным парсером на основе регулярных выражений для валидации и извлечения параметров. Для оценки качества трансляции использовалась метрика *BLEU* на специально созданном датасете из 200 пар «запрос – эталон». Эффективность подхода сравнивалась с базовым методом, основанным на правилах.

Результаты. Эксперимент показал высокую точность формализации запросов в сравнении с базовым подходом. Качественный анализ подтвердил способность системы корректно интерпретировать синонимы и сленг, извлекать неявно заданные параметры и отфильтровывать нерелевантные команды, что доказывает робастность предложенного подхода.

Заключение. Сделан вывод о жизнеспособности и практической значимости предложенного подхода для снижения трудоемкости и повышения эффективности процессов мерчандайзинга. Разработанная система представляет собой фундамент для создания нового поколения интеллектуальных инструментов управления торговым пространством.

Ключевые слова: большие языковые модели, парсер, метаязык, планограмма, семантический транслятор, естественный язык

Для цитирования. Курочка, К. С. Технология трансляции естественно-языковых правил мерчандайзинга в цифровые планограммы / К. С. Курочка, Ю. Д. Ёвженко // Информатика. – 2025. – Т. 22, № 4. – С. 55–64. – DOI: 10.37661/1816-0301-2025-22-4-55-64.

Конфликт интересов. Авторы заявляют об отсутствии конфликта интересов.

Поступила в редакцию | Received 26.09.2025
Подписана в печать | Accepted 17.10.2025
Опубликована | Published 30.12.2025

The technology of translating natural language merchandising rules into digital planograms

Konstantin S. Kurochka, Yury D. Youzhanka[✉]

*Gomel State Technical University named after P. O. Sukhoi,
av. Oktyabrya, 48, Gomel, 246029, Belarus*

[✉]E-mail: yuevzhenko@gmail.com

Abstract

Objectives. The aim of the research is to develop and test an approach for the automatic transformation of product layout rules, formulated in natural language, into formalized machine-readable instructions to bridge the gap between business requirements and their technical implementation.

Methods. A hybrid approach is proposed in which a large language model performs the function of a semantic translator, converting a user query into a command in a specialized domain-specific language. The resulting command is then processed by a deterministic parser based on regular expressions for validation and parameter extraction. The BLEU metric was used to evaluate the quality of the translation on a specially created dataset of 200 «query-reference» pairs. The effectiveness of the approach was compared with a baseline rule-based method.

Results. The experiment showed high accuracy in the formalization of queries compared to the baseline approach. A qualitative analysis confirmed the system's ability to interpret correctly synonyms and slang, extract implicitly specified parameters, and filter out irrelevant commands, which proves the robustness of the proposed approach.

Conclusion. The conclusion is made about the viability and practical significance of the proposed approach for reducing labor intensity and increasing the efficiency of merchandising processes. The developed system represents a foundation for creating a new generation of intelligent tools for managing retail space.

Keywords: large language models, parser, meta-language, planogram, semantic translator, natural language

For citation. Kurochka K. S., Youzhanka Y. D. *The technology of translating natural language merchandising rules into digital planograms*. Informatika [Informatics], 2025, vol. 22, no. 4, pp. 55–64 (In Russ.). DOI: 10.37661/1816-0301-2025-22-4-55-64.

Conflict of interest. The authors declare of no conflict of interest.

Введение. Планограмма, по сути, представляет собой визуальную схему или диаграмму, детально регламентирующую размещение конкретных товаров на торговом оборудовании – стеллажах, полках, витринах или холодильниках в розничной точке продаж [1]. Основная цель создания планограмм заключается в оптимизации использования торгового пространства, максимизации продаж и прибыли, улучшении покупательского опыта и обеспечении соответствия выкладки общим стратегиям категории и бренда [2]. Процесс формирования этих схем, известный как планограммирование, исторически развивался и на сегодняшний день включает в себя различные подходы, отличающиеся степенью автоматизации, используемыми инструментами и объемом анализируемых данных.

На заре развития розничной торговли и мерчандайзинга составление планограмм зачастую носило интуитивный и ручной характер. Специалисты по выкладке или категорийные менеджеры могли использовать простые эскизы от руки, масштабные чертежи на бумаге или базовые таблицы в электронных редакторах, таких как *Excel*, для схематичного отображения расположения товаров. При этом сотрудники должны были опираться на свой опыт и интуицию, а также выступать в роли контролеров, обеспечивающих соответствие выкладки многочисленным требованиям. Требования делились на две большие группы: во-первых, обязательные к исполнению государственные нормативно-правовые акты, регулирующие правила соседства и условия продажи отдельных товарных групп^{1,2}, и, во-вторых, внутренние корпоративные регламенты, брендбуки

¹О мерах реализации товаров, произведенных в Республике Беларусь : постановление Совета Министров Респ. Беларусь от 4 февр. 2015 г. № 72. – URL: <https://etalonline.by/document/?regnum=C21500072> (дата обращения: 17.07.2025).

²О государственном регулировании производства, оборота и потребления табачного сырья и табачных изделий : Декрет Президента Респ. Беларусь от 17 дек. 2002 г. № 28. – URL: <https://etalonline.by/document/?regnum=Pd0200028> (дата обращения: 17.07.2025).

и коммерческие условия договоров с поставщиками. Вся ответственность за интерпретацию, запоминание и ручное применение этого массива неструктурированной информации полностью ложилась на сотрудника. В результате такой подход страдал от ряда системных недостатков: высокой вероятности ошибок и несоответствия нормативным требованиям, колоссальной трудоемкости, сложности внесения оперативных изменений, низкой точности визуализации, трудностей в тиражировании на сеть магазинов и, главное, отсутствия системного анализа объективных данных о продажах и прибыльности каждого товара в привязке к занимаемому им месту.

С развитием информационных технологий появились специализированные программные решения для планограммирования. Эти программные комплексы предлагают значительно более совершенные инструменты для создания, редактирования и анализа планограмм. Они позволяют создавать точные двумерные и даже трехмерные визуализации торгового оборудования с размещенными на нем товарами, используя базы данных с изображениями упаковок и их габаритами. Важным преимуществом такого программного обеспечения является возможность интеграции с учетными системами ритейлера, что позволяет привязывать к каждой товарной позиции (*SKU*) на планограмме актуальные данные о продажах, прибыли, оборачиваемости, остатках и других ключевых показателях эффективности (*KPI*). Это дает возможность категорийным менеджерам принимать более обоснованные решения о размещении товаров, основываясь не только на визуальной привлекательности, но и на конкретных цифрах. Программное обеспечение также облегчает процесс тиражирования утвержденных планограмм по сети магазинов и контроль за их исполнением.

Несмотря на значительный прогресс, связанный с использованием специализированного программного обеспечения, процесс составления планограмм во многом остается зависимым от человеческого фактора и требует значительных временных затрат. Специалисту необходимо не только владеть инструментарием программы, но и уметь интерпретировать большие объемы данных, учитывать множество факторов и ограничений. К таким факторам относятся как количественные показатели продаж, так и качественные аспекты: стратегические цели категорий, договорные обязательства с поставщиками (например, требования к доле полки или определенному месту размещения), правила соседства товаров (например, не размещать бытовую химию рядом с продуктами питания), принципы визуального мерчандайзинга (цветовые блоки, размещение товаров-лидеров на уровне глаз, учет потоков покупателей), а также физические ограничения самого торгового оборудования и конкретного магазина.

Особую сложность представляют интеграция и учет требований, изложенных в неструктурированных или полуструктурных источниках, таких как маркетинговые гайдлайны, брендбуки, условия промоакций или контрактные обязательства с поставщиками. Эти требования часто формулируются в текстовом виде и содержат специфические правила выкладки, которые сложно напрямую транслировать в параметры программного обеспечения для планограммирования. Специалисту приходится вручную анализировать эти документы и следить за соблюдением всех изложенных в них условий при формировании схемы выкладки. Это не только увеличивает трудоемкость и время создания планограммы, но и повышает риск ошибок или неполного учета всех релевантных требований, что может привести к неоптимальной выкладке, нарушению договоренностей или снижению эффективности маркетинговых инициатив.

Таким образом, возникает очевидный разрыв между высокоуровневыми бизнес-требованиями, часто выраженными на естественном языке, и низкоуровневой технической реализацией в специализированном программном обеспечении. Этот разрыв является ключевым барьером на пути к полной автоматизации и повышению эффективности процесса планограммирования. Для его преодоления необходимы новые подходы, способные «понимать» и формализовать неструктурированные текстовые инструкции [3]. В связи с этим целями настоящей работы являются создание и исследование подхода и архитектуры системы для автоматизации формирования планограмм на основе семантического анализа правил выкладки, заданных на естественном языке.

Методы. В основе предлагаемого решения лежит концепция разделения задач между стохастической, семантически мощной системой, в качестве которой выступает большая языковая мо-

дель (large language model, *LLM*), и детерминированной, логически строгой системой (обработчик инструкций на основе регулярных выражений). Авторы рассматривают большую языковую модель не как конечный генератор планограммы, а как семантический транслятор. Ее ключевая задача – преобразовать запрос пользователя, сформулированный на гибком и потенциально неоднозначном естественном языке, в инструкцию на строго формализованном, однозначном и машиночитаемом метаязыке.

Такой подход позволяет использовать сильные стороны *LLM* – понимание контекста, извлечение сущностей и их взаимосвязей [4, 5] – и одновременно нивелировать ее слабые стороны, такие как недетерминированность, возможность «галлюцинаций» и генерации синтаксически некорректных или логически противоречивых данных.

Для реализации предложенной концепции был разработан специализированный метаязык команд – предметно-ориентированный язык, созданный специально для описания операций с планограммами. Ключевой особенностью такого языка является его строгий и однозначный характер. Каждая команда имеет императивную форму (например, РАЗМЕСТИ, УСТАНОВИ, СОРТИРОВКА), четко указывающую на выполняемое действие. Эта директивность подкрепляется строгим синтаксисом, который регламентирует точный порядок ключевых слов и их параметров. В свою очередь, каждый параметр типизированный: он должен соответствовать определенному типу данных, такому как целое число или строка, и может иметь дополнительные ограничения, например обязательность или принадлежность к допустимому диапазону значений.

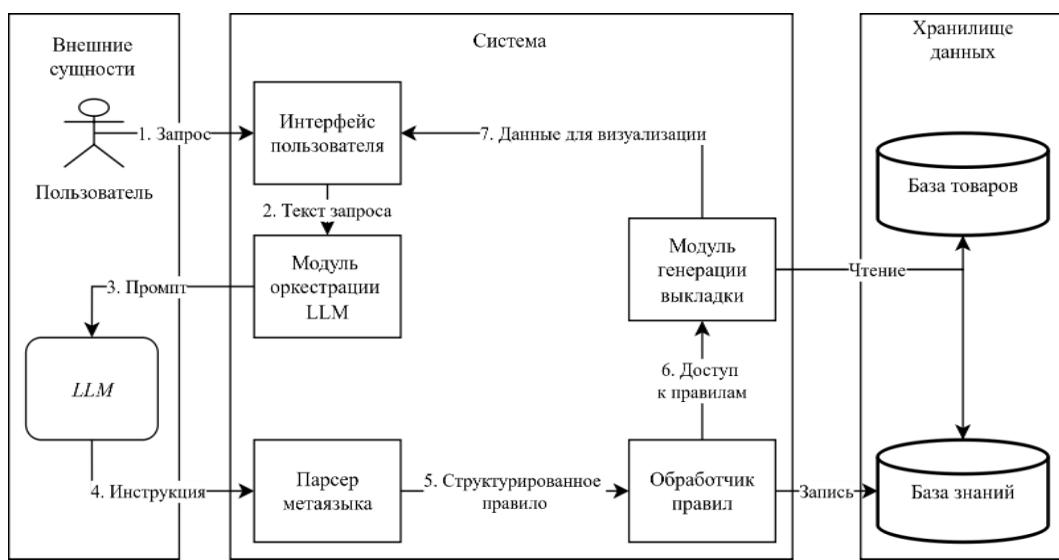
Примером команды метаязыка является РАЗМЕСТИ КАТЕГОРИЮ. Согласно спецификации (представлена ниже) она требует обязательного указания названия категории и номера полки, а также позволяет опционально задать долю занимаемого места и ограничения по весу:

```
{
  "command": "РАЗМЕСТИ КАТЕГОРИЮ",
  "syntax": "РАЗМЕСТИ КАТЕГОРИЮ <название> НА ПОЛКЕ <номер> [ЗАНЯВ <%> МЕСТА] [ВЕС [ОТ <мин> КГ] [ДО <макс> КГ]]",
  "parameters": [
    {"name": "название_категории", "type": "string", "required": true},
    {"name": "номер_полки", "type": "integer", "required": true},
    {"name": "доля_процентов", "type": "float", "required": true},
    {"name": "минимальный_вес", "type": "float", "required": false, "min": 0},
    {"name": "максимальный_вес", "type": "float", "required": false, "min": 0}
  ]
}
```

Таким образом, *LLM* получает задачу не сгенерировать планограмму, а перевести фразу «Положи молочные продукты на первую полку, пусть они займут 70 % места» в команду «РАЗМЕСТИ КАТЕГОРИЮ 'Молочные продукты' НА ПОЛКЕ 1 ЗАНЯВ 70 % МЕСТА». Это делает результат работы *LLM* предсказуемым и верифицируемым.

Предлагается следующая архитектура системы, построенная по модульному принципу, который обеспечивает сквозную обработку запроса пользователя от ввода на естественном языке до финальной визуализации. Весь процесс можно разделить на несколько логических этапов, каждый из которых реализуется отдельными компонентами (рисунок).

Начальный этап взаимодействия с системой начинается в пользовательском интерфейсе, который предоставляет специалисту (мерчандайзеру, категорийному менеджеру) текстовый чат для ввода команд. Введенный запрос передается в модуль оркестрации *LLM*, выполняющий роль связующего звена с нейросетью. Данный компонент формирует расширенный промпт, который включает системную инструкцию и полную спецификацию метаязыка для «обучения» модели в рамках одного запроса (*few-shot learning*) [6], и непосредственно текст пользователя. Затем этот промпт обрабатывается внешней большой языковой моделью, которая возвращает строку, формализованную согласно синтаксису метаязыка.



Далее сгенерированная строка поступает в критически важный компонент – парсер метаязыка, первоочередная задача которого – валидация. С помощью набора скомпилированных регулярных выражений парсер проверяет соответствие строки одному из известных шаблонов команд, что гарантирует детерминированность и высокую производительность этого этапа. В случае успеха парсер извлекает из строки все параметры, преобразует их в требуемые типы данных (*int*, *float*) и проверяет на соответствие заданным ограничениям. В результате на выходе формируется гарантированно корректный структурированный объект, готовый к дальнейшей обработке. На заключительном этапе этот верифицированный объект передается в обработчик правил – хранилище, которое аккумулирует все правила выкладки. После обновления базы знаний активируется модуль генерации выкладки. Этот алгоритм, основываясь на всем наборе правил и данных о товарах, рассчитывает конкретное положение каждого объекта на полках. Финальный результат его работы отображается пользовательским интерфейсом в виде интерактивной трехмерной модели планограммы, замыкая таким образом цикл обработки запроса.

Рассмотрим сквозной процесс выполнения запроса на конкретном примере:

Шаг 1. Пользователь вводит в чат команду на естественном языке: «Выкладка брендов внутри товарных категорий происходит по возрастанию рейтинга брендов от меньшего к большему по ходу движения покупательского потока».

Шаг 2. Модуль оркестрации формирует промпт, включающий спецификацию метаязыка и запрос пользователя, и отправляет его в *LLM*.

Шаг 3. *LLM*, проанализировав запрос и предоставленную ей спецификацию команд, генерирует ответ в виде строки на метаязыке: «СОРТИРОВКА БРЕНДОВ 'Рейтинг бренда (по возрастанию)'».

Шаг 4. Стока поступает на вход парсера. Парсер находит совпадение со своим регулярным выражением для команды «СОРТИРОВКА БРЕНДОВ». С помощью именованных групп извлекаются значения параметров, которые приводятся к нужным типам. На выходе формируется структурированный объект:

```
{
    "command": "СОРТИРОВКА БРЕНДОВ",
    "parameters": {
        "тип_сортировки": "Рейтинг бренда (по возрастанию)"
    }
}
```

Шаг 5. Структурированный объект передается в обработчик правил, который обновляет глобальное правило сортировки. Теперь при любой последующей генерации планограммы модуль генерации выкладки будет применять этот принцип: внутри каждой категории товары будут сгруппированы по брендам, а сами бренды будут расположены в порядке возрастания их рейтинга. Отображенный результат позволит пользователю увидеть, что бренды с низким рейтингом находятся в начале блока, а с высоким – в конце.

Для оценки преимуществ предложенного гибридного подхода было проведено сравнение с базовым методом, представляющим собой детерминированный парсер на основе набора регулярных выражений, разработанных для извлечения параметров непосредственно из текста запроса, минуя *LLM*. В отличие от *LLM*-подхода данный метод не способен обрабатывать синонимы или неявные указания, если они не были заранее жестко запрограммированы.

Результаты эксперимента. Для оценки эффективности предложенной системы и, в частности, ключевого ее компонента – способности *LLM* к качественной формализации естественноязыковых запросов – был проведен вычислительный эксперимент. Целью эксперимента являлась количественная и качественная оценка точности преобразования неструктурированных пользовательских команд в строго типизированные инструкции на метаязыке. Для этого был сформирован валидационный набор данных, состоящий из 200 уникальных пар «запрос – эталон», отражающих широкий спектр возможных пользовательских сценариев. Набор данных, использованный в настоящем исследовании, доступен по запросу у авторов. Следует отметить, что данный датасет, созданный для доказательства концепции (*Proof-of-Concept*), является относительно небольшим, что обусловлено высокой трудоемкостью ручной разметки. Набор был спроектирован таким образом, чтобы можно было проверить систему на устойчивость к различным лингвистическим конструкциям: от прямых и однозначных команд («Установи промежуток между товарами в 1 см») до более разговорных форм с использованием сленга и синонимов («На полке 1: омывайка (40 %/до 4 кг), антифриз (60 %/от 1 кг)»), а также запросов, требующих логического вывода («На полке 11 размести три категории поровну...»). Кроме того, в датасет были включены заведомо нерелевантные и бессмысленные фразы («Какая сегодня погода?», «абракадабра»), чтобы оценить способность системы отфильтровывать нецелевой ввод.

В качестве оцениваемых *LLM* были выбраны три ведущие модели: *GPT-4* от *OpenAI*, *Gemini 2.5 Pro* от *Google* и *Qwen3* от *Alibaba*. Для количественной оценки качества трансляции была выбрана метрика *BLEU* (*Bilingual Evaluation Understudy*), традиционно применяемая для оценки качества машинного перевода. Использование *BLEU* в данном контексте оправдано, поскольку задача, по сути, является переводом с гибкого естественного языка на формальный язык команд. Метрика *BLEU* оценивает степень совпадения *n*-граммов (последовательностей слов) между сгенерированной моделью строкой и эталонной строкой, что позволяет получить объективную численную меру их близости. При этом авторы осознают, что *BLEU* не всегда отражает семантическую корректность. Результаты эксперимента по расчету среднего значения *BLEU* для каждой модели на всем тестовом наборе данных представлены в таблице. Для сравнения также приведена точность (*Accuracy*) – доля полностью корректно распознанных парсером команд.

Результаты оценки качества формализации запросов

Results of the evaluation of the quality requests formalization

Модель <i>Model</i>	Значение <i>BLEU</i> <i>BLEU Value</i>	Точность, % <i>Accuracy, %</i>	Время выполнения запроса, с <i>Time to complete the request, s</i>
<i>GPT-4</i>	0,8227	98,5	~ 2
<i>Gemini 2.5 Pro</i>	0,7315	100,0	
<i>Qwen3</i>	0,7051	100,0	
Базовый метод (запросы пользователя)	<i>N/A</i>	0,5	< 0,1
Базовый метод (эталоны)	<i>N/A</i>	100,0	

Полученные значения *BLEU*, превышающие 0,7 для всех моделей, свидетельствуют о высоком общем качестве формализации. Это означает, что в большинстве случаев сгенерированные командами строки были очень близки или полностью идентичны эталонным. Модель *GPT-4* показала несколько лучший результат, что может быть связано с особенностями ее архитектуры и наборами данных для обучения. Однако близость результатов говорит о том, что предложенный подход является робастным и зависит не от конкретной реализации *LLM*, а, скорее всего, от самой концепции семантического транслятора.

Высокие показатели точности распознавания транслированных команд подтверждают, что модели не просто генерируют синтаксически похожий текст, но и корректно извлекают семантику запроса. Для более объективной оценки в таблицу также добавлены результаты для базового метода, точность которого для необработанных пользовательских запросов составила всего 0,5 %. Это подчеркивает неспособность жестких правил справляться с лингвистической вариативностью, синонимами и сленгом в отличие от предлагаемого *LLM*-подхода.

Более глубокий анализ результатов выходит за рамки единой метрики. Качественный анализ работы моделей на тестовом датасете позволил выявить несколько ключевых закономерностей. Во-первых, все модели продемонстрировали превосходную способность к синонимии и распознаванию сленга. Например, запросы, содержащие слова «охлаждайка» или «антифриз», были корректно преобразованы в инструкцию с использованием каноничного термина «охлаждающие жидкости». Это является критически важным преимуществом использования *LLM* по сравнению с системами, основанными на жестких правилах и ключевых словах.

Во-вторых, система успешно справлялась с извлечением неявно выраженных параметров. Так, запрос «На полке 1 размести охлаждайку и омывайку поровну» был верно интерпретирован как «ПОЛКА 1: охлаждающие жидкости (50 %), стеклоомыватели (50 %)», а инструкция «размести три категории поровну» привела к корректному распределению долей 33, 33 и 34 % для соблюдения суммарного ограничения в 100 %. Это демонстрирует не простое извлечение сущностей, а способность модели к базовым логическим операциям в контексте поставленной задачи [6].

В-третьих, важным результатом является способность системы к самофильтрации. Все нерелевантные запросы, такие как «Как дела?» или «Закажи мне пиццу», были правильно классифицированы и привели к генерации специальной команды НЕОПОЗНАННАЯ КОМАНДА, что предотвращает попадание «мусорных» данных в правила. Это подтверждает надежность системы в реальных условиях эксплуатации, где пользовательский ввод не всегда является идеальным.

Таким образом, проведенный анализ результатов подтверждает эффективность и адекватность предложенного метода. Связка из *LLM*, выступающей в роли семантического транслятора, и строгого парсера метаязыка позволяет достичь высокой точности формализации правил выкладки, сохраняя при этом гибкость и удобство естественно-языкового интерфейса для пользователя. Это доказывает жизнеспособность архитектуры и открывает путь к созданию интуитивно понятных и мощных инструментов для автоматизации мерчандайзинга.

Анализ ошибок и ограничения. Несмотря на высокие общие показатели, качественный анализ выявил ряд систематических недостатков, определяющих текущие границы применимости предложенного подхода. Наиболее часто ошибки проявлялись при обработке сложных инструкций, содержащих несколько логически связанных или взаимоисключающих условий. В таких случаях модель успешно идентифицировала основное намерение, но не справлялась с декомпозицией и обработкой всех второстепенных ограничений. Так, например, запрос «На полке 2 остаток заполни 'тормозными жидкостями', а сортировку брендов установи по рейтингу.» либо транслировался частично: система генерировала команду на размещение категории «Тормозные жидкости», но полностью игнорировала инструкцию по сортировке, либо генерировалась команда «НЕОПОЗНАННАЯ КОМАНДА». Это демонстрирует ограничение в способности *LLM* к композиционному анализу – разложению сложной задачи на последовательность более простых.

Другая существенная трудность возникала, когда запрос содержал субъективные или неоднозначные понятия, не имеющие прямого эквивалента в формализованном метаязыке. Пытаясь интерпретировать такие команды, модель могла делать недопустимые предположения, приводящие

к семантическим ошибкам или «галлюцинациям». Например, на просьбу «На полке 1 размести охлаждающие жидкости рядом со стеклоомывателями.» система могла сгенерировать команду на размещение категорий последовательно, делая произвольный выбор их порядкового расположения. Такое поведение недопустимо в системах, требующих высокой точности и детерминизма.

Основной причиной указанных недостатков является опора исключительно на метод «обучения на нескольких примерах» (*few-shot learning*), который не всегда позволяет модели уловить все тонкости предметной области. Для устранения этих ошибок планируется работа в двух направлениях. В краткосрочной перспективе предполагается использование более сложных техник промпт-инжиниринга, например *chain-of-thought*, которые заставляют модель «рассуждать» перед выдачей ответа, что повышает точность при работе со сложной логикой. В долгосрочной перспективе ключевым решением является дообучение (*fine-tuning*) *open-source*-модели на расширенном датасете, включающем большое количество примеров сложных и неоднозначных запросов. Такой подход позволит адаптировать ее к специфике задачи и значительно снизить вероятность подобных ошибок.

Заключение. В ходе данного исследования была успешно разработана и апробирована система для автоматизации формирования планограмм на основе семантического анализа текстовых данных. Ключевым результатом работы стало подтверждение эффективности предложенной гибридной архитектуры, в которой большая языковая модель выступает в роли семантического транслятора с естественного языка на формализованный метаязык. Такой подход позволил использовать сильные стороны *LLM* в понимании контекста и семантики, одновременно нивелировав риски недетерминированности за счет последующей обработки строго детерминированным парсером. Разработанный предметно-ориентированный язык выступил в качестве надежного контракта между компонентами системы, гарантируя, что в правила поступают только валидные и однозначные инструкции. Эффективность данного подхода была количественно подтверждена в ходе вычислительного эксперимента: значения метрики *BLEU* свидетельствуют о высокой точности трансляции широкого спектра пользовательских запросов, включая разговорные формы и команды с неявными параметрами.

Практическая значимость полученных результатов заключается в создании инструмента, способного изменить подходы к управлению торговым пространством. Внедрение подобной системы позволяет значительно снизить трудоемкость и временные затраты на создание и обновление планограмм, повысить скорость реакции ритейлера на маркетинговые изменения и минимизировать количество ошибок, обусловленных человеческим фактором. Это, в свою очередь, позволит сделать процесс планографирования доступным для более широкого круга сотрудников.

Следует отметить и ограничения данного исследования. Во-первых, валидация проводилась на статичном датасете и не включала тестирование на реальных производственных нагрузках с участием конечных пользователей. Во-вторых, как показал анализ ошибок, текущая реализация может испытывать трудности с интерпретацией неоднозначных или логически сложных правил.

Настоящая работа открывает широкие перспективы для дальнейшего развития. Следующим логическим шагом является переход от реактивного инструмента, выполняющего команды, к проактивной интеллектуальной системе, способной самостоятельно предлагать оптимальные решения. Этого можно достичь путем глубокой интеграции с аналитическими данными о продажах, что позволит системе генерировать планограммы на основе *ABC*-анализа и других метрик эффективности. Дальнейшее развитие может включать создание полноценного диалогового ассистента, способного не только выполнять задачи, но и объяснять свои решения; расширение до мультимодальной системы, анализирующей визуальную информацию с фотографий реальных полок [7], и проведение полноценного тестирования с участием категорийных менеджеров для оценки эффективности и удобства системы в реальных практических условиях. Переход на дообученные (*fine-tuned*) *open-source*-модели позволит повысить точность, снизить операционные издержки и обеспечить полную конфиденциальность данных.

При внедрении подобных систем важно учитывать и этические аспекты. Существует риск, что *LLM*, обученные на огромных массивах интернет-данных, могут обладать скрытыми

предубеждениями, например, отдавая предпочтения более известным брендам при неоднозначных формулировках. Это требует дополнительного контроля и, возможно, дообучения моделей на специфических корпоративных данных для минимизации подобных рисков и обеспечения беспристрастной выкладки.

Таким образом, представленная работа закладывает фундамент для создания нового поколения интеллектуальных систем управления торговым пространством, делая процесс мерчандайзинга более гибким, быстрым и основанным на данных.

Вклад авторов. К. С. Курочка сформулировал цель и идею исследования, Ю. Д. Ёвженко разработал программное обеспечение и провел его тестирование, оформил текст рукописи. Оба автора участвовали в написании и редактировании текста рукописи.

Список использованных источников

1. Швецова, А. М. Разработка планограммы как эффективный инструмент маркетинга / А. М. Швецова, С. Н. Хрипунов // Молодые ученые – развитию Национальной технологической инициативы (ПОИСК-2018) : материалы межвузовской (с международным участием) молодежной науч.-техн. конф., 24–27 апр. 2018 г. – Иваново : ИВГПУ, 2018. – С. 129–130.
2. Shelf space reallocation for out of stock reduction / E. Frontoni, F. Marinelli, R. Rosetti, P. Zingaretti // Computers & Industrial Engineering. – 2017. – Vol. 106. – P. 32–40. – DOI: 10.1016/j.cie.2017.01.006.
3. Курочка, К. С. Применение генеративно-ассоциативных нейросетевых моделей для создания планограмм / К. С. Курочка, Ю. С. Башаримов // Актуальные вопросы экономической науки в XXI веке : материалы Междунар. науч.-практ. конф. – X чтения, посвященные памяти известного белорусского и российского ученого-экономиста Михаила Вениаминовича Научителя, Гомель, 24 окт. 2024 г. / М-во образования Республики Беларусь, Гомельский гос. ун-т им. Ф. Скорины ; редкол.: А. К. Костенко [и др.]. – Гомель, 2024. – С. 139–141.
4. Jurafsky, D. Speech and Language Processing / D. Jurafsky, J. H. Martin. – 3rd ed. draft. – Stanford University, 2023. – 613 p.
5. Artificial intelligence in supply chain management: A systematic literature review / R. Toorajipour, V. Sohrabpour, A. Nazarpour [et al.] // Journal of Business Research. – 2021. – Vol. 122. – P. 502–517. – DOI: 10.1016/j.jbusres.2020.09.016.
6. Language models are few-shot learners / T. B. Brown, B. Mann, N. Ryder [et al.] // Advances in Neural Information Processing Systems (NeurIPS). – 2020. – Vol. 33. – P. 1877–1901.
7. Масалитина, Н. Н. Математическая модель принятия решений при лечении остеохондроза поясничного отдела позвоночника / Н. Н. Масалитина, К. С. Курочка, Е. Л. Цитко // Информатика. – 2019. – Т. 16, № 1. – С. 24–35.

References

1. Shvetsova A. M., Khripunov S. N. *Development of a planogram as an effective marketing tool*. Molodye uchenye – razvitiyu Natsional'noi tekhnologicheskoi initsiativy (POISK-2018): materialy mezhvuzovskoj (s mezhdunarodnym uchastiem) molodezhnoj nauchno-tehnicheskoy konferencii, 24–27 aprelja 2018 g. [Young Scientists for the Development of the National Technology Initiative (POISK-2018): Proceedings of the Inter-university (with International Participation) Youth Scientific and Technical Conference, 24–27 April 2018]. Ivanovo, Ivanovskij gosudarstvennyj politehnicheskij universitet, 2018, pp. 129–130 (In Russ.).
2. Frontoni E., Marinelli F., Rosetti R. Shelf space reallocation for out of stock reduction. *Computers & Industrial Engineering*, 2017, vol. 106, pp. 32–40. DOI: 10.1016/j.cie.2017.02.004.
3. Kurochka K. S., Basharimov Yu. S. *Application of generative-associative neural network models for creating planograms*. Aktual'nye voprosy ekonomicheskoi nauki v XXI veke : materialy Mezhdunarodnoj nauchno-prakticheskoy konferencii – X chteniiia, posviashchennye pamiati izvestnogo belorusskogo i rossiiskogo uchenogo-ekonomista Mikhaila Veniaminovicha Nauchitelia, Gomel', 24 oktyabrya 2024 g. [Topical Issues of Economic Science in the XXI Century: Proceedings of the International Scientific and Practical Conference – X Readings Dedicated to the Memory of the Famous Belarusian and Russian Economist Mikhail Veniaminovich Nauchitel, Gomel, 24 October 2024]. Gomel, 2024, pp. 139–141 (In Russ.).
4. Jurafsky D., Martin J. H. *Speech and Language Processing*, 3rd ed. draft. Stanford University, 2023, 613 p.

5. Toorajipour R., Sohrabpour V., Nazarpour A., Oghazi P., Fischl M. Artificial intelligence in supply chain management: A systematic literature review. *Journal of Business Research*, 2021, vol. 122, pp. 502–517. DOI: 10.1016/j.jbusres.2020.09.009.
6. Brown T. B., Mann B., Ryder N., Subbiah M., Kaplan J. D., Amodei D. Language models are few-shot learners. *Advances in Neural Information Processing Systems (NeurIPS)*, 2020, vol. 33, pp. 1877–1901.
7. Masalitina N. N., Kurocha K. S., Tsitko E. L. *Mathematical model of decision making in the treatment of osteochondrosis of the lumbar spine*. Informatika [Informatics], 2019, vol. 16, no. 1, pp. 24–35 (In Russ.).

Информация об авторах

Курочка Константин Сергеевич, кандидат технических наук, доцент, Гомельский государственный технический университет им. П. О. Сухого.

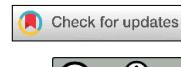
Ёвженко Юрий Дмитриевич, магистрант, Гомельский государственный технический университет им. П. О. Сухого.

E-mail: yuevzhenko@gmail.com

Information about the authors

Konstantin S. Kurochka, Ph. D. (Eng.), Assoc. Prof., Gomel State Technical University named after P. O. Sukhoi.
Yury D. Youzhanka, Master's Student, Gomel State Technical University named after P. O. Sukhoi.
E-mail: yuevzhenko@gmail.com

ЗАЩИТА ИНФОРМАЦИИ И НАДЕЖНОСТЬ СИСТЕМ INFORMATION PROTECTION AND SYSTEM RELIABILITY



УДК 004.312
DOI: 10.37661/1816-0301-2025-22-4-65-81

Оригинальная статья
Original Article

Генерирование детерминированных идентификаторов и случайных чисел на основе схемы конфигурируемого кольцевого осциллятора

А. А. Иванюк[✉], Л. А. Бурко

Белорусский государственный университет
информатики и радиоэлектроники,
ул. П. Бровки, 6, 220013, Минск, Беларусь
[✉]E-mail: ivaniuk@bsuir.by

Аннотация

Цели. Целью работы является рассмотрение особенностей функционирования цифровой схемы, анализирующей частоту выходного сигнала конфигурируемого кольцевого осциллятора в пределах фиксированного окна измерения.

Методы. Используются методы синтеза и анализа цифровых устройств, в том числе на программируемых логических интегральных схемах (ПЛИС), основы цифровой схемотехники, методы анализа случайных нормально распределенных величин.

Результаты. Разработана цифровая схема регистрации периода конфигурируемого кольцевого осциллятора в зависимости от временного окна измерения и значения его конфигурации. Проведены экспериментальные исследования периодов вырабатываемых сигналов при реализации разработанной схемы на программируемых логических интегральных схемах FPGA Xilinx ZYNQ 7000. Показано, что при многократном повторении измерения периода для регистрирующего счетчика можно выделить три группы разрядов: группу G_2 стабильных разрядов, значения которых остаются неизменными на протяжении всех измерений; группу G_1 слабо стабильных разрядов, искажения которых незначительны, и группу G_0 сильно нестабильных разрядов, вероятность искажения которых от измерения к измерению близка к 1. Было выдвинуто предположение, что группа разрядов G_0 представляет собой оцифрованные значения шумовой составляющей значения измеряемого периода. Предполагается, что в силу наличия многих независимых компонентов схемы конфигурируемого кольцевого осциллятора и цифрового регистратора, девиаций питающего напряжения, температуры кристалла и окружающей среды, ошибок квантования и др. данная шумовая составляющая нормально распределена. Аналитически было доказано, что нормально распределенная величина, квантованная многоразрядными двоичными числами, при определенных значениях математического ожидания μ и среднеквадратического отклонения σ порождает только две группы – G_2 и G_0 . Доказано, что вероятность появления единичного символа на всех разрядах группы G_0 близка к 0,5, а размерность группы можно оценить как $3 + \lfloor \log_2 \sigma \rfloor$. Разряды группы G_1 можно привести к группе G_2 различными

способами, в том числе методом максимального правдоподобия либо нормализацией значения каждого измерения до теоретически обоснованного разделения на группы G_2 и G_0 . Значения разрядов группы G_2 можно интерпретировать как детерминированный ответ на запрос, представляющий собой конфигурацию схемы кольцевого осциллятора в заданном окне измерения, формируя новый тип многоразрядных физически неклонируемых функций, обладающих высокой стабильностью. В свою очередь, разряды G_0 могут быть использованы как однобитные источники случайных величин, распределение которых близко к равномерному, формируя основу для построения генераторов случайных чисел.

Заключение. Полученные результаты могут быть применены во встроенных средствах обеспечения неклонируемой идентификации цифровых систем и генерации случайных данных. Применение синхронного двоичного счетчика в качестве схемы регистратора частоты конфигурируемого кольцевого осциллятора открывает новые возможности для построения многоразрядных схем физически неклонируемых функций с улучшенными показателями стабильности, уникальности и случайности.

Ключевые слова: физическая криптография, физически неклонируемые функции, конфигурируемый кольцевой осциллятор, стабильная идентификация, генерация случайных данных

Благодарности. Авторы выражают искреннюю благодарность резиденту ПВТ компании «Инженерный Центр Ядро», которая является одним из центров разработки YADRO, за предоставленное оборудование для проведения экспериментов в рамках работы совместной учебной лаборатории с Белорусским государственным университетом информатики и радиоэлектроники.

Для цитирования. Иванюк, А. А. Генерирование детерминированных идентификаторов и случайных чисел на основе схемы конфигурируемого кольцевого осциллятора / А. А. Иванюк, Л. А. Бурко // Информатика. – 2025. – Т. 22, № 4. – С. 65–81. – DOI: 10.37661/1816-0301-2025-22-4-65-81.

Конфликт интересов. Авторы заявляют об отсутствии конфликта интересов.

Поступила в редакцию | Received 22.09.2025

Подписана в печать | Accepted 13.10.2025

Опубликована | Published 30.12.2025

Generation of deterministic identifiers and random numbers using a configurable ring oscillator circuit

Alexander A. Ivaniuk[✉], Liana A. Burko

Belarusian State University of
Informatics and Radioelectronics,
st. P. Brovki, 6, Minsk, 220013, Belarus
✉E-mail: ivaniuk@bsuir.by

Abstract

Objectives. The purpose of the study is to examine the operational characteristics of a digital circuit designed to analyze the output signal frequency of a configurable ring oscillator within a fixed measurement window.

Methods. Methods of digital device synthesis and analysis were employed, including implementation on programmable logic integrated circuits (FPGAs), fundamentals of digital circuit design, and methods for analyzing normally distributed random variables.

Results. A digital circuit for recording the period of a configurable ring oscillator, depending on the measurement time window and the value of its configuration has been developed. Experimental studies of the generated signal periods were conducted using Xilinx ZYNQ 7000 series FPGAs. It was demonstrated that upon repeated period measurements the bits of the recording counter can be categorized into three groups: group G_2 : stable bits retaining constant values across all measurements; group G_1 : weakly stable bits exhibiting minor distortions; group G_0 : strongly unstable bits with a distortion probability approaching 1 between measurements. It was hypothesized that group G_0 represents the digitized noise component of the measured period value. Due to numerous independent factors (components within the configurable ring oscillator and digital recorder, supply voltage deviations, die and ambient temperature variations, quantization errors, etc.), it is assumed that this noise component follows a normal

distribution. Analytical proof established that a normally distributed variable, quantized using multi-bit binary numbers under specific values of mathematical expectation μ and standard deviation σ , generates only two groups: G_2 and G_0 . It was proven that the probability of a '1' appearing in any bit of group G_0 approaches 0,5, and the group size can be estimated as $3 + \lfloor \log_2 \sigma \rfloor$. The bits of group G_1 can be converted to group G_2 using various methods, such as maximum likelihood estimation or by normalizing each measurement value to a theoretically justified separation into G_2 and G_0 . The values of group G_2 bits can be interpreted as a response to a challenge defined by the ring oscillator circuit configuration and measurement window, forming a novel type of multi-bit Physical Unclonable Function (PUF) characterized by high stability. Conversely, the G_0 bits can serve as single-bit sources of random variables with near-uniform distribution, providing a foundation for building random number generators.

Conclusion. The obtained results can be utilized in embedded systems for providing unclonable identification of digital devices and for random data generation. The application of a synchronous binary counter as a frequency recording circuit for a configurable ring oscillator opens new avenues for designing multibit physical unclonable function architectures with enhanced performance metrics in terms of stability, uniqueness, and randomness.

Keywords: physical cryptography, physically unclonable functions, configurable ring oscillator, stable identification, random numbers generation

Acknowledgments. The authors sincerely thank "Engineering Center Yadro" (a resident of the Belarus High-Tech Park / HTP and an R&D center within the YADRO Group) for providing the experimental equipment used in this work through our joint educational laboratory with the Belarusian State University of Informatics and Radioelectronics.

For citation. Ivaniuk A. A., Burko L. A. *Generation of deterministic identifiers and random numbers using a configurable ring oscillator circuit*. Informatika [Informatics], 2025, vol. 22, no. 4, pp. 65–81 (In Russ.). DOI: 10.37661/1816-0301-2025-22-4-65-81.

Conflict of interest. The authors declare of no conflict of interest.

Введение. Цифровые физически неклонируемые функции (ФНФ) [1–3], составляющие основу современной аппаратной криптографии, лежат в основе методов уникальной идентификации, аутентификации, защиты от клонирования и генерации случайных чисел и по-прежнему находятся в фокусе внимания как проектировщиков, так и исследователей. Среди большого разнообразия видов схем ФНФ выделяются схемы типа «Арбитр» [4, 5] и ФНФ кольцевых осцилляторов (КО) [6], основные характеристики которых наряду с простотой реализации привлекают разработчиков защищенных цифровых устройств. Данные типы ФНФ основаны на сравнении задержек распространения сигналов по симметричным путям, уникально выбираемым по поступающему запросу CH . По результату сравнения формируется бинарный ответ R , а сами схемы реализуют функцию вида $R = PUF(CH)$. Все множество пар «запрос – ответ» является уникальным, неклонируемым, непредсказуемым и случайным, что характеризует конкретный воплощенный в кремнии экземпляр схемы ФНФ.

Развитием классической архитектуры ФНФ КО являются схемы, основанные на применении конфигурируемых кольцевых осцилляторов (ККО) [7–12]. Это дает возможность извлекать уникальность не только по результату сравнения частот, выбранных из реализованного множества КО, но и из каждой схемы КО путем конфигурации внутренних путей прохождения сигнала. Данная особенность позволяет применять ФНФ КО как для заказных СБИС, так и для ПЛИС, обеспечивая для последних приемлемые показатели стабильности, надежности, случайности и внутрикристальной уникальности [10]. Компактность схем ФНФ ККО наряду с расширенными возможностями извлечения уникальных признаков делает их перспективными для дальнейших исследований и реализаций на различных технологиях.

Одним из основных блоков цифровых схем ФНФ КО и ККО является блок измерения частоты выбранного осциллятора. Как правило, такие блоки представляют собой синхронные двоичные счетчики, регистрирующие число фронтов генерируемого КО сигнала в фиксированном временном окне MW . В классической схеме ФНФ КО по определенному запросу CH выбирается пара осцилляторов из множества имеющихся, для которых производятся измерения, а их результаты сравниваются между собой, формируя бинарный ответ R всей схемы ФНФ. В схеме ФНФ ККО,

помимо воспроизведения более компактной версии ФНФ КО [9], запросом могут выступать значение конфигурации CF самой схемы и размер временного окна измерения MW , при этом многоразрядный результат измерения числа фронтов является уникальным ответом. В данном случае схема реализует функцию вида $R = PUF(CF, MW)$. В силу многих факторов получаемое значение R является нестабильным, что проявляется при повторяющихся измерениях с одинаковыми же значениями CF и MW и негативно сказывается на таких характеристиках ФНФ, как надежность и стабильность [13, 14].

Настоящая статья посвящена исследованию поведения регистрирующего счетчика ФНФ КО с целью выделения группы стабильных разрядов, значения которых в дальнейшем могут быть применены для повышения надежности ответов ФНФ (например, для генерации детерминированных идентификаторов), и группы нестабильных (шумовых) разрядов, которые могут стать основой для генерации случайных чисел.

Схема конфигурируемого кольцевого осциллятора. Объектом исследования является схема, состоящая из ККО и регистратора количества передних фронтов, сгенерированных ККО в фиксированном временном окне измерения MW . Период вырабатываемого сигнала зависит от управляемых и неуправляемых (случайных) факторов. Управляемым фактором является конфигурация ККО, задаваемая значением на многоразрядной входнойшине CF . К неуправляемым факторам можно отнести уникальные задержки сигналов на структурных элементах ККО, зависящие от технологии изготовления, девиации уровня питающего напряжения, температуры кристалла и т. п. [15]. В совокупности это приводит к неуправляемым случайным разбросам фронтов генерируемых сигналов [16], что можно наблюдать на многоразрядном выходе схемы регистратора. При этом многократно регистрируемое значение числа фронтов при фиксированных MW и CF можно представить как сумму постоянной и случайной (шумовой) составляющей, к которой в том числе можно отнести несовершенство схемы регистратора, порождающей шум квантования. Если в качестве схемы регистратора использовать синхронный многоразрядный ($N=C+L$ разрядов) двоичный счетчик с разрешением, то по окончании измерения старшая часть разрядов C будет характеризовать постоянную составляющую результата измерения R_{id} , а младшая часть разрядов L – шумовую составляющую R_{rnd} (рис. 1).

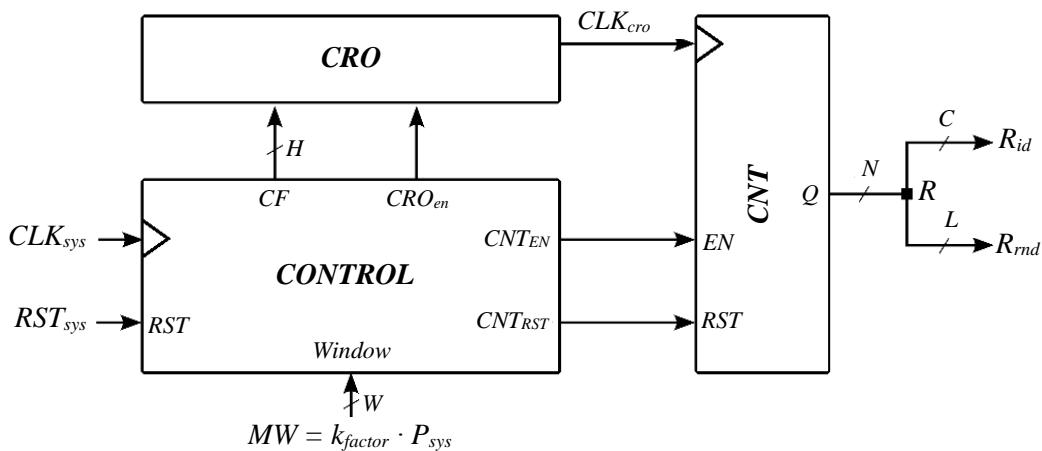


Рис. 1. Структурная схема измерения сигналов ККО

Fig. 1. Block diagram of the CRO signal measurement

В общем случае регистрация числа фронтов выполняется в следующей последовательности: сначала на устройство управления **CONTROL** по W -разрядной шине подается значение желаемого окна измерения MW , далее выбранным значением CF конфигурируется схема ККО (configurable ring oscillator, CRO) одной из 2^H возможных конфигураций, после чего схема переходит в режим генерирования периодического сигнала CLK_{cro} , который подается на вход синхронизации двоичного счетчика **CNT** (Counter). Сигналы разрешения функционирования ККО CRO_{en} и счетчика CNT_{en} являются активными на протяжении всего временного окна измерения, по-

окончании которого значение R на счетчике равно числу зарегистрированных фронтов сигнала CLK_{cro} . При многократном измерении в значениях R можно выделить старшие разряды R_{id} со стабильными значениями и младшие разряды R_{rnd} с нестабильными оцифрованными значениями шумовой составляющей. Таким образом, предложенная схема может выполнять роль двух схем: ФНФ с многоразрядным ответом $R_{id} = PUF(CF, MW)$, обладающей достаточно высоким уровнем стабильности, и ФНФ, выполняющей роль генератора случайных чисел $R_{rnd} = PUF(CF, MW)$.

Покажем на практике, что подобным свойством обладают схемы ККО, реализованные на ПЛИС типа FPGA, а шумовая составляющая R_{rnd} может быть представлена как случайная нормально распределенная величина.

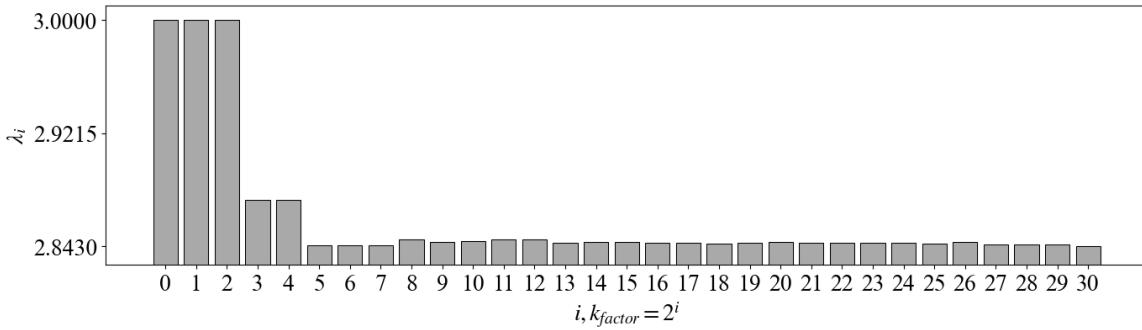
Экспериментальное исследование периодов сигналов ККО. Во многих цифровых системах с кристаллами ПЛИС сигналы системной частоты $CLK_{sys} = P_{sys}^{-1}$ выбираются со значениями, которые гораздо ниже, чем максимально возможная рабочая частота реализуемых схем, $F_{sch} > CLK_{sys}$. Так, на плате быстрого прототипирования Digilent ZYBO-Z7 для части кристалла с программируемой логикой $CLK_{sys} = 50$ МГц, а максимальная частота функционирования 32-разрядного синхронного счетчика, реализованного на программируемой логике, оценивается как $F_{sch} \approx 180$ МГц (для кристалла Xilinx ZYNQ xc7z010-1, выполненного по 28-нанометровому техпроцессу). Согласно теореме Котельникова системная частота может быть применена для измерения периодов схем ККО, частоты выходных сигналов которых не превышают 25 МГц ($P_{cro} = 40$ нс). Для повышения разрешающей способности схемы измерения будем производить оценку периода P_{cro} (частоты $F_{cro} = P_{cro}^{-1}$) сигнала ККО, подавая его на вход синхронизации двоичного синхронного счетчика во временном окне измерения $MW = k_{factor} \cdot P_{sys}$, кратном стационарному периоду P_{sys} системного сигнала синхронизации, где k_{factor} – целочисленный множитель, задаваемый системой измерения. Подобный метод измерения позволяет оценивать частоты $F_{cro} \leq 180$ МГц.

На точность измерения периода сигнала ККО будут влиять ширина окна измерения (множитель k_{factor}) и разрядность регистрирующего счетчика $N=C+L$ (см. рис. 1). Предположим, что $\lambda \cdot P_{cro} = P_{sys}$. Если $\lambda \geq 1$, для регистрации первого фронта сигнала ККО достаточно минимального окна измерения при $k_{factor} = 1$. В случае если $\lambda < 1$, для регистрации первого фронта необходимо, чтобы $k_{factor} = 2 \cdot \lceil \lambda \rceil$. Так как значение P_{cro} заранее неизвестно, измерение можно начинать со значения $k_{factor} = 1$. Если при этом значение регистрирующего счетчика $R \geq 1$, то принимается, что $\lambda \geq 1$. В противном случае, если $R = 0$, то измерение повторяется для большего значения k_{factor} до тех пор, пока не будет достигнуто условие $R \geq 1$. Далее временное окно измерения можно увеличивать, экспоненциально изменяя множитель $k_{factor} = 2^i$, $i \in \mathbb{Z}$. На каждой новой итерации происходит более точное измерение периода ККО $P_{cro} = \frac{MW_i}{R_i} = 2^i \cdot \frac{P_{sys}}{R_i}$ либо значения

$\lambda_i = \frac{P_{sys}}{P_{cro}} = \frac{R_i}{2^i}$. Увеличение степенного показателя i заканчивается при достижении необходимого значения точности: $\epsilon \leq |\lambda_{i+1} - \lambda_i|$.

Пример 1. Для реализованной схемы ККО, конфигурации $CH=0$ и для $k_{factor} = 1$ ($i=0$) значение счетчика $R_0 = 3$. При этом $P_{cro0} = \frac{MW_0}{R_0} = \frac{P_{sys}}{3}$, т. е. $\lambda_0 = 3$. Дальнейшие измерения на увеличенных окнах MW_i дают следующие результаты: $\lambda_1 = \frac{6}{2} = 3$, $\lambda_2 = \frac{12}{4} = 3$, $\lambda_3 = \frac{23}{8} = 3 - \frac{1}{8}$,

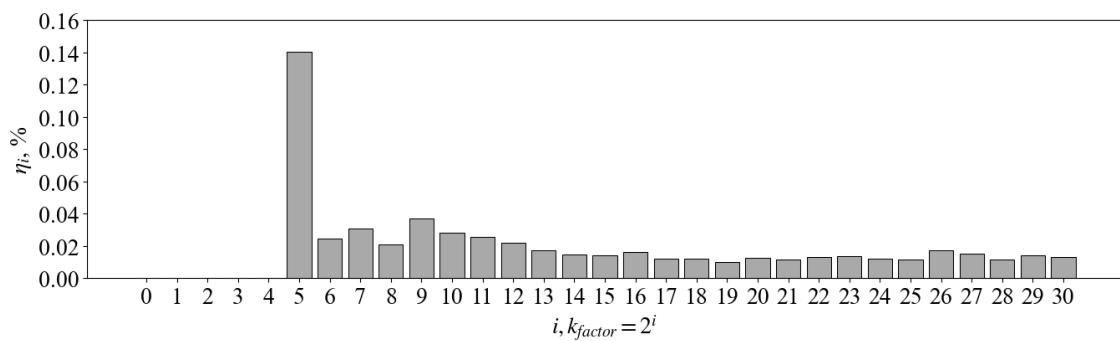
$\lambda_4 = \frac{46}{16} = 3 - \frac{1}{8}$, $\lambda_5 = \frac{91}{32} = 3 - \frac{5}{32}$ и т. д. На рис. 2 приведены значения λ_i для $i \in [0, 30]$. Видно, что с увеличением i значение λ_i асимптотически приближается к 2,84 по мере увеличения окна измерения и для $i = 30$ принимает значение 2,842 954.

Рис. 2. Значения λ^i для $i \in [0, 30]$ Fig. 2. λ^i values for $i \in [0, 30]$

В примере 1 были использованы значения, полученные путем однократных измерений в соответствующих временных окнах.

При многократно повторяемых измерениях M на одном и том же временном окне MW_i наблюдаются различные значения регистрирующего счетчика, вызванные разными факторами, к которым можно отнести нестабильность питающего напряжения, тепловой шум окружающей среды и кристалла FPGA, несовершенство измерительного оборудования и т. п. Так, для $M = 10^3$ повторяющихся измерений и для различных окон MW_i были оценены математические ожидания (МО) $\mu(\lambda_i)$ и среднеквадратические отклонения (СКО) $\sigma(\lambda_i)$ регистрируемых величин λ_i .

На рис. 3 изображена столбчатая гистограмма значений величины $\eta_i = \frac{\sigma(\lambda_i)}{\mu(\lambda_i)} \cdot 100\%$, определяющей удельный вес шумовой составляющей λ_i при измерениях на различных окнах MW_i .

Рис. 3. Зависимость величины η_i от окна измеренияFig. 3. Dependence of the η_i value on the measurement window

Как показали проведенные эксперименты, регистрируемые значения на M повторяющихся измерениях можно представить вероятностями появления единичного символа в каждом j -м разряде счетчика R_i^j :

$$P_i^j = \frac{1}{M} \sum_{m=0}^{M-1} R_{i,m}^j, \quad (1)$$

где $R_{i,m}^j \in \{0,1\}$ – значение j -го разряда счетчика ($j \in [0, N_i - 1]$) на m -м измерении в фиксированном окне MW_i . В качестве примера на рис. 4 приведены значения P_{17}^j для $M = 10^3$, $i=17$ ($k_{factor} = 2^{17}$), $N_i = 19$, $j \in [0, 18]$.

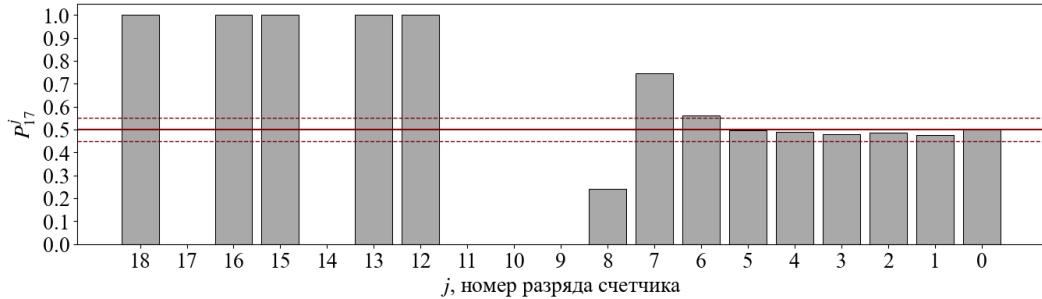
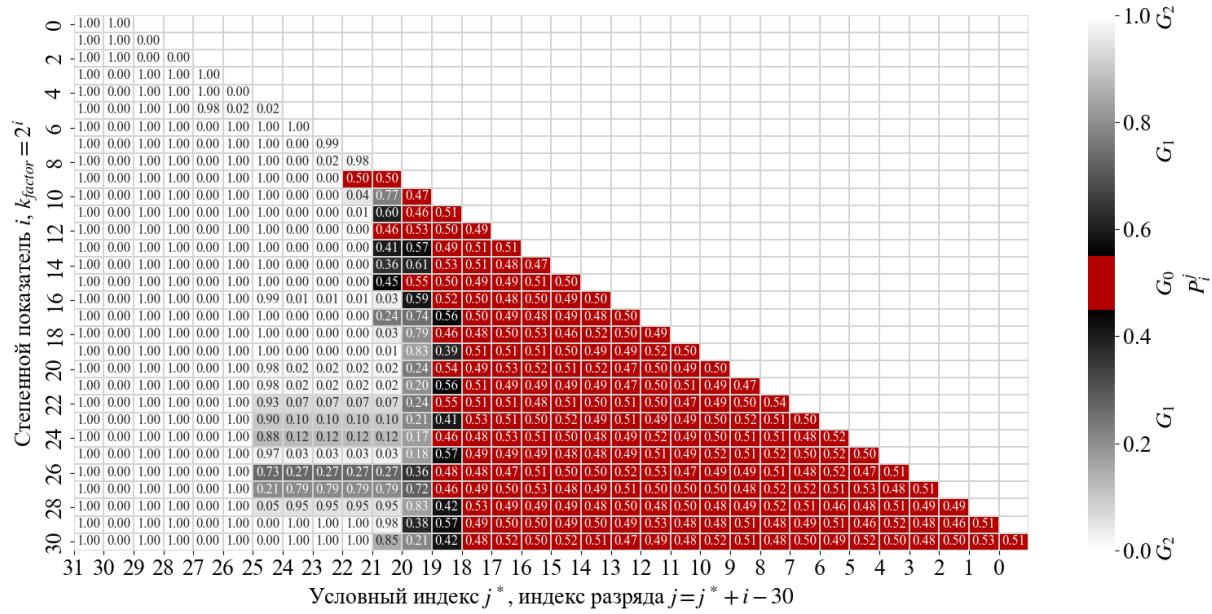


Рис. 4. Значения P_{17}^j для каждого из разрядов регистрирующего счетчика при $k_{factor} = 2^{17}$

Fig. 4. Values of P_{17}^j for each of the digits of the register counter at $k_{factor} = 2^{17}$

На рис. 5 показана тепловая карта значений P_i^j для $i \in [0, 30]$.



3. Группу сильно нестабильных разрядов G_0 (отмечены красным цветом), для которых $0,5 + \xi \geq P_i^j \geq 0,5 - \xi$.

Для данных на рис. 4 и значения $\xi = 0,05$ группы разрядов выглядят следующим образом: $G_2 = \{R_{17}^{18}, R_{17}^{17}, R_{17}^{16}, R_{17}^{15}, R_{17}^{14}, R_{17}^{13}, R_{17}^{12}, R_{17}^{11}, R_{17}^{10}, R_{17}^0\}$, $G_1 = \{R_{17}^8, R_{17}^7, R_{17}^6\}$ и $G_0 = \{R_{17}^5, R_{17}^4, R_{17}^3, R_{17}^2, R_{17}^1, R_{17}^0\}$.

Сумма размерностей представленных групп равна минимально необходимому числу разрядов счетчика N_i , регистрирующего целое число периодов ККО в заданном окне измерения:

$N_i = |G_2| + |G_1| + |G_0| = \lceil \log_2 R_i \rceil = \lceil \log_2(\lambda_i \cdot 2^i) \rceil = \lceil \log_2 \lambda_i \rceil + i$. Для рассматриваемого примера ($i=17$) $N_i = \lceil \log_2 \lambda_i \rceil + i = 2 + 17 = 19$, а $|G_2| = 10$, $|G_1| = 3$ и $|G_0| = 6$, что также в сумме дает 19 разрядов.

Как видно из представленных данных (см. рис. 5), ненулевое число разрядов в группе G_1 начинается с индекса $i=5$, что коррелируется с первым ненулевым значением η_i (см. рис. 3). Начиная с индекса $i=9$, наблюдается ненулевая численность сильно нестабильных разрядов группы G_0 , которая асимптотически линейно возрастает по мере увеличения окна MW_i . При этом общее число разрядов двух групп $|G_2| + |G_1|$, начиная с индекса $i=13$, практически остается стационарным (принимает значение 12 либо 13). Кроме того, начиная с того же индекса $i=13$, распределение сгенерированных значений счетчика становится близким к нормальному распределению.

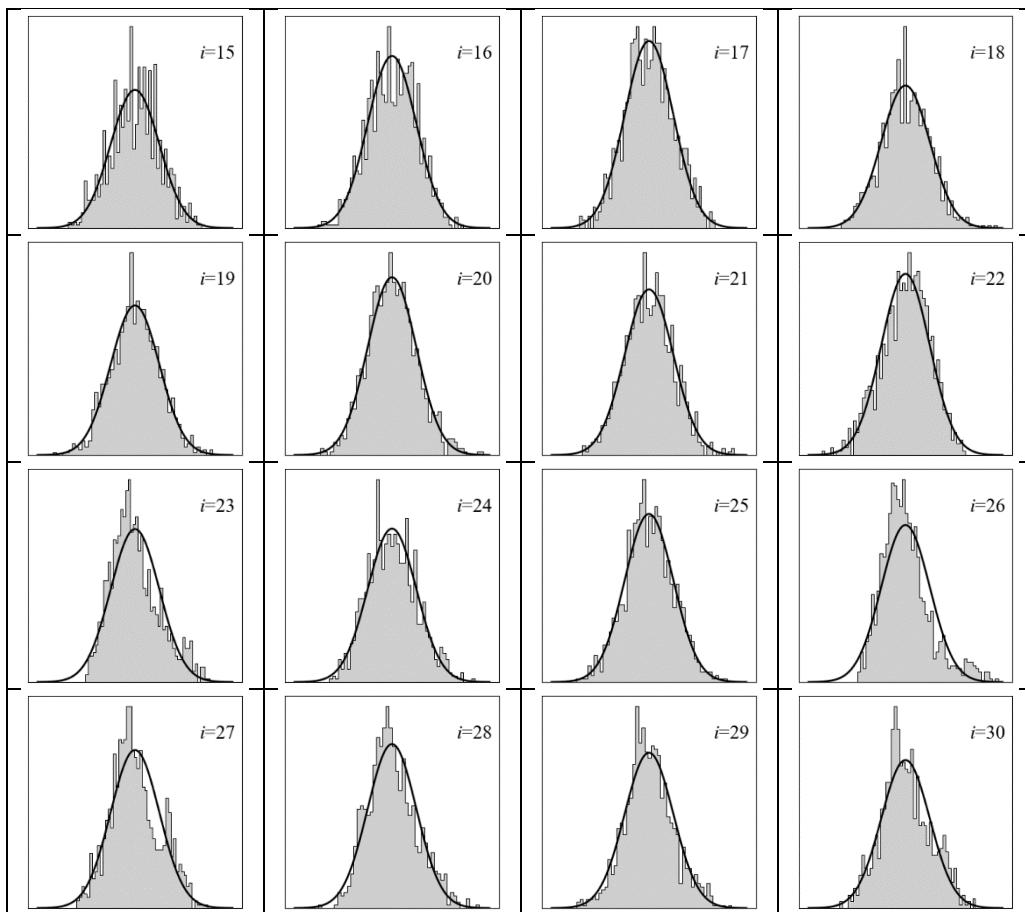


Рис. 6. Гистограммы распределения значений счетчика на различных окнах измерения для $i \in [15, 30]$

Fig. 6. Histograms of the distribution of counter values on different measurement windows for $i \in [15, 30]$

На рис. 6 представлены гистограммы распределения значений счетчика в различных окнах измерения для $i \in [15, 30]$ совместно с графиками нормального распределения, построенными по экспериментальным значениям МО и СКО. Большое сходство с нормальным распределением дает основание использовать разряды регистрирующего счетчика из группы G_0 для генерирования случайных данных.

Проведем аналитическое исследование случайных нормально распределенных положительных целочисленных значений, алгоритмически сгенерированных по экспериментально полученным величинам МО и СКО. Для этого создадим программную модель, позволяющую оценивать распределение разрядов регистрирующего счетчика на группы G_2 , G_1 и G_0 , и сравним полученные распределения с реальными экспериментальными.

Моделирование поведения групп разрядов регистрирующего счетчика. Целью данного исследования будет служить установление закономерностей по распределению групп G_2 , G_1 и G_0 на множестве положительных целочисленных значений, полученных путем квантования и округления случайных нормально распределенных величин, в зависимости от задаваемых значений МО и СКО. Для этого была реализована программная модель на языке Python функционального поведения двоичного счетчика, регистрирующего передние фронты сигналов ККО. Входными данными для модели служат параметры $\mu(R_i)$ (МО) и $\sigma(R_i)$ (СКО) значений регистрирующего счетчика R_i , полученных экспериментальным путем для $i \in [9, 30]$. Для имитации поведения двоичного счетчика параметры $\mu(R_i)$, $\sigma(R_i)$ и вырабатываемые значения округлялись до ближайших меньших целых чисел. При этом первое ненулевое значение $\sigma(R_i)$ появляется для $i=9$, что коррелируется с первой ненулевой размерностью группы G_0 (см. рис. 5).

Для генерирования $M = 10^3$ целочисленных нормально распределенных значений по заданным параметрам использовалась функция `truncnorm` из библиотечного пакета `scipy.stats`. Далее сгенерированные значения переводились в двоичное представление. Для этих значений вычислялись вероятности P_i^j и проводилась оценка размерностей групп G_2 , G_1 и G_0 .

Определим отклонение модельных оценок от экспериментальных при помощи расстояния Евклида. Пусть $V_i^e = (P_i^0; P_i^1; P_i^2; \dots; P_i^{N_i-1})$ – вектор, компонентами которого $V_i^e(j) = P_i^j$, $j \in [0, N_i - 1]$, являются экспериментальные данные, полученные для i -го степенного коэффициента окна измерения. Тогда V_i^m – вектор той же размерности, сформированный из модельных оценок значений P_i^j .

С учетом того что все значения $P_i^j \in [0, 1]$, максимальное расстояние Евклида между векторами V_i^e и V_i^m можно выразить как $D_{em}^{\max}(i) = \sqrt{N_i}$. Введем меру сходства двух векторов V_i^e и V_i^m через их нормированное расстояние Евклида:

$$S_{em}(i) = 1 - D_{em}^{norm}(i) = 1 - \frac{1}{D_{em}^{\max}(i)} \cdot \sqrt{\sum_{j=0}^{N_i-1} (V_i^e(j) - V_i^m(j))^2}. \quad (2)$$

Очевидно, что $S_{em}(i) \in [0, 1]$ и $S_{em}(i) = 0$ соответствуют случаю максимального расстояния (различия) векторов, а $S_{em}(i) = 1$ – полного подобия векторов $V_i^e = V_i^m$.

Пример 2. Для $i=9$ ($k_{factor} = 2^9$, $D_{em}^{\max}(i) = \sqrt{11} \approx 3,3166$) векторы принимают следующие значения: $V_i^e = (1; 0; 1; 1; 0; 1; 0; 0; 0,496; 0,503)$, $V_i^m = (1; 0; 1; 1; 0; 1; 0,943; 0,057; 0,068; 0,334; 0,486)$.

Как видно, пять компонент векторов из 11 различаются, но их отличия незначительные, что выражается в значениях нормированного расстояния $D_{em}^{norm}(9) \approx 0,0585$ и меры сходства $S_{em}(9) = 0,9415$.

На рис. 7 показаны значения $S_{em}(i)$, вычисленные для $i \in [9, 30]$.

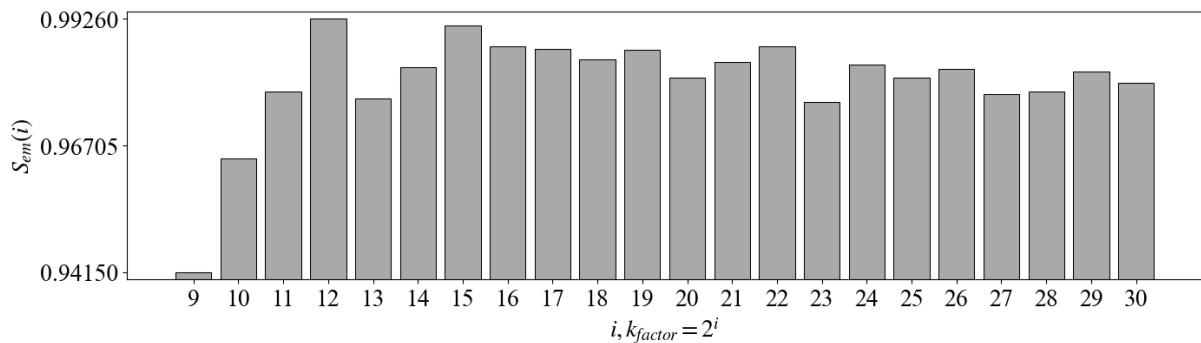


Рис. 7. Значения меры сходства $S_{em}(i)$ для $i \in [9, 30]$

Fig. 7. Similarity measure values $S_{em}(i)$ for $i \in [9, 30]$

Из представленных значений видно, что для всех окон измерений ($i \in [9, 30]$) наблюдаются незначительные отличия экспериментальных данных от данных, полученных на программной модели. Это может служить подтверждением гипотезы о распределении многократно измеряемых значений R_i , близком к нормальному распределению случайной величины.

Рассмотрим непрерывную случайную нормально распределенную величину Q : $Q \sim N(\mu_Q, \sigma_Q^2)$. Пусть Q равномерно квантована $T = 2^L$ уровнями и кодируется L -разрядными положительными двоичными числами, каждое из которых будем рассматривать в следующей нотации: $b^t = (b_{L-1}^t, b_{L-2}^t, b_{L-3}^t, \dots, b_0^t)$, где $b_l^t \in \{0, 1\}$ есть l -й разряд двоичного числа b^t , кодирующего уровень квантования t , при этом $t \in [0, T-1]$, $l \in [0, L-1]$, $T \in \mathbb{Z}$, $L \in \mathbb{Z}$. В общем случае b^t есть двоичное представление уровня квантования t . Например, для $L=4$ и $t=5$ имеется $T=16$ уровней квантования, а $b^5 = (b_3^5, b_2^5, b_1^5, b_0^5) = (0, 1, 0, 1)$.

Диапазон значений, в который будут попадать большинство чисел (99,9937 %), представим как $D_Q = [\mu_Q - 4 \cdot \sigma_Q, \mu_Q + 4 \cdot \sigma_Q]$, тогда шаг квантования будет равен $\Delta_Q = \frac{8 \cdot \sigma_Q}{T} = 2^{3-L} \cdot \sigma_Q$.

Произвольное значение q случайной величины Q принадлежит одному из уровней квантования t из T возможных и лежит в диапазоне $\mu_Q - 4\sigma_Q + t \cdot \Delta_Q \leq q < \mu_Q - 4\sigma_Q + (t+1) \cdot \Delta_Q$. Если применить квантование с использованием усечения, то значение q будет соответствовать целому числу t либо его двоичному представлению b^t . Так, для предыдущего примера случайная величина q из диапазона $[\mu_Q - \frac{3}{2}\sigma_Q, \mu_Q - \sigma_Q)$ будет представлена двоичным числом b^5 .

С учетом сказанного выше можно рассматривать эквивалентную случайную величину $B \sim N(\mu_B, \sigma_B^2)$, для которой диапазон большинства принимаемых значений равен $D_B = [\mu_B - 4 \cdot \sigma_B, \mu_B + 4 \cdot \sigma_B] = [0, 2^L)$, при этом $\mu_B = 2^{L-1}$, $\sigma_B = 2^{L-3}$, $\Delta_B = 1$. Произвольное значение b случайной величины B принадлежит одному из диапазонов $D_B^t = [t, t + \Delta_B)$, а квантованное с усечением значение будет равно t (b^t в двоичном представлении) (рис. 8).

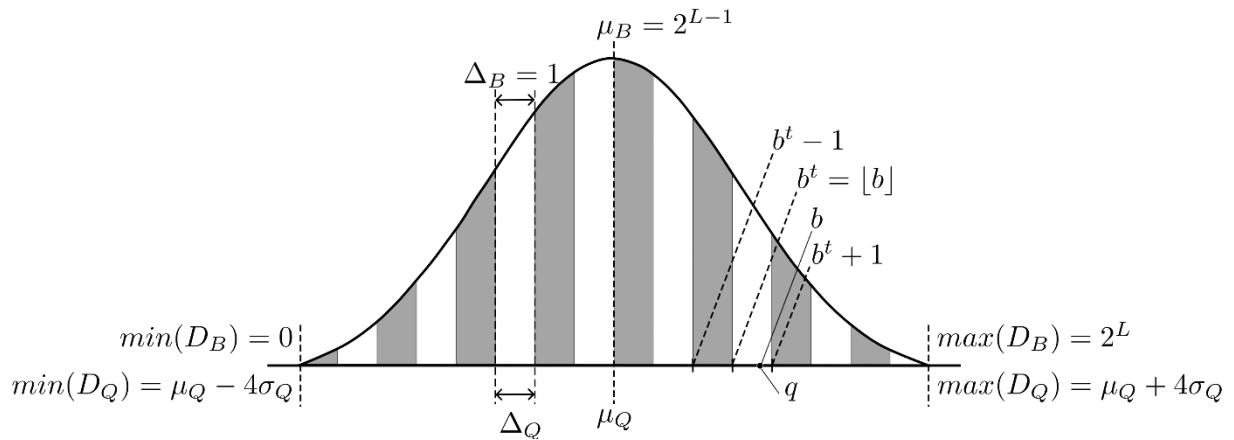


Рис. 8. Основные параметры величин Q и B
Fig. 8. Basic parameters of the quantities Q and B

Так как $\Delta_B = 1$, то записи вида $t + \Delta_B$ и $t + 1$ являются эквивалентными. В силу симметрии функции распределения случайной величины B диапазон D_B можно представить двумя равновеликими диапазонами $D_B = D_{LB} \cup D_{HB}$, где $D_{LB} = [0, \mu_B]$, а $D_{HB} = [\mu_B, 2^L]$, для которых выполняется равенство вероятностей принадлежности $P(B \in D_{LB}) = P(B \in D_{HB})$.

Оценим вероятность появления единичного бита в старшем разряде $P(B_{L-1} = 1)$ при квантовании и двоичном кодировании величины Q .

Утверждение 1. *Вероятность появления единичного бита в старшем разряде $P(B_{L-1} = 1) = P(B \in D_{HB}) = 0,5$.*

Доказательство. Все двоичные числа, для которых $b_{L-1}^t = 1$, принадлежат диапазону D_{HB} , что для непрерывной величины B соответствует вероятности попадания в диапазон $P(\mu_B \leq B < \mu_B + 4\sigma_B)$. Стандартизируем случайную величину B относительного стандартного нормального распределения $Z \sim N(0,1)$: $Z = \frac{B - \mu_B}{\sigma_B}$. Для нижней границы диапазона $Z_{\min} = \frac{\mu_B - \mu_B}{\sigma_B} = 0$, а для верхней $- Z_{\max} = \frac{\mu_B + 4\sigma_B - \mu_B}{\sigma_B} = 4$.

Тогда вероятность попадания случайной величины в заданный диапазон вычисляется как $P(0 \leq Z < 4) = \Phi(Z_{\max}) - \Phi(Z_{\min}) = \Phi(4) - \Phi(0) = 0,9999 - 0,5 = 0,4999$, где Φ – функция стандартного нормального распределения. Таким образом, $P(B_{L-1} = 1) \approx 0,5$, что и требовалось доказать.

Изменим принимаемые кодированные значения величины B в двоичном представлении путем добавления одного старшего разряда с константным значением $B_L = 1$: $b^t = (b_L^t, b_{L-1}^t, b_{L-2}^t, \dots, b_0^t)$, $\forall t \in [0, T-1]$, при этом число T квантованных уровней и значение σ_B остаются прежними. В таком случае значение матожидания будет равным $\mu_B^* = 2^L + \mu_B$, а диапазон принимаемых значений будет выглядеть следующим образом: $D_{HB}^* = [2^L, 2^{L+1})$.

Оценим вероятность $P(B_L = 1)$ как вероятность попадания величины B в диапазон D_{HB}^* : $P(B \in D_{HB}^*) = P(2^L \leq B < 2^{L+1})$. Стандартизируем случайную величину B относительного стандартного нормального распределения $Z \sim N(0,1)$: $Z = \frac{B - \mu_B^*}{\sigma_B}$.

Для нижней границы диапазона $Z_{\min} = \frac{2^L - 2^L - \mu_B}{\sigma_B} = -\frac{\mu_B}{\sigma_B} = -4$, а для верхней – $Z_{\max} = \frac{2^{L+1} - 2^L - \mu_B}{\sigma_B} = \frac{\mu_B}{\sigma_B} = 4$. Тогда вероятность $P(2^L \leq B < 2^{L+1}) = \Phi(4) - \Phi(-4) = 1 - 1 + \Phi(4) = 1$. Таким образом, $P(B^L = 1) = 1$.

В общем случае расширение кодированных значений величины B на C константных разрядов с произвольными значениями $b^t = (b_{L+C-1}^t, b_{L+C-2}^t, \dots, b_{L+1}^t, b_L^t, b_{L-1}^t, b_{L-2}^t, \dots, b_0^t)$, $\forall t \in [0, T-1]$ также дадут $P(B_{l-1} = 1) = 1$ либо $P(B_{l-1} = 0) = 1$, $\forall l \in [L, L+C-1]$. Иными словами, все старшие добавленные константные разряды b_i^t образуют группу G_2 . На основании этого сформулируем следующее следствие.

Следствие 1. Если случайная нормально распределенная величина $B \sim N(\mu_B, \sigma_B^2)$, для которой $\sigma_B = 2^{L-3}$, $\mu_B = H_C + 2^{L-1}$, где $H_C > 2^{L-1}$, и $\mu_B \equiv 0 \pmod{2^L}$, квантуется $T = 2^L$ уровнями, кодирующимися $(C+L)$ -разрядными двоичными числами, то C старших разрядов (кодирующих значение H_C) являются неизменными и образуют группу G_2 .

Для следующего утверждения введем понятие инверсного диапазона: для произвольного диапазона $D_B^t = [b^t, b^t + 1]$ его инверсия $\overline{D}_B^t = [\overline{b}^t, \overline{b}^t + 1]$, где \overline{b}^t есть поразрядная инверсия значения b^t , $\forall t \in [0, T-1]$.

Утверждение 2. Вероятности принадлежности случайной величины B к произвольному диапазону D_B^t и его инверсии \overline{D}_B^t равны: $P(B \in D_B^t) = P(B \in \overline{D}_B^t)$.

Данное утверждение верно в силу симметрии функции распределения относительно значения $\mu_B = 2^{L-1}$.

Доказательство. Инверсию произвольного двоичного L -разрядного числа b^t можно представить выражением $\overline{b}^t = (b_{L-1}^t, b_{L-2}^t, b_{L-3}^t, \dots, b_0^t)$ либо в десятичном представлении выражением $\overline{t} = 2^L - 1 - t = 2\mu_B - 1 - t$.

В общем случае вероятность $P(B \in D_B^t)$ представим как равенство

$$P(B \in D_B^t) = P(t \leq B < t + 1) = \Phi(Z_{\max}) - \Phi(Z_{\min}), \quad (3)$$

где $Z_{\min} = \frac{t - \mu_B}{\sigma_B}$, а $Z_{\max} = \frac{t + 1 - \mu_B}{\sigma_B} = Z_{\min} + \frac{1}{\sigma_B}$ – стандартизованные значения.

Выразим аналогичным образом вероятность $P(B \in \overline{D}_B^t)$:

$$P(B \in \overline{D}_B^t) = P(2\mu_B - 1 - t \leq B < 2\mu_B - t) = \Phi(Z_{\max}^*) - \Phi(Z_{\min}^*). \quad (4)$$

Рассмотрим стандартизованные значения инверсного диапазона Z_{\max}^* и Z_{\min}^* :

$$\begin{aligned} Z_{\max}^* &= \frac{2\mu_B - t - \mu_B}{\sigma_B} = -\frac{t - \mu_B}{\sigma_B} = -Z_{\min}, \\ Z_{\min}^* &= \frac{2\mu_B - 1 - t - \mu_B}{\sigma_B} = -\frac{t + 1 - \mu_B}{\sigma_B} = -Z_{\max} = -(Z_{\min} + \frac{1}{\sigma_B}). \end{aligned} \quad (5)$$

Тогда согласно симметрии функции стандартного нормального распределения относительно нуля вероятность (4) можно выразить следующим образом: $\Phi(Z_{\max}^*) - \Phi(Z_{\min}^*) = \Phi(-Z_{\min}) - \Phi(-Z_{\max}) = \Phi(Z_{\max}) - \Phi(Z_{\min}) = P(B \in D_B^t)$, что и требовалось доказать.

Обозначим составной диапазон $D_B^{(l,1)} = \bigcup_{t=0}^{T-1} D_B^t$ как совокупность диапазонов квантования, для

которых выполняется условие $b_l^t = 1, l \in [0, L-2]$. По аналогии введем составной диапазон

$D_B^{(l,0)} = \bigcup_{t=0}^{T-1} D_B^t, \forall b_l^t = 0$. Произвольный диапазон $D_B^{(l,1)}$ можно представить следующим образом:

$D_B^{(l,1)} = D_{LB}^{(l,1)} \cup D_{HB}^{(l,1)}$, где $D_{LB}^{(l,1)} = \bigcup_{t=0}^{\mu_B-1} D_B^t$ при $b_{l-1}^t = 0$, а $D_{HB}^{(l,1)} = \bigcup_{t=\mu_B}^{T-1} D_B^t$ при $b_{l-1}^t = 1, \forall b_l^t = 1$,

$l \in [0, L-2]$. Тогда вероятность принадлежности случайной величины диапазону $D_B^{(l,1)}$ можно выразить как $P(B \in D_B^{(l,1)}) = P(B \in D_{LB}^{(l,1)}) + P(B \in D_{HB}^{(l,1)})$.

Утверждение 3. *Вероятность принадлежности случайной величины B диапазону $D_B^{(l,1)}$ равна $P(B \in D_B^{(l,1)}) = 0,5$ для всех $l \in [0, L-2]$.*

Доказательство. Произведем инверсию составного диапазона $D_{LB}^{(l,1)}$: $\overline{D_{LB}^{(l,1)}} = \bigcup_{t=0}^{\mu_B-1} \overline{D_B^t} = \bigcup_{t=\mu_B}^{M-1} D_B^t = D_{HB}^{(l,0)}$. В свою очередь, объединение диапазонов $D_{HB}^{(l,1)}$ и $D_{HB}^{(l,0)}$ дает диапазон D_{HB} : $D_{HB}^{(l,1)} \cup D_{HB}^{(l,0)} = D_{HB}$. Основываясь на утверждении 2, $P(B \in D_{LB}^{(l,1)}) = P(B \in \overline{D_{LB}^{(l,1)}}) = P(B \in D_{HB}^{(l,0)})$. Тогда в разложении $P(B \in D_B^{(l,1)}) = P(B \in D_{LB}^{(l,1)}) + P(B \in D_{HB}^{(l,1)})$ произведем замену вероятности $P(B \in D_{LB}^{(l,1)})$ на вероятность $P(B \in D_{HB}^{(l,0)})$ в силу их равенства: $P(B \in D_{HB}^{(l,0)}) + P(B \in D_{HB}^{(l,1)}) = P(B \in D_{HB})$. Согласно утверждению 1 $P(B \in D_{HB}) = 0,5$, что и требовалось доказать.

Пример 3. Пусть $L=3$, тогда $T=8$, $\mu_B=4$, $\sigma_B=1$, $D_B=[0,8]$. Оценим вероятность $P(B \in D_B^{(l,1)})$ для $l=0$. Весь диапазон принимаемых значений D_B условно разделим на симметричные диапазоны $D_{LB}=[0,4]$ и $D_{HB}=[4,8]$. Тогда $D_B^{(0,1)} = D_B^1 \cup D_B^3 \cup D_B^5 \cup D_B^7$, $D_{LB}^{(0,1)} = D_B^1 \cup D_B^3$, а $D_{HB}^{(0,1)} = D_B^5 \cup D_B^7$. Инверсия составного диапазона выглядит как $\overline{D_{LB}^{(0,1)}} = D_B^6 \cup D_B^4 = D_{HB}^{(0,0)}$. Согласно утверждению 2 следующее равенство является верным: $P(B \in D_{LB}^{(0,1)}) = P(B \in \overline{D_{LB}^{(0,1)}}) = P(B \in D_{HB}^{(0,0)})$. Тогда вероятность принадлежности составному диапазону $D_B^{(0,1)}$ представим как $P(B \in (D_B^6 \cup D_B^4 \cup D_B^5 \cup D_B^7)) = P(B \in D_{HB}^{(0,0)}) = 0,5$.

Следствие 2. *Если случайная нормально распределенная величина $B \sim N(\mu_B, \sigma_B^2)$, для которой $\sigma_B = 2^{L-3}$, а $\mu_B \geq 2^{L-1}$ и $\mu_B \equiv 0 \pmod{2^L}$, квантуется $T = 2^L$ уровнями, кодирующимися $(C+L)$ -разрядными двоичными числами, то L младших разрядов образуют группу G_0 .*

Следствие 2 дополняет следствие 1 на основании утверждения 3.

Для подтверждения приведенных теоретических выводов были проведены эксперименты на программной модели, позволяющие оценить разделение на группы G_2 и G_0 при параметрах случайной величины, указанных в следствиях 1 и 2.

Рассмотренный случай является идеализированным по таким параметрам, как матожидание и СКО, значения которых на практике могут быть произвольными. Так, если в рассмотренной модели значение $\mu_Q = H_C + 2^{L-1} + \chi$, $H_C > 2^{L-1}$, $\chi < 2^{L-1}$, то при подсчете числа передних фронтов на синхронном счетчике могут возникать случаи, когда происходит арифметический перенос в разряды, порядковые индексы которых больше чем $\lceil \log_2 \chi \rceil$. Это влияет на значения вероятностей P_i^j (1) для разрядов в окрестности $\mu_B = 2^{L-1}$.

Искажения могут затронуть и разряды из группы G_0 , в этом случае $P_i^j \neq 0,5$, и (или) разряды из группы G_2 , что даст $1,0 < P_i^j < 0,5$ либо $0,5 < P_i^j < 1,0$, формируя группу G_1 .

Обобщая вышесказанное, можно утверждать, что для определенных значений множителя многоократно регистрируемое значение счетчика, вероятнее всего, будет содержать три ненулевые группы разрядов G_2 , G_1 и G_0 . Методом максимального правдоподобия группу G_1 можно преобразовать в часть дополнительных разрядов группы G_2 либо вычесть значение шумовой составляющей матожидания $\chi < 2^{L-1}$ из регистрируемых счетчиком значений для приведения к виду (G_2, G_0) .

Эксперимент по устраниению группы G_1 . Основной целью эксперимента является искусственное устранение условно стабильной группы G_1 (нахождение значения $\chi < 2^{L-1}$). Ранее (см. рис. 5) были получены экспериментальные данные со счетчика для $\forall i \in [0, 30]$. Для каждой выборки значений счетчика R_i на повторяющихся k измерениях $(R_{i,1}, R_{i,2}, R_{i,3}, \dots, R_{i,k})$ производилась оценка математического ожидания (округление до ближайшего целого) μ_i и среднеквадратичного отклонения σ_i . Согласно следствию 2 возможна оценка размерности нестабильной части G_0 , которая показывает, какое количество младших бит можно считать шумовой составляющей: $|G_0| = 3 + \lfloor \log_2 \sigma_i \rfloor$. После этого выделяется информативная, или стабильная, часть математического ожидания H_C , размерность которой можно оценить как $|H_C| = N_i - |G_0|$. Эта часть демонстрирует стабильность в рамках всей выборки и не изменяется под действием случайных шумов. Далее информативные биты H_C используются для формирования нового модифицированного среднего значения μ_i^* , состоящего из трех частей: $|H_C|$ значащих старших бит, разделяющего бита «1» и $|G_0| - 1$ нулевых младших бит ($\mu_i^* = [H_C : 1 : 0^{|G_0|-1}]_2$). Разделительный бит «1» предотвращает влияние шумовой составляющей на стабильные разряды. Таким образом, результирующее значение μ_i^* представляет собой математическое ожидание, очищенное от влияния условно нестабильной группы G_1 , а полученная ненулевая разница $\Delta_i = \mu_i - \mu_i^*$ интерпретируется как влияние этой группы. Вычитание из каждого элемента выборки R_i значения Δ_i равносильно исключению группы G_1 из статистики, поскольку ее вклад компенсируется в каждом значении, образуя новую модифицированную выборку, состоящую только из двух групп: стабильной G_2 и нестабильной G_0 .

Пример 4. Для полученных ранее значений имеем $i = 19$, $\mu_{19} = 1491607$, $\sigma_{19} = 148,18$. Оценим размерность нестабильной группы: $|G_0| = 3 + \lfloor \log_2 \sigma_{19} \rfloor = 10$. Переведем μ_{19} в бинарный вид: $\mu_{19} = [101101100001010010111]_2$, $N_{19} = 21$.

Для формирования нового значения μ_{19}^* берется стабильная μ_{19} часть размером $|H_C| = N_i - |G_0| = 11$ бит: $H_C = [10110110000]_2$. Тогда $\mu_{19}^* = [H_C : 1 : 0^9]$, а значение $\Delta_{19} = \mu_{19} - \mu_{19}^* = 151_{10}$. Скорректировав каждое значение выборки R_{19} на Δ_{19} , получим распределение по группам (рис. 9).

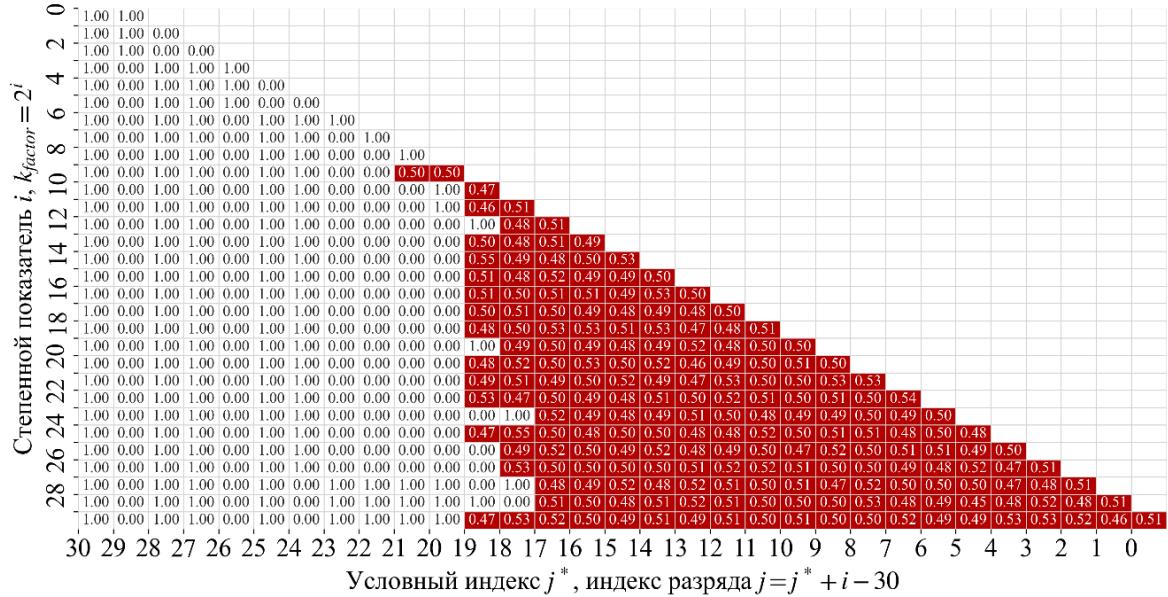


Рис. 9. Значения P_i^j , $\forall i \in [0, 30]$ и $\xi = 0, 05$ после обработки

Fig. 9. Values of P_i^j , $\forall i \in [0, 30]$ and $\xi = 0, 05$ after postprocessing

На рис. 9 видно, что для всех значений i удалось нивелировать группу G_1 , оставив распределение значений регистрируемого счетчика на две группы: G_2 и G_0 . Следует отметить, что для значений $3 + \log_2 \sigma_i \leq 0$ коррекция значения выборок не проводилась.

Заключение. В статье представлены результаты исследования поведения схемы регистрации числа импульсов конфигурируемого кольцевого генератора в фиксированном временном окне. Было показано, что разряды регистрируемого счетчика можно условно разделить на три группы: старшие стабильные разряды, условно стабильные разряды и младшие сильно нестабильные разряды, вероятность появления единичного символа в которых близка к значению 0,5. Анализ собранных экспериментальных данных и проведенное исследование программной модели рассматриваемой схемы показали схожесть генерируемых значений со значениями случайной нормально распределенной величины. Теоретически и практически была показана возможность нивелирования группы условно стабильных разрядов, что открывает перспективы для исследования новых подходов к генерации уникальных идентификаторов, повышения стабильности ФНФ и генерации случайных чисел на основе цифровых схем.

Вклад авторов. А. А. Иванюк выдвинул гипотезу об использовании разрядов двоичного счетчика, регистрирующего сигналы от схемы конфигурируемого кольцевого осциллятора для генерирования уникальных стабильных идентификаторов и случайных чисел. Л. А. Бурко провела экспериментальные исследования и приняла участие в обобщении, анализе и оформлении полученных результатов.

Список использованных источников

1. Secure System Design and Trustable Computing / ed.: Ch. H. Chang, M. Potkonjak – Switzerland : Springer, 2016. – 549 p. – DOI: 10.1007/978-3-319-14971-4.
2. Ярмолик, В. Н. Физически неклонируемые функции / В. Н. Ярмолик, Ю. Г. Вашинко // Информатика. – 2011. – Т. 30, № 2. – С. 92–103.
3. Vinagreiro Gutierrez, S. Physical Unclonable Functions (PUFs): foundations, evaluation, and testing for secure hardware systems / S. Vinagreiro Gutierrez, G. Di Natale, I. Vatajelu // 30th IEEE European Test Symp. (ETS 2025), Tallinn, Estonia, May 2025. – URL: <https://hal.science/hal-05111870> (date of access: 28.08.2025).
4. Hemavathy, S. Arbiter PUF – a review of design, composition, and security aspects / S. Hemavathy, V. S. K. Bhaaskaran // IEEE Access. – 2023. – Vol. 11. – P. 33979–34004. – DOI: 10.1109/ACCESS.2023.3264016.
5. Иванюк, А. А. Синтез симметричных путей физически неклонируемой функции типа арбитр на FPGA / А. А. Иванюк // Информатика. – 2019. – Т. 16, № 2. – С. 99–108.
6. Maiti, A. Improved ring oscillator PUF: an FPGA-friendly secure primitive / A. Maiti, P. Schaumont // Journal of Cryptology. – 2011. – Vol. 24. – P. 375–397. – DOI: 10.1007/s00145-010-9088-4.
7. Configurable ring oscillator PUF using hybrid logic gates / D. Deng, S. Hou, Z. Wang, Y. Guo // IEEE Access. – 2020. – Vol. 8. – P. 161427–161437. – DOI: 10.1109/ACCESS.2020.3021205.
8. Иванюк, А. А. Конфигурируемый кольцевой осциллятор с управляемыми межсоединениями / А. А. Иванюк, В. Н. Ярмолик // Безопасность информационных технологий. – 2024. – Т. 31, № 2. – С. 121–133. – DOI: 10.26583/bit.2024.2.08.
9. Иванюк, А. А. Физически неклонируемые функции на базе управляемого кольцевого осциллятора / А. А. Иванюк, В. Н. Ярмолик // Безопасность информационных технологий. – 2023. – Т. 30, № 3. – С. 90–103. – DOI: 10.26583/bit.2023.3.06.
10. Иванюк, А. А. Исследование физически неклонируемой функции конфигурируемого кольцевого осциллятора / А. А. Иванюк // Информатика. – 2025. – Т. 22, № 1. – С. 73–89. – DOI: 10.37661/1816-0301-2025-22-1-73-89.
11. Xin, X. A configurable ring-oscillator-based PUF for Xilinx FPGAs / X. Xin, J.-P. Kaps, K. Gaj // 14th Euromicro Conf. on Digital System Design, Oulu, Finland, 31 Aug. – 02 Sept. 2011. – Oulu, 2011. – P. 651–657. – DOI: 10.1109/DSD.2011.88.
12. Hardware-efficient configurable ring-oscillator-based physical unclonable function/true random number generator module for secure key management / S. Sánchez-Solano, L. F. Rojas-Muñoz, C. Martínez-Rodríguez, P. Brox // Sensors. – 2024. – Vol. 24. – P. 5674–5707. – DOI: 10.3390/s24175674.
13. A large scale characterization of RO-PUF / A. Maiti, J. Casarona, L. McHale, P. Schaumont // 2010 IEEE Intern. Symp. on Hardware-Oriented Security and Trust (HOST), Anaheim, CA, USA, 13–14 June 2010. – Anaheim, 2010 – P. 94–99. – DOI: 10.1109/HST.2010.5513108.
14. Kulagin V. Optimizing RO-PUFs: a filtering approach to reliability and entropy trade-offs / V. Kulagin, G. Di Natale, I. Vatajelu // IEEE 30th European Test Symp. (ETS 2025), Tallinn, Estonia, May 2025. – URL: <https://hal.science/hal-05111852v1> (date of access: 28.08.2025).
15. Vinagreiro Gutierrez, S. On-line method to limit unreliability and bit-aliasing in RO-PUF / S. Vinagreiro Gutierrez, G. Di Natale, I. Vatajelu // IEEE 29th Intern. Symp. on On-Line Testing and Robust System Design (IOLTS 2023), Crete, Greece, July 2023. – URL: https://hal.science/hal-04193294/file/IOLTS__RO_Reliability_and_Bitaliasing.pdf (date of access: 28.08.2025). – DOI: 10.1109/IOLTS59296.2023.10224877.
16. Jitter and phase noise in ring oscillators // The Design of Low Noise Oscillators. – Boston, MA : Springer, 1999. – P. 79–110. – DOI: 10.1007/0-306-48199-5_5.

References

1. Chang Ch. H., Potkonjak M. (eds.). *Secure System Design and Trustable Computing*. Switzerland, Springer, 2016, 549 p. DOI: 10.1007/978-3-319-14971-4.
2. Yarmolik V. N., Vashynko Yu. G. *Physically unclonable functions*. Informatika [Informatics], 2011, vol. 30, no. 2, pp. 92–103 (In Russ.).
3. Vinagreiro Gutierrez S., Di Natale G., Vatajelu I. Physical Unclonable Functions (PUFs): foundations, evaluation, and testing for secure hardware systems. *30th IEEE European Test Symposium (ETS 2025), Tallinn, Estonia, May 2025*. Available at: <https://hal.science/hal-05111870> (accessed 28.08.2025).
4. Hemavathy S., Bhaaskaran V. S. K. Arbiter PUF – a review of design, composition, and security aspects. *IEEE Access*, 2023, vol. 11, pp. 33979–34004. DOI: 10.1109/ACCESS.2023.3264016.

5. Ivaniuk A. A. *Synthesis of symmetric paths of arbiter physically unclonable function on FPGA*. Informatika [Informatics], 2019, vol. 16, no. 2, pp. 99–108 (In Russ.).
6. Maiti A., Schaumont P. Improved ring oscillator PUF: an FPGA-friendly secure primitive. *Journal of Cryptology*, 2011, vol. 24, pp. 375–397. DOI: 10.1007/s00145-010-9088-4.
7. Deng D., Hou S., Wang Z., Guo Y. Configurable ring oscillator PUF using hybrid logic gates. *IEEE Access*, 2020, vol. 8, pp. 161427–161437. DOI: 10.1109/ACCESS.2020.3021205.
8. Ivaniuk A. A., Yarmolik V. N. *Configurable ring oscillator with controlled interconnections*. Bezopasnost` informatsionnykh tekhnologiy [IT Security (Russia)], 2024, vol. 31, no. 2, pp. 121–133 (In Russ.). DOI: 10.26583/bit.2024.2.08.
9. Ivaniuk A. A., Yarmolik V. N. *Physically unclonable functions based on a controlled ring oscillator*. Bezopasnost` informatsionnykh tekhnologiy [IT Security (Russia)], 2023, vol. 30, no. 3, pp. 90–103 (In Russ.). DOI: 10.26583/bit.2023.3.06.
10. Ivaniuk A. A. *Investigation of the physically unclonable function of a configurable ring oscillator*. Informatika [Informatics], 2025, vol. 22, no. 1, pp. 73–89 (In Russ.). DOI: 10.37661/1816-0301-2025-22-1-73-89.
11. Xin X., Kaps J.-P., Gaj K. A configurable ring-oscillator-based PUF for Xilinx FPGAs. *14th Euromicro Conference on Digital System Design, Oulu, Finland, 31 August – 02 September 2011*. Oulu, 2011, pp. 651–657. DOI: 10.1109/DSD.2011.88.
12. Sánchez-Solano S., Rojas-Muñoz L. F., Martínez-Rodríguez C., Brox P. Hardware-efficient configurable ring-oscillator-based physical unclonable function/true random number generator module for secure key management. *Sensors*, 2024, vol. 24, pp. 5674–5707. DOI: 10.3390/s24175674.
13. Maiti A., Casarona J., McHale L., Schaumont P. A large scale characterization of RO-PUF. *2010 IEEE International Symposium on Hardware-Oriented Security and Trust (HOST), Anaheim, CA, USA, 13–14 June 2010*. Anaheim, 2010, pp. 94–99. DOI: 10.1109/HST.2010.5513108.
14. Kulagin V., Di Natale G., Vatajelu I. Optimizing RO-PUFs: a filtering approach to reliability and entropy trade-offs. *30th IEEE European Test Symposium (ETS 2025), May 2025, Tallinn, Estonia*. Available at: <https://hal.science/hal-05111852v1> (accessed 28.08.2025).
15. Vinagreiro Gutierrez S., Di Natale G., Vatajelu I. On-line method to limit unreliability and bit-aliasing in RO-PUF. *IEEE 29th International Symposium on On-Line Testing and Robust System Design (IOLTS 2023), Crete, Greece, July 2023*. Available at: https://hal.science/hal-04193294/file/IOLTS__RO_Reliability_and_Bitaliasing.pdf (accessed 28.08.2025). DOI: 10.1109/IOLTS59296.2023.10224877.
16. Jitter and phase noise in ring oscillators. *The Design of Low Noise Oscillators*. Springer, Boston, MA, 1999, pp. 79–110. DOI: 10.1007/0-306-48199-5_5.

Информация об авторах

Иванюк Александр Александрович, доктор технических наук, профессор, профессор кафедры информатики, Белорусский государственный университет информатики и радиоэлектроники.

E-mail: ivaniuk@bsuir.by

Бурко Лиана Александровна, магистрант, факультет компьютерных систем и сетей, Белорусский государственный университет информатики и радиоэлектроники.

E-mail: burkoliana@gmail.com

Information about the authors

Alexander A. Ivaniuk, D. Sc. (Eng.), Prof., Prof. of Comp. Sci. Department, Belarusian State University of Informatics and Radioelectronics.

E-mail: ivaniuk@bsuir.by

Liana A. Burko, Undergraduate, the Faculty of Computer Systems and Networks, Belarusian State University of Informatics and Radioelectronics.

E-mail: burkoliana@gmail.com

ЛОГИЧЕСКОЕ ПРОЕКТИРОВАНИЕ LOGICAL DESIGN



УДК 519.714.5
DOI: 10.37661/1816-0301-2025-22-4-82-93

Оригинальная статья
Original Article

Распознавание подсхем трехстабильных элементов в КМОП СБИС

Л. Д. Черемисинова

Объединенный институт проблем информатики
Национальной академии наук Беларусь,
ул. Сурганова, 6, Минск, 220012, Беларусь
E-mail: cld@newman.bas-net.by

Аннотация

Цели. Рассматривается проблема экстракции высокоуровневой структуры на уровне логических элементов из схемы на транзисторном уровне. Получение такого представления существенно снижает время выполнения верификации топологии СБИС на стадии ее проектирования и служит основой для перепроектирования интегральных схем. Целью исследования является разработка метода и программных средств выделения в КМОП-схемах блоков, представляющих трехстабильные элементы.

Методы. Предлагаются методы распознавания подсхем, представляющих собой элементы с тремя состояниями, в частности трехстабильные инверторы. Задача сводится к поиску сначала КМОП-подсхем и подсхем передаточных элементов, а затем структур инверторов на их основе.

Результаты. Разработаны программы на языке C++, реализующие методы извлечения элементов с тремя состояниями в плоском SPICE-описании транзисторной схемы и включения описаний соответствующих им блоков в иерархическое описание генерируемой логической сети.

Заключение. Разработанные программы поиска трехстабильных инверторов включены в программу декомпиляции транзисторных КМОП-схем и протестированы в ее составе на практических примерах схем транзисторного уровня.

Ключевые слова: обратное проектирование, экстракция транзисторных подсхем, КМОП-схема, проходная транзисторная логика, формат SPICE

Для цитирования. Черемисинова, Л. Д. Распознавание подсхем трехстабильных элементов в КМОП СБИС / Л. Д. Черемисинова // Информатика. – 2025. – Т. 22, № 4. – С. 82–93. – DOI: 10.37661/1816-0301-2025-22-4-82-93.

Конфликт интересов. Автор заявляет об отсутствии конфликта интересов.

Поступила в редакцию | Received 13.11.2025
Подписана в печать | Accepted 08.12.2025
Опубликована | Published 30.12.2025

Recognition of subcircuits of tristable elements in CMOS VLSI

Ljudmila D. Cheremisinova

*The United Institute of Informatics Problems
of the National Academy of Sciences of Belarus,
st. Surganova, 6, Minsk, 220012, Belarus
E-mail: cld@newman.bas-net.by*

Abstract

Objectives. The problem of extracting high-level structure at the level of logical elements from a transistor-level circuit is considered. Obtaining such a representation significantly reduces the execution time of VLSI topology verification at the design stage and serves as the basis for integrated circuit redesign. The goal of the study is to develop a method and software tools for extracting blocks representing tristable elements in CMOS circuits.

Methods. Methods for recognizing subcircuits representing tri-state elements, in particular tristable inverters, are proposed. The task is reduced to first searching for CMOS subcircuits and transmission gates subcircuits, and then for inverter structures based on them.

Results. C++ programs have been developed that implement methods for extracting three-state elements from a flat SPICE description of a transistor circuit and including descriptions of the corresponding blocks in the hierarchical description of the generated logical network.

Conclusion. The developed programs for searching the tristable inverters are included in the program for decompiling transistor CMOS circuits and tested as part of it on practical examples of transistor-level circuits.

Keywords: reverse engineering, transistor subcircuit extraction, CMOS circuit, pass transistor logic, SPICE format

For citation. Cheremisinova L. D. *Recognition of subcircuits of tristable elements in CMOS VLSI*. Informatika [Informatics], 2025, vol. 22, no. 4, pp. 82–93 (In Russ.). DOI: 10.37661/1816-0301-2025-22-4-82-93.

Conflict of interest. The author declares of no conflict of interest.

Введение. Современные сверхбольшие интегральные схемы (СБИС) могут содержать до миллиарда транзисторов, и их сложность быстро увеличивается. Рост размеров и сложности микросхем, которые можно разместить на одном кристалле, а также рост стоимости их проектирования привели к необходимости разработки средств анализа и тестирования топологии, позволяющих на ранних стадиях установить, соответствует ли описание проекта на уровне транзисторов исходной спецификации на проектирование [1]. Эффективным инструментом автоматизации тестирования и проектирования СБИС является обратное проектирование, или обратный инжиниринг (англ. hardware reverse engineering), схем на уровне транзисторов [1, 2], который позволяет заменить трудоемкое схемотехническое моделирования СБИС (или части СБИС) менее затратным логическим моделированием путем распознавания (экстракции) высокуюровневых структур в схемах на транзисторном уровне. Целью данного преобразования является замена представления схемы на транзисторном (низком) уровне функционально эквивалентным представлением на уровне логических элементов (более высокуюровневым) [3]. Для построения такого описания нужно выделить в схеме наборы взаимосвязанных транзисторов в качестве отдельных компонентов, т. е. найти подсхемы, представляющие собой логические элементы. Замена найденных транзисторных подсхем логическими элементами позволяет получить схему на логическом уровне, функционально эквивалентную исходной транзисторной схеме.

Операция, в результате которой по плоскому (одноуровневому) описанию транзисторной схемы строится иерархическое описание, инверсна задаче проектирования в смысле направления процесса преобразования от спецификации до реализации интегральной схемой. По аналогии с программированием этот процесс можно считать декомпиляцией описания транзисторной

схемы [4]. Структурному анализу цифровых схем путем их декомпиляции посвящены многие работы [5–7]. Если при декомпиляции СБИС известна библиотека логических элементов, которая была использована при компиляции топологии СБИС, то задача декомпиляции сводится к поиску этих подсхем в транзисторной схеме. В более сложном в теоретическом плане случае библиотека логических элементов неизвестна. Тогда задача сводится к распознаванию в транзисторной схеме подсхем, представляющих логические элементы, и формированию библиотеки элементов, встречающихся в схеме.

В работе [4] рассмотрена задача декомпиляции плоского описания КМОП-схемы с целью построения иерархического структурного описания, компоненты которого представляют собой логические вентили и их подсхемы. Предложены метод и программа распознавания в описании транзисторной схемы комплементарных (стандартных) КМОП-вентилей статического стиля как самого распространенного, а также элементы на основе проходной логики – передаточные вентили и схемы на их основе, такие как мультиплексоры. В процессе декомпиляции выполняются следующие основные этапы: выделение подсхем, представляющих передаточные вентили и схемы на их основе; распознавание подсхем, представляющих КМОП-вентили и псевдоэлементы; выделение групп функционально и топологически идентичных подсхем и формирование библиотеки элементов; выделение подсхем с обратными связями и построение логической сети. Исходным объектом для программы декомпиляции служит плоское (одноуровневое) описание КМОП-схемы в формате SPICE (Simulation Program with Integrated Circuit Emphasis – программа моделирования с акцентом на интегральные схемы) [1], который является одним из основных форматов для обмена электрическими схемами между системами автоматизации проектирования. Результат работы программы – иерархическое SPICE-описание, в которое включены модели всех идентифицированных при декомпиляции логических элементов.

Настоящая работа продолжает эти исследования. Предлагаются методы (и их практическая реализация) выявления в МОП-схеме подсхем, реализующих трехстабильные элементы, в частности трехпозиционные инверторы.

Задание транзисторных схем. При распознавании элементов в процессе декомпиляции анализируется основная часть описания транзисторной схемы в формате SPICE – Data Statements, которая описывает компоненты схемы в виде «моделей» и задает взаимосвязи между ними. Используются два типа моделей: модели устройств (device models) и модели подсхем (subcircuit models). Первые модели представляют собой, по сути, примитивные компоненты, такие как транзисторы, диоды и т. д. Модель устройства (.subckt в SPICE-описании) определяет схему непримитивного элемента, которая задается в виде блока в разделе Data Statements и описывает структуру соединений элементов (примитивных и непримитивных) в этой подсхеме. Тип модели используемого в SPICE-описании элемента задается первой буквой его имени. Например, имена примитивных элементов, являющихся МОП-транзисторами, должны начинаться с латинской буквы «M», непримитивных элементов (подсхем) – с буквы «X». Следует заметить, что описания на SPICE нечувствительны к регистру символов.

Схема задается списком соединений ее элементов, в котором указаны связи между элементами (а точнее, между их выводами), осуществляющиеся с помощью электрических цепей (nets). При этом для каждого элемента указываются соединения всех его выводов путем задания имен связанных с ними цепей.

Главной частью плоского описания схемы в формате SPICE (листинг 1) является список транзисторов, в котором для каждого из четырех выводов транзистора указано имя цепи, соединяющей его с остальными частями схемы:

M<name> <nd> <ng> <ns> <nb> <model-name>,

где name – имя транзистора (ему предшествует буква «m») в схеме; nd, ng, ns и nb – идентификаторы цепей, связанных с выводами стока (drain), затвора (gate), истока (source) и подложки (substrate) соответственно; model-name – имя модели (для n-МОП- или p-МОП-транзистора – это обычно nmos или pmos).

Например, плоское описание транзисторной схемы, приведенной на рис. 1, в формате SPICE представлено на листинге 1.

Листинг 1. SPICE-описание транзисторной схемы (рис. 1)

```
.GLOBAL Gnd Vdd
.SUBCKT GG0 in1 in2 in3 out
Mn1 out in3 t2 Gnd nmos
Mn2 t2 t3 Gnd Gnd nmos
Mn3 t3 in1 t4 Gnd nmos
Mn4 t4 in2 Gnd Gnd nmos
Mp1 out t3 Vdd Vdd pmos
Mp2 out in3 Vdd Vdd pmos
Mp3 t3 in1 Vdd Vdd pmos
.ENDS
Circuit GG0 contains 7 device instances.
  Class: pmos           instances: 3
  Class: nmos           instances: 4
Circuit contains 9 nets.
```

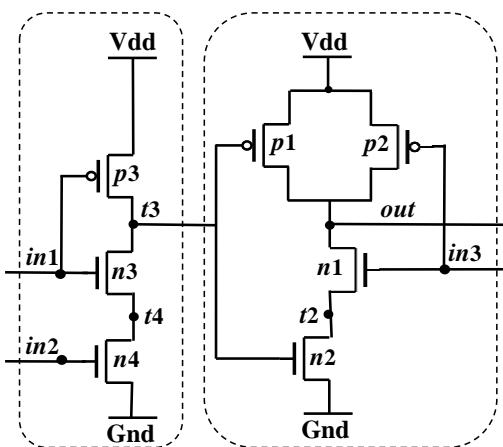


Рис. 1. Пример транзисторной схемы

Fig. 1. Transistor circuit example

Общая форма описания связей непримитивного элемента, модель которого представляется подсхемой с n выводами, в формате SPICE имеет вид

$X<\text{name}> <\text{P1}> <\text{P2}> \dots <\text{Pn}> <\text{model-name}>$,

где P_i – идентификаторы цепей, связанных с выводами элемента.

Схемы на основе комплементарной МОП-логики. Основным элементом для изготовления цифровых интегральных микросхем в настоящее время является полевой МОП-транзистор (структуря: металл-окисел-полупроводник) в силу высокой плотности интеграции и относительной простоты изготовления [8]. МОП-транзистор имеет четыре вывода: сток, затвор, исток и подложка. На переключательном уровне МОП-транзистор работает как управляемый ключ: если к затвору транзистора прикладывается напряжение, значение которого превышает пороговое, ключ замыкается и пропускает ток между истоком и стоком. В противном случае ток не течет и ключ является разомкнутым.

Доминирующим стилем логики при разработке современных цифровых СБИС является комплементарная МОП-структура – КМОП-структура, состоящая из МОП-транзисторов двух типов: n-МОП и p-МОП, различающихся типами носителей тока и уровнями сигнала, которые необходимо подать на затворы для протекания через них тока (и замыкания ключа). Элемент на основе стандартной (комплементарной) КМОП-логики включает две подсхемы, которые состоят соответственно из p-МОП- и n-МОП-транзисторов. Блоки обеспечивают связь выхода подсхемы элемента с источником питания Vdd, и тогда значение сигнала на выходе предполагается равным 1, или с землей Gnd в противном случае.

При распознавании КМОП-подсхем, реализующих логические вентили, используется структурный подход [9] к декомпиляции транзисторных схем, который сначала разбивает транзисторную схему на непересекающиеся подсхемы, представляющие группы связанных по постоянному току транзисторов. Одной из таких групп является подсхема из МОП-транзисторов с тремя типами внешних соединений: входы группы подаются только на затворы ее транзисторов, выходы группы подаются только на затворы транзисторов других групп, имеются связи транзисторов группы с шинами питания Vdd и земли Gnd. Например, схема, приведенная на рис. 1, содержит две такие группы, выделенные пунктирными линиями.

Среди найденных групп транзисторов находятся правильные подсхемы, являющиеся статическими КМОП-вентилями, значение сигнала на выходе которых представляется булевой функцией от сигналов, поступающих на выводы транзисторов (если игнорировать переходные эффекты во время переключения). Группа транзисторов, связанных по току, реализует КМОП-вентиль, если удовлетворяет следующим условиям: р-МОП-транзисторы образуют р-блок, расположенный между цепями питания Vdd и выхода группы; н-МОП-транзисторы образуют н-блок между цепями выхода группы и земли Gnd; р- и н-блоки группы имеют одинаковое количество транзисторов и реализуют взаимно инверсные функции. Каждая из подсхем, не распознанных как КМОП-вентиль, объявляется псевдоэлементом, его функциональное описание структурным методом декомпиляции в общем случае неопределено. Например, правая из двух выделенных на рис. 1 групп транзисторов, связанных по току, удовлетворяет приведенным условиям и представляет собой вентиль И-НЕ, реализующий функцию $f = \bar{f}_n = \overline{in_3 \wedge t_3}$. Левая группа транзисторов не распознается как КМОП-вентиль и объявляется псевдоэлементом (3, 4), состоящим из трех транзисторов и четырех цепей.

При декомпиляции множество распознанных КМОП-вентилей делится на классы функционально эквивалентных (реализующих одинаковые функции), а множество псевдоэлементов разбивается на классы элементов, характеризующихся одинаковыми парами чисел (p_i, p_n) транзисторов и цепей. Каждый из полученных классов элементов (и вентилей, и псевдоэлементов) разбивается затем на подклассы топологически эквивалентных [10]. Например, в результате декомпиляции плоского SPICE-описания (для транзисторной схемы на рис. 1), представленного на листинге 1, выделен двухходовый элемент G0, реализующий функцию И-НЕ, и псевдоэлемент P0_0 (3, 4). Полученное двухуровневое описание приведено на листинге 2.

Листинг 2. SPICE-описание двухуровневой транзисторной схемы

```
Contents of circuit cldG0.sp: Circuit: 'GG0'
Circuit GG0 contains 7 device instances.
  Class: pmos           instances: 3
  Class: nmos           instances: 4
Circuit contains 9 nets.
Connected Componens = 2
Invalid comps
Valid Components = 1
Pass fets = 0
Psevdo Componens = 1 nets =4
Unclassified fets = 0 nets = 0
(A AND B)           1
Psevdo
(3) (4) 1
Defining cell: GG0_gen
  Defining global node: Gnd
  Defining global node: Vdd
Start of Computation: 15h10m40s 22/08/2025
End of Computation: 15h10m40s 22/08/2025
Computation Time (s): 0.0051
.SUBCKT G0 A B Y
```

```

* (A AND B)
M1 Y A 2 Gnd nmos
M2 2 B Gnd Gnd nmos
M3 Y B Vdd Vdd pmos
M4 Y A Vdd Vdd pmos
.ENDS
.SUBCKT P0_0 P0 P1 P2
* (3) (4)
M1 P0 P1 2 Gnd nmos
M2 2 P2 Gnd Gnd nmos
M3 P0 P1 Vdd Vdd pmos
.MOS
.SUBCKT GG0_gen in1 in2 in3 out
XM0I1 in3 t3 out G0 Fets=nmosn1+nmosn2+pmosp1+pmosp2
XMP0I1 t3 in1 in2 P0_0 Fets=nmosn3+nmosn4+pmosp3
.ENDS

```

Достоинствами КМОП-стиля являются сравнительно небольшое рассеивание мощности, высокая степень интеграции (за счет плотной упаковки схем), низкое энергопотребление и масштабируемость напряжения питания. Недостаток заключается в значительном количестве транзисторов: для реализации элемента с k входами требуются $2k$ транзисторов, половину из которых составляют затратные по площади р-МОП-транзисторы.

Схемы на основе проходной транзисторной логики. Наряду со стандартной КМОП-логикой при построении интегральных схем часто используется МОП-логика на проходных транзисторах (pass transistor logic, PTL), которая позволяет существенно сократить число транзисторов, необходимых для реализации логических функций, а также снизить энергопотребление схемы [8].

Простейшая цифровая схема состоит из одного МОП-транзистора – управляемого двоичного ключа, представляющего собой двунаправленный элемент, проводимость которого управляетяется сигналом на выводе затвора. Например, n-МОП-транзистор (рис. 2, a) включается (и пропускает сигнал с вывода A на B), когда на его затвор подается сигнал высокого уровня ($C = 1$), и запирается (не пропускает ток), когда на его затвор поступает сигнал низкого уровня ($C = 0$). В последнем случае вывод B переходит в третье (высокоимпедансное) состояние. р-МОП-транзистор, наоборот, включается (и пропускает сигнал с вывода A на B) при сигнале $C = 0$ на его затворе и запирается при $C = 1$.

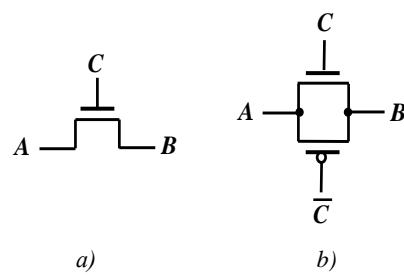


Рис. 2. Элементы проходной логики: a) n-МОП-транзистор; b) передаточный КМОП-вентиль
Fig. 2. Elements of pass transistor logic: a) n-MOS transistor; b) CMOS transmission gate

В проходной логике сигналы с входных портов схемы допускается использовать для питания не только затворов транзисторов, но и выводов стока и истока n-МОП- и р-МОП-транзисторов. На эти выводы могут подаваться внутренние сигналы схемы, что дает возможность достаточно просто реализовывать некоторые достаточно сложные логические функции. Логика на проходных транзисторах позволяет реализовывать более эффективно, чем стандартная КМОП-логика, схемы с большим количеством вентилей, такие как мультиплексоры, вентили исключающее ИЛИ, сумматоры. Однако транзисторы n-МОП- и р-МОП-типа по отдельности не являются идеально проводящими ключами при передаче сигналов логических 1 и 0. Транзистор n-МОП-типа

служит почти идеальным переключателем при передаче сигнала логического 0, но сильно снижает уровень напряжения, соответствующий логической единице. Аналогично p-MOP-транзистор при нулевом напряжении на затворе хорошо передает высокие уровни напряжения (сигналы логической 1), а при передаче низких уровней напряжения (сигналов логического 0) выходное напряжение может не снизиться до приемлемого уровня.

Распространенным методом устранения этих недостатков является использование не одного проходного транзистора, а передаточного вентиля, который состоит из пары транзисторов – n-MOP- и p-MOP-типа, связанных параллельно своими выводами истока и стока (рис. 2, б) [8, 11]. Передаточный вентиль действует как двунаправленный ключ, для которого используется параллельное управление: на затворы n-MOP- и p-MOP-транзисторов должны подаваться комплементарные сигналы C и \bar{C} . Если $C = 1$ (тогда $\bar{C} = 0$), оба транзистора открыты, позволяя сигналу пройти через элемент от вывода A к B или наоборот (от B к A), поскольку передача сигнала в таком элементе может идти в обоих направлениях и ни одно из них не является предпочтительным. В результате цепи A и B имеют одинаковый потенциал ($A = B$) при $C = 1$. В случае $A = 0$ вывод B замкнут на A через n-MOP-транзистор, а при $A = 1$ вывод B замкнут через p-MOP-транзистор. Если же $C = 0$ (тогда $\bar{C} = 1$), оба транзистора запираются, при этом вывод A не имеет связи с выводом B . Это свойство передаточного вентиля позволяет использовать его для построения схем с тремя состояниями, которые осуществляют блокировку передачи сигналов на некоторые участки схем.

Передаточные вентили, несмотря на проблемы, связанные с деградацией уровня сигнала при их последовательном соединении, широко используются при построении сложных цифровых микросхем, так как схемы на их основе состоят из меньшего количества транзисторов, работают быстрее и требуют меньше энергии, чем аналогичные схемы, реализованные на основе КМОП-логики. Логика на основе передаточных вентилей широко применяется, как правило, при проектировании сложных регулярных структур там, где существенна скорость работы и число транзисторов. Например, ее использование более предпочтительно по сравнению с элементами на основе стандартной КМОП-логики при проектировании схем мультиплексоров, регистров и сумматоров.

В процессе декомпиляции транзисторных структур подсхемы, представляющие передаточные вентили, находятся на этапе анализа SPICE-описания транзисторной схемы. На этом этапе собирается информация, связанная с каждым синтаксическим элементом анализируемой транзисторной схемы, и строится ее графовая модель. Передаточные элементы распознаются путем поиска пар связанных n-MOP- и p-MOP-транзисторов. В работе [12] предлагается метод поиска таких пар, который имеет линейную сложность. Он основан на построении и анализе хеш-таблицы, которая содержит информацию о связях всех примитивных элементов. Ключами таблицы являются хеши (последовательности битов), вычисленные по текстовым строкам, которые образованы из имен цепей, связанных с истоком и стоком каждого из транзисторов анализируемой схемы. Запись в таблице состоит из хеша (ссылки на запись в таблице имен элементов схемы) имени транзистора с такими связями истока и стока или из хешей имен транзисторов, если транзисторов с такими связями несколько.

В процессе просмотра SPICE-описания для каждого транзистора формируется текстовая строка T из имен цепей, связанных с его истоком и стоком. Вычисленный для транзистора хеш $h(T)$ является ключом хеш-таблицы H . Запись в таблице H , выбираемая по некоторому ключу $h(T)$, состоит из хешей имен тех транзисторов, для которых значение хеша $h(T_i)$ равно h . Таким образом, запись хеш-таблицы H задает связанные транзисторы, истоки (и стоки) которых связаны с одной и той же цепью. Ситуация, когда для очередного рассматриваемого транзистора сгенерирован ключ (хеш $h(T)$), для которого в хеш-таблице H уже есть вход, означает, что найдена пара связанных транзисторов, представляющая передаточный вентиль.

Для транзисторной схемы, фрагмент описания которой приведен на листинге 3, имеются два передаточных элемента, представленных парами связанных транзисторов: (4, 19), (5, 20). Имена транзисторов, образующих передаточные вентили, а также цепей, связанных с их стоками и истоками, выделены на листинге 3 жирным шрифтом.

Листинг 3. Фрагмент SPICE-описания транзисторной схемы

```
.SUBCKT rdrl r1 s1 c d vcc gnd q
M0 130 8 9 gnd MN
M1 gnd r1 130 gnd MN
M2 131 s1 gnd gnd MN
M3 13 g 131 gnd MN
M4 8 c 13 gnd MN
M5 d 10 8 gnd MN
M6 gnd c 10 gnd MN
.
.
.
M18 13 g vcc vcc MP
M19 8 10 13 vcc MP
M20 d c 8 vcc MP
M21 vcc c 10 vcc MP
M22 12 10 g vcc MP
.
```

После распознавания передаточных вентилей исходное плоское описание транзисторной схемы преобразуется в иерархическое описание, в котором появляется описание модели передаточного вентиля (см. рис. 2, б) с именем CN2 (листинг 4). Вентиль состоит из двух транзисторов и имеет семь выводов: А и В – имена цепей, связанных со стоками и истоками; NC (соответствует сигналу \bar{C}) и С – имена цепей, соединенных с затворами р-МОП- и н-МОП-транзисторов соответственно; PB и NB – имена цепей, связанных с их подложками. Фактически сначала указываются цепи, связанные со стоком, затвором и истоком р-МОП-транзистора, а затем цепь затвора н-МОП-транзистора. Пары транзисторов (5, 20) и (4, 19) из плоского SPICE-описания (листинг 3) заменяются в иерархическом описании передаточными элементами XM2I1 и XM2I4 типа CN2 (листинг 5).

Листинг 4. SPICE-описание передаточного вентиля

```
.SUBCKT CN2 A NC B C NB PB
M1 A NC B PB MP
M2 A C B NB MN
.ENDS
```

Листинг 5. Фрагмент SPICE-описания, задающий связи передаточных вентилей

```
XM2I1 d c 8 10 gnd vcc CN2 Fets=pmos20+nmox5
XM2I4 8 10 13 c gnd vcc CN2 Fets=pmos19+nmox4
```

Элементы с тремя состояниями выхода. Элементы с тремя состояниями на выходе применяются в цифровых системах [8, 13], когда необходимо передавать информацию от разных источников сигнала на одну и ту же схему или шину данных. Выходы таких элементов кроме обычных состояний 0 и 1 могут иметь «третье логическое состояние», называемое состоянием высокого импеданса (обозначаемое обычно как High-Z) и соответствующее высокому сопротивлению, которое эквивалентно электрически отключенному (открытым) состоянию выхода. Состояние высокого импеданса электрически изолирует устройства от линии передачи данных. Простейшие реализации элементов с тремя состояниями (трехстабильных элементов) рассмотрены выше и приведены на рис. 2.

В практических схемах для передачи данных от разных устройств с помощью трехстабильных элементов обычно используются трехпозиционные инверторы. На рис. 3 изображены две схемы инверторов с тремя состояниями. Инвертор первого типа реализован путем каскадирования КМОП-инвертора с передаточным вентилем (на основе проходной логики). Очевидно, что передаточный вентиль должен быть именно на выходе такого инвертора. Инвертор второго типа представляет собой псевдоэлемент (4, 5), состоящий из четырех транзисторов и пяти цепей.

Проводимость обоих инверторов зависит от значения сигнала S на их входе, который является для них управляющим. Схемы инверторов при высоком уровне управляющего сигнала ($S = 1$) работают как обычные инверторы значения сигнала In на их входах, т. е. реализуют на выходе

сигнал, инверсный входному ($Out = \bar{In}$). При низком уровне управляющего сигнала ($S = 0$) выход схем инверторов переходит в высокоимпедансное (высокоомное) состояние, когда сигнал на выходах их схем отсутствует.

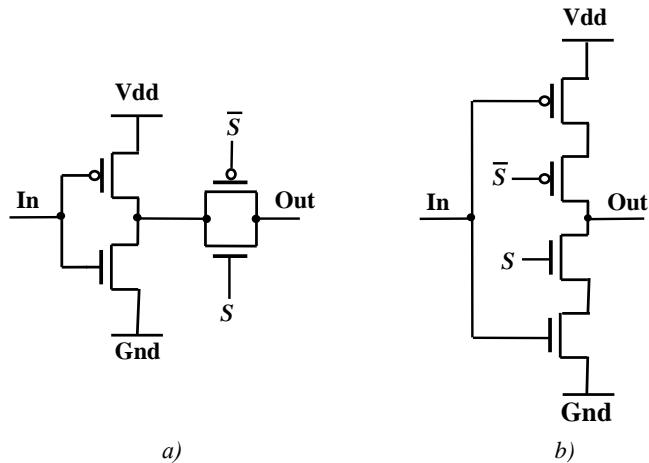


Рис. 3. Схемы инверторов с тремя состояниями на основе: a) передаточного вентиля; b) псевдоэлемента (4, 5)

Fig. 3. Three-state inverter circuits based on: a) transmission gate; b) pseudoelement

Распознавание инверторов с тремя состояниями выхода. В процессе декомпиляции плоского описания транзисторной схемы сначала выполняются разбор и анализ исходного SPICE-описания. На этой стадии, как говорилось выше, находятся и выделяются передаточные вентиля, а также схемы на их основе, такие как мультиплексоры и демультиплексоры. Передаточные вентиля, не вошедшие в состав этих элементов, являются кандидатами на вхождение в состав инверторов с тремя состояниями.

После выделения передаточных вентиляй и схем на их основе SPICE-описание транзисторной схемы становится двухуровневым. Первый уровень образуют модели найденных элементов и модель оставшейся части схемы, для которой генерируется графовое представление. Этот граф разбивается на связные подграфы. Им соответствуют подсхемы из связанных по току транзисторов, среди которых находятся КМОП-вентиля и псевдоэлементы [9]. Выделенные элементы группируются сначала на функционально эквивалентные (в случае КМОП-вентиляй) или структурно схожие (в случае псевдоэлементов должны быть равны пары (k, l) чисел транзисторов и цепей). Каждый из полученных таким образом классов элементов разбивается затем на подклассы топологически эквивалентных [10]. Данным этапом заканчивается формирование библиотеки элементов.

Схемы выделенных типов трехстабильных инверторов находятся уже после построения библиотеки распознанных в процессе декомпиляции элементов. Схема инвертора первого типа (рис. 3, a) находится как связанная пара «инвертор – передаточный вентиль»: выход КМОП-инвертора связан со входом передаточного вентиля одной цепью.

Схема трехстабильного инвертора второго типа (рис. 3, b) ищется среди распознанных псевдоэлементов (4, 5), имеющих четыре транзистора и пять цепей. Псевдоэлемент (4, 5) представляет собой инвертор с тремя состояниями в следующих случаях:

1) затворы пары транзисторов р-МОП- и н-МОП-типов, соединенных (выводами стоков или истоков) соответственно с цепями питания и земли, связаны с одной и той же цепью (цепью In на рис. 3, b);

2) на затворы остальных двух р-МОП- и н-МОП-транзисторов подаются взаимно инверсные сигналы (\bar{S} и S на рис. 3, b).

Например, на листинге 6 приведены SPICE-описания двух топологически разных псевдоэлементов GZ_0 (4, 5) и GZ_1 (4, 5), распознанных в описании транзисторной схемы CPU. При их анализе на предмет распознавания инвертора с тремя состояниями выделяются н-МОП- и р-МОП-транзисторы, истоки или стоки которых связаны с цепями питания (таких вариантов

может быть четыре, так как сток и исток каждого транзистора равноправны). Псевдоэлемент считается инвертором с тремя состояниями, если для него выполняется первое условие: затворы выделенных транзисторов связаны с одной и той же цепью (на рис. 3, б это цепь In).

Листинг 6. SPICE-описание псевдоэлементов (4, 5)

```
.SUBCKT GZ_0 Y A P2 P3
* A
M1 2 P2 Y GND N W=0.15u L=0.06u
M2 2 A GND GND N W=0.15u L=0.06u
M3 4 A VCC VCC P W=0.15u L=0.06u
M4 Y P3 4 VCC P W=0.15u L=0.06u
.EDNS
.SUBCKT GZ_1 Y A P2 P3
* A
M1 1 A GND GND N W=0.15u L=0.06u
M2 Y P2 1 GND N W=0.15u L=0.06u
M3 4 A VCC VCC P W=0.15u L=0.06u
M4 Y P3 4 VCC P W=0.15u L=0.06u
```

Псевдоэлемент GZ_1 (листинг 5) является инвертором с тремя состояниями, так как затворы его транзисторов 1 и 3 (n-МОП- и p-МОП-типов) соединены с одной и той же цепью A, а их истоки связаны соответственно с цепями земли и питания. Аналогично псевдоэлемент GZ_0 также является трехстабильным инвертором, так как на затворы транзисторов 2 и 3 поступает сигнал с одной и той же цепи A, а их истоки связаны с цепями земли и питания. Стоит заметить, что инверторы GZ_0 и GZ_1 функционально эквивалентны, но топологически неидентичны: выход Y инвертора GZ_0 связан с истоком транзистора 1, а инвертора GZ_1 – со стоком транзистора 2.

Второе условие, согласно которому сигнал, соответствующий цепи P2, является инверсным сигналу цепи P3, проверяется на основе анализа логической схемы, формируемой на последнем этапе декомпиляции транзисторной схемы.

Обсуждение практических результатов. Описанные в работе методы экстракции подсхем передаточных вентилей и псевдоэлементов реализованы в составе программы выполнения первого этапа декомпиляции плоских SPICE-описаний транзисторных схем. На этом этапе строится библиотека распознанных элементов и генерируется иерархическое SPICE-описание. На втором этапе декомпиляции в иерархическом SPICE-описании выделяются подсхемы инверторов с тремя состояниями и вентили исключающее ИЛИ, а затем строится логическая схема, реализующая исходную транзисторную схему [14].

В таблице приведены данные о примере декомпиляции плоского SPICE-описания транзисторной схемы, которое содержит значительное число трехстабильных инверторов на основе псевдоэлемента (4, 5). В некоторых строках таблицы приведены значения параметров через слеш. В строках «Типы КМОП-вентилей» и «Число типов псевдоэлементов» через слеш указаны числа разных классов функционально и топологически эквивалентных элементов. Например, «63/212» в строке «Типы КМОП-вентилей» означает, что найденные 54 323 КМОП-вентиля разделены на 63 класса функционально эквивалентных элементов. Каждый из этих классов в свою очередь разбит на подклассы топологически эквивалентных элементов, и общее число таких классов равно 212. В строке «Процессорное время декомпиляции» через слеш приведены значения времени выполнения первого и второго этапов декомпиляции.

На первом этапе декомпиляции рассматриваемой транзисторной схемы найдены 2392 псевдоэлемента (4, 5). Все эти псевдоэлементы на втором этапе декомпиляции распознаны как инверторы с тремя состояниями, разделенными на две группы топологически эквивалентных схем: GZ_1 (листинг 6) и GZ_0, содержащих 26 и 2366 инверторов соответственно.

Экспериментальные данные

Experimental Data

Категории <i>Categories</i>	Количество <i>Quantity</i>
Примитивные элементы	243 806
МОП-транзисторы	243 806
Цепи	123 017
Типы КМОП-вентилей	63/212
КМОП-вентили (всех типов)	54 323
Типы псевдоэлементов	1/2
Псевдоэлементы (всех типов)	2392
Типы инверторов с тремя состояниями	2
Инверторы с тремя состояниями	2392
Передаточные элементы	6902
Типы мультиплексоров	АА, ВВ
Мультиплексоры	1559
Типы триггеров	8
Триггеры (всех типов)	4924
Процессорное время декомпиляции	98,128/20,051

На этапе построения логической схемы на основе распознанных элементов в ней были найдены 4924 триггера восьми типов. Все найденные триггеры не имеют в своем составе мультиплексоров, но два типа триггеров (их всего 2366) содержат по одному трехстабильному инвертору типа GZ_0 с тремя состояниями.

Заключение. Предложенные методы распознавания подсхем инверторов с тремя состояниями реализованы на языке C++ как часть программы декомпиляции плоских описаний транзисторных схем в формате SPICE. Программа декомпиляции была протестирована на практических примерах описаний КМОП-схем. Декомпилированные схемы прошли проверку на соответствие исходной топологии транзисторной схемы с помощью разработанных средств верификации [20], а также средств Mentor Graphics Calibre nmLVS. В программе на каждом этапе декомпиляции используются внутренние представления транзисторной схемы, оптимальные с точки зрения требуемого объема памяти и скорости обработки. Испытания показали, что программа имеет достаточное быстродействие, чтобы обрабатывать схемы более чем со 100 тыс. транзисторов за несколько минут на ПЭВМ.

Список использованных источников

1. Baker, R. J. CMOS Circuit Design, Layout, and Simulation / R. J. Baker. – Third ed. – Wiley-IEEE Press, 2010. – 1214 p.
2. Hunt, V. D. Reengineering: Leveraging the Power of Integrated Product Development / V. D. Hunt. – Wiley, 1993. – 283 p.
3. Zhang, N. The subcircuit extraction problem / N. Zhang, D. C. Wunsch, F. Harary // Proceedings IEEE International Behavioral Modeling and Simulation Workshop. – 2005. – Vol. 33, no. 3. – P. 22–25.
4. Черемисинов, Д. И. Извлечение сети логических элементов из КМОП-схемы транзисторного уровня / Д. И. Черемисинов, Л. Д. Черемисинова // Микроэлектроника. – 2019. – Т. 48, № 3. – С. 224–234. – DOI: 10.1134/S0544126919030037.
5. Kundu, S. GateMaker: A transistor to gate level model extractor for simulation, automatic test pattern generation and verification / S. Kundu // Proc. of the Intern. Test Conf., Washington, USA, 18–23 Oct. 1998. – Washington, 1998. – P. 372–381.
6. Yang, L. FROSTY: A program for fast extraction of high-level structural representation from circuit description for industrial CMOS circuits / L. Yang, C.-J. R. Shi // Integration the VLSI Journal. – 2006. – Vol. 39, no. 4. – P. 311–339.
7. Han, M. Efficient subgraph matching: Harmonizing dynamic programming, adaptive matching order, and failing set together / M. Han, H. Kim, G. Gu // Proc. of Intern. Conf. on Management of Data (SIGMOD '19), Amsterdam, Netherlands, 30 June – 5 July 2019. – Amsterdam, 2019. – P. 1429–1446.
8. Рабай, Ж. М. Цифровые интегральные схемы. Методология проектирования / Ж. М. Рабай, А. Чандраксан, Б. Николич ; пер с англ. – 2-е изд. – М. : ООО «И.Д. Вильямс», 2007. – 912 с.

9. Черемисинов, Д. И. Распознавание логических вентилей в плоской транзисторной схеме / Д. И. Черемисинов, Л. Д. Черемисинова // Информатика. – 2021. – Т. 18, № 4. – С. 96–107. – DOI: 10.37661/1816-0301-2021-18-4-96-107.
10. Черемисинов, Д. И. Канонизация графов при декомпиляции транзисторных схем / Д. И. Черемисинов, Л. Д. Черемисинова // Информатика. – 2022. – Т. 19, № 3. – С. 25–39. – DOI: 10.37661/1816-0301-2022-19-3-25-39.
11. Combinational circuits using transmission gate logic for power optimization / G. N. Balaji, V. Aathira, K. Ambhikavathi [et al.] // International Research Journal of Engineering and Technology (IRJET). – 2016. – Vol. 3, no. 5. – Р. 649–654.
12. Черемисинов, Д. И. Методы структурного распознавания подсхем проходной транзисторной логики в КМОП-схемах / Д. И. Черемисинов, Л. Д. Черемисинова // Известия высших учебных заведений. Электроника. – 2025. – Т. 30, № 1. – С. 51–63. – DOI: 10.24151/1561-5405-2025-30-1-51-63.
13. Угрюмов, Е. П. Цифровая схемотехника / Е. П. Угрюмов. – СПб. : БХВ-Петербург, 2000. – 528 с.
14. Черемисинов, Д. И. Обратное проектирование СБИС для обеспечения безопасности аппаратуры / Д. И. Черемисинов, Л. Д. Черемисинова // Проблемы разработки перспективных микро- и наноэлектронных систем : сб. тр. / под общ. ред. акад. РАН А. Л. Стемпковского. – М. : ИППМ РАН, 2022. – Вып. III. – С. 10–17.

References

1. Baker R. J. *CMOS Circuit Design, Layout, and Simulation*, third edition. Wiley-IEEE Press, 2010, 1214 p.
2. Hunt V. D. *Reengineering: Leveraging the Power of Integrated Product Development*. Wiley, 1993, 283 p.
3. Zhang N., Wunsch D. C., Harary F. The subcircuit extraction problem. *Proceedings IEEE International Behavioral Modeling and Simulation Workshop*, 2005, vol. 33, no. 3, pp. 22–25.
4. Cheremisinov D. I., Cheremisinova L. D. Extracting a logic gate network from a transistor-level CMOS circuit. *Mikroelektronika [Russian Microelectronics]*, 2019, vol. 48, no. 3, pp. 224–234 (In Russ.). DOI: 10.1134/S0544126919030037.
5. Kundu S. GateMaker: A transistor to gate level model extractor for simulation, automatic test pattern generation and verification. *Proceedings of the International Test Conference, Washington, USA, 18–23 October 1998*. Washington, 1998, pp. 372–381.
6. Yang L., Shi C.-J. R. FROSTY: A program for fast extraction of high-level structural representation from circuit description for industrial CMOS circuits. *Integration the VLSI Journal*, 2006, vol. 39, no. 4, pp. 311–339.
7. Han M., Kim H., Gu G. Efficient subgraph matching: Harmonizing dynamic programming, adaptive matching order, and failing set together. *Proceedings of International Conference on Management of Data (SIGMOD '19), Amsterdam, Netherlands, 30 June – 5 July 2019*. Amsterdam, 2019, pp. 1429–1446.
8. Rabaev J. M., Chandrakasan A., Nikolic B. *Digital Integrated Circuits*, 2nd edition. Pearson, 2002, 800 p.
9. Cheremisinov D. I., Cheremisinova L. D. Logical gates recognition in a flat transistor circuit. *Informatika [Informatics]*, 2021, vol. 18, no. 4, pp. 96–107 (In Russ.). DOI: 10.37661/1816-0301-2021-18-4-96-107.
10. Cheremisinov D. I., Cheremisinova L. D. Canonization of graphs during transistor circuits decompilation. *Informatika [Informatics]*, 2022, vol. 19, no. 3, pp. 25–39 (In Russ.). DOI: 10.37661/1816-0301-2022-19-3-25-39.
11. Balaji G. N., Aathira V., Ambhikavathi K., Geethiga S., Havin R. Combinational circuits using transmission gate logic for power optimization. *International Research Journal of Engineering and Technology (IRJET)*, 2016, vol. 3, no. 5, pp. 649–654.
12. Cheremisinov D. I., Cheremisinova L. D. Recognition of pass transistor logic subcircuits in CMOS circuits. *Izvestija vysshih uchebnyh zavedenij. Jelektronika [Proceedings of Universities. Electronics]*, 2025, vol. 30, no. 1, pp. 51–63 (In Russ.). DOI: 10.24151/1561-5405-2025-30-1-51-63.
13. Uglyumov Ye. P. Cifrovaja shemotekhnika. *Digital Circuit Design*. Saint Petersburg, BHV-Peterburg, 2000, 528 p. (In Russ.).
14. Cheremisinov D. I., Cheremisinova L. D. *Reverse engineering of VLSI for equipment safety*. Problemy razrabotki perspektivnykh mikro- i nanoelektronnykh system [Problems of Developing Promising Micro- and Nanoelectronic Systems]. In A. L. Stempkovskij (ed.). Moscow, Institut problem proektirovaniya v mikroelektronike Rossijskoj akademii nauk, 2022, iss. III, pp. 10–17 (In Russ.).

Информация об авторе

Черемисинова Людмила Дмитриевна, доктор технических наук, профессор, главный научный сотрудник, Объединенный институт проблем информатики Национальной академии наук Беларусь.
E-mail: cld@newman.bas-net.by

Information about the author

Ljudmila D. Cheremisinova, D. Sc. (Eng.), Prof., Chief Researcher, The United Institute of Informatics Problems of the National Academy of Sciences of Belarus.
E-mail: cld@newman.bas-net.by

УЧЕНЫЕ БЕЛАРУСИ SCIENTISTS OF BELARUS

**Крючков Александр Николаевич
(к 75-летию со дня рождения)**

**Kruchkov Alexander Nikolaevich
(on the 75th anniversary of his birth)**



16 ноября 2025 г. заслуженному ученому в области цифровой картографии и геоинформатики, заведующему лабораторией картографических систем и технологий государственного научного учреждения «Объединенный институт проблем информатики Национальной академии наук Беларусь» (ОИПИ НАН Беларусь) Крючкову Александру Николаевичу исполнилось 75 лет.

А. Н. Крючков является известным ученым в области автоматизации создания и обновления цифровых карт, создания сложных картографических и геоинформационных систем различного назначения, обработки и распознавания картографических изображений по данным дистанционного зондирования Земли, автором более 110 научных работ, включая одну монографию и четыре учебных пособия. На основе разработанных им методов, моделей пространственной информации и технологий создан ряд автоматизированных картографических систем и комплексов, которые внедрены на предприятиях Беларусь, России и других стран СНГ, а работы, выполняемые под его руководством в интересах Министерства обороны Республики Беларусь (МО РБ), и в настоящее время имеют большое значение для обороноспособности нашей страны.

Научные интересы Александра Николаевича охватывают направления информатики, связанные с решением прикладных задач в области цифровой картографии, геоинформационных систем, обработки и анализа цифровых изображений местности по данным дистанционного зондирования Земли, и распространяются на разработку методов и алгоритмов автоматизации процессов создания, обновления и изготовления цифровых моделей местности, а также интеллектуальных технологий, моделей и алгоритмов поддержки принятия решений на основе свойств местности и данных дистанционного зондирования Земли.

Биография и научная деятельность. А. Н. Крючков родился в 1950 г. в городе Одесса. В 1973 г. закончил Минский радиотехнический институт и был направлен на работу в Научно-исследовательский институт средств автоматизации, в котором работал с 1973 по 1976 г. старшим инженером.

С 1976 по 1989 г. – ведущий конструктор, заведующий сектором, заведующий отделом Специального конструкторско-технологического бюро с опытным производством Института технической кибернетики АН БССР.

С 1989 по январь 2001 г. – заведующий отделом Института технической кибернетики АН БССР.

С 2001 г. по настоящее время – заведующий лабораторией картографических систем и технологий ОИПИ НАН Беларусь.

За время работы в институте Александр Николаевич проявил себя специалистом высокой квалификации. В течение 1984–1990 гг. являлся главным конструктором особо важных опытно-конструкторских работ (ОКР), выполняемых в интересах МО СССР, России, Украины и Беларусь. Под его руководством создан ряд автоматизированных картографических систем и технологий, предназначенных для создания, обновления и изготовления цифровых карт. Разработанные системы и технологии обладали высокой надежностью и новизной примененных научных и технических решений.

В настоящее время основное место работы Александра Николаевича – ОИПИ НАН Беларусь, общий стаж работы – 52 года. Созданный и возглавляемый им коллектив является одним из старейших в институте и в декабре 2025 г. отметит свое 50-летие. Александр Николаевич пользуется заслуженным уважением и признанием коллег на многих предприятиях Беларусь и далеко за ее пределами.

Основные научные результаты. Работы, выполненные А. Н. Крючковым в интересах МО, имеют большое значение для обороноспособности страны. В 1977 г. постановлением ЦК КПСС головным исполнителем по разработке и изготовлению комплекса технических и программных средств создания, хранения, обновления и выдачи цифровых карт местности и автоматизированного управления базой цифровых картографических данных (БЦКД) был определен Институт технической кибернетики АН БССР. Перед головным исполнителем были поставлены три важные научно-технические задачи:

- разработка методов и технологий цифрового моделирования рельефа земной поверхности и создание крупномасштабных цифровых карт местности;
- создание базы цифровых картографических данных;
- содержательная обработка изображений земной поверхности из космоса с целью обновления и актуализации цифровых карт местности и осуществления операций зондирования и мониторинга.

По большому счету речь шла о полном перевооружении всей военно-топографической службы Вооруженных Сил СССР на основе новейших технологий. Впервые в истории АН БССР академический институт брался за выполнение столь масштабной опытно-конструкторской разработки, результаты которой должны были непосредственно использоваться в системе обороны страны. Для работы приглашаются А. В. Старцев (главный конструктор комплексов АРМ-К), А. Н. Крючков, С. Б. Бочаров и другие молодые специалисты.

При создании комплекса БЦКД были разработаны: методы и средства комплексирования ЭВМ различных классов и периферийных устройств машинной графики (две высокопроизводительные ЭВМ, 20 ЭВМ СМ, свыше 90 графических терминалов) в единую интегрированную информационно-вычислительную систему, что дало возможность организовать поточную технологию создания, хранения, обновления, выдачи цифровых карт по тематическим запросам различных категорий пользователей, многоуровневую специализированную систему программного обеспечения, позволившую достичь высокой степени автоматизации процессов обработки больших (гигабайтных) массивов цифровых картографических данных.

В 1983 г. были завершены опытно-конструкторские работы и изготовлен головной образец комплекса БЦКД. На предприятии заказчика монтируются крупные вычислительные мощности, ведутся комплексирование и отладка прикладного и общесистемного программного обеспечения БЦКД в целом. Начинается опытное серийное производство крупномасштабных цифровых карт на большие территории и государственные испытания БЦКД. В апреле 1984 г. приказом министра обороны СССР комплекс БЦКД по итогам государственных испытаний был принят на вооружение Советской Армии. Это была победа всего коллектива института. За успешное выполнение задания указом Президиума Верховного Совета СССР в 1984 г. институт награжден орденом Трудового Красного Знамени. Ряд сотрудников, в том числе и А. Н. Крючков, непосредственно принимавших участие в работах по созданию комплекса БЦКД, были удостоены высоких государственных наград: награждены почетными грамотами, орденами и медалями.

Институт технической кибернетики АН БССР завоевал авторитет в стране своими работами в области картографии и обработки изображений, создания соответствующих математических и технических средств.

В период с 1977 по 1990 г. Александр Николаевич выполнил большой объем работ по развертыванию картографических производств в воинских частях Военно-топографической службы МО СССР (Голицыно, Ногинск, Симферополь, Киев). С 1986 по 1992 г. являлся главным конструктором систем и комплексов, разрабатываемых в интересах Военно-топографической службы МО СССР (ОКР «ТИПАЖ-90», «КАРТА-ИО»), Главного разведывательного управления МО СССР (ОКР «АРКА-90»), Главного управления навигации и океанографии МО СССР в части автоматизированных рабочих мест создания цифровых морских карт.

А. Н. Крючков внес большой личный вклад в разработку методов, алгоритмов и технологий создания цифровых моделей местности и их использования в системах управления войсками и оружием. Руководил рядом контрактов с Украиной, Россией, Индией, Китаем в области создания технологий и систем обработки картографической информации.

В 2010 г. созданные по контракту с МО Украины картографические комплексы «Карта-Ц» и «Издание-Ц» приняты на снабжение топографической службы Вооруженных сил Украины.

На протяжении последних 15 лет А. Н. Крючков является главным конструктором и научным руководителем работ в области цифровой картографии и геоинформатики в интересах МО Республики Беларусь и Государственного комитета по имуществу Республики Беларусь. Под его руководством выполнены такие ОКР, как «Система-ЦИМ» – разработка автоматизированной картографической системы распределенного банка данных цифровой информации о местности, «ГИС-ВН» – разработка комплекса поддержки принятия решений на основе геоинформационной системы военного назначения, «Формула-Т» – разработка программно-информационных комплексов для подвижного навигационно-топографического комплекса, «Фараон-К», «Составление-Ц» и ряд других ОКР и НИР.

Александр Николаевич Крючков являлся научным руководителем заданий в программах Союзного государства «КОСМОС-РБ», «КОСМОС-СГ», «КОСМОС-НТ», «МОНИТОРИНГ-СГ», а также ряда отдельных заданий государственных программ научных исследований.

Ряд программно-информационных комплексов используются в учебном процессе в БГУ, БГУИР, Военной академии Республики Беларусь.

Под руководством Александра Николаевича выполнены и выполняются ряд зарубежных контрактов и разработок в интересах гражданских организаций и ведомств. В 2012–2015 гг. выполнены работы по теме «Старт-5» по заказу МО Китая, где А. Н. Крючков руководил разработкой комплекса программных средств подготовки цифровых моделей местности для системы моделирования оперативной обстановки.

Разработанные под руководством А. Н. Крючкова картографические и геоинформационные системы и комплексы были приняты на вооружение в МО РБ и в промышленную эксплуатацию на предприятия Государственного комитета по имуществу Республики Беларусь. Промышленная эксплуатация комплекса «Составление-Ц» на предприятии «Белгеодезия» Государственного комитета по имуществу Республики Беларусь обеспечила увеличение эффективности составления топографических карт производного масштаба в среднем в пять–шесть раз.

Принятые на снабжение МО РБ комплексы «Издание-М» и «ГИС-ВН-2М» обеспечивают замещение зарубежных аналогов и имеют высокий потенциал по экспорту в страны СНГ. В настоящее время программно-информационный комплекс «Издание-М» поставлен в Армению, подготовлены договоры на поставку «Издание-М» и «ГИС-ВН-2М» в Узбекистан и Казахстан.

Высокий профессионализм и исключительная работоспособность Александра Николаевича позволяют ему постоянно находиться на переднем плане научных исследований, использовать инновации и активно обеспечивать выполнение проектов на самом высоком научном уровне. По результатам исследований А. Н. Крючкова опубликовано около 110 научных работ, докладов на международных и республиканских конференциях.

Александр Николаевич активно участвует в научно-организационной работе, является членом ученого и проблемного советов ОИПИ НАН Беларуси, членом межведомственной комиссии по вопросам топогеодезического и навигационного обеспечения военной организации государства при Министерстве обороны.



Премьер-министр Беларуси А. В. Кобяков вручил А. Н. Крючкову медаль «За трудовые заслуги» Указом Президента Республики Беларусь от 02.10.2015 г. № 411 за многолетний плодотворный труд и заслуги в развитии науки



Группа главных специалистов, непосредственно принимавших участие в разработке комплекса БЦКД, на праздновании 60-летия института 10 февраля 2025 г.
Слева направо: А. В. Старцев, А. К. Сутурин, А. Н. Крючков, С. А. Шавров

Награды и признание. Научные достижения Александра Николаевича многократно отмечены государственными наградами:

1984 г. – почетная грамота Верховного Совета БССР (указ Президиума Верховного Совета БССР).

1987 г. – почетные грамоты ЦК КПСС, Совета Министров СССР, ЦК ВЛКСМ.

1998, 2000, 2005 и 2014 гг. – почетные грамоты НАН Беларуси.

2007 г. – благодарность Председателя Президиума НАН Беларуси за разработку программно-информационных комплексов создания и обновления цифровых карт местности и внедрения картографических комплексов в различных отраслях народного хозяйства Беларуси и за рубежом.

2007 г. – медаль Ю. А. Гагарина Федерального космического агентства России.

2009 г. – нагрудный знак «Юбилейная медаль «В честь 80-летия НАН Беларуси».

2012 г. – ежемесячная персональная надбавка на 2012 г. за выдающийся вклад в социально-экономическое развитие республики (распоряжение Президента Республики Беларусь).

2015 г. – медаль «За трудовые заслуги» за многолетний плодотворный труд, заслуги в развитии науки (Указ Президента Республики Беларусь).

2018 г. – медаль «100 лет Навигационно-топографической службы Вооруженных Сил Республики Беларусь» за успехи в области научно-технических разработок.

2019 г. – почетная грамота Государственного комитета по имуществу Республики Беларусь.

2022 г. – юбилейная медаль «В память двухсотлетия корпуса военных топографов» (РФ).

2025 г. – нагрудный знак «За заслугі перад Нацыянальнай акадэміяй навук Беларусі».

В 2025 г. за активное участие в деятельности в интересах обеспечения обороноспособности Республики Беларусь и в связи с 60-летним юбилеем ГНУ «Объединенный институт проблем информатики Национальной академии наук Беларуси» награжден ценным подарком – макетом 7,62 мм пистолета ТТ (выписка из приказа начальника Генерального штаба Вооруженных Сил – первого заместителя министра обороны Республики Беларусь).

Юбилей Александра Николаевича – это повод вспомнить о его заслугах и возможность выразить ему благодарность за вклад в науку, труд и упорство, за умение поддерживать и вдохновлять на новые свершения не только словом, но и делом.

Компетентность и ответственность за порученное дело Александра Николаевича, доброжелательность и открытость в общении, прекрасные человеческие качества вызывают всеобщее уважение и признательность. От всего сердца желаем замечательному и талантливому человеку, заслуженному ученому Александру Николаевичу Крючкову крепкого здоровья, благополучия, долгих лет жизни, неиссякаемой энергии, новых творческих идей и реализованных проектов на благо Беларуси.

E. E. Сотикова, E. H. Афонова, B. K. Утекалко

Правила для авторов

Редакция журнала «Информатика» просит авторов руководствоваться приведенными ниже правилами.

I. Статьи принимаются в редакцию через электронную систему подачи по адресу <http://inf.grid.by> в формате файлов текстовых редакторов Microsoft Word. Объем оригинальной статьи – от 8 до 16 стр., включая рисунки, таблицы и достаточное количество наиболее актуальных ссылок; объем обзорной статьи – от 16 до 32 стр., включая все основные ссылки. Текст набирается с переносами, шрифт Times New Roman 11 пт, интервал между строками одинарный, абзацный отступ 0,5 см, поля по 2,5 см со всех сторон.

Материал статьи должен быть четко структурированным: Введение; основные разделы, в которых изложены цели и задачи, методы, результаты; Заключение (выводы).

II. Статьи о результатах работ, проведенных в научных учреждениях, должны иметь разрешение на публикацию (сопроводительное письмо за подпись руководителя или выписку из заседания ученого совета, отдела или кафедры, акт экспертизы).

III. Статьи в обязательном порядке должны включать аннотацию, ключевые слова, список литературы, информацию об авторах на русском и английском языках.

На заглавной странице располагаются следующие метаданные:

1. Индекс по универсальной десятичной классификации (УДК); на русском и английском языках тип статьи (оригинальная или обзорная), название статьи, инициалы и фамилии всех авторов, полное наименование учреждений, где работают авторы, с указанием почтового адреса, при наличии указывается ученая степень и ORCID, e-mail ответственного лица.

2. Аннотация (Abstract) объемом 150–250 слов в оригинальной статье должна быть структурирована отдельными подразделами: Цели, Методы, Результаты, Заключение, а также максимально характеризовать содержательную часть рукописи. Сюда не следует включать впервые введенные термины, аббревиатуры (за исключением общезвестных), ссылки на литературу.

3. Ключевые слова (Keywords) – наиболее значимые слова или словосочетания по теме работы, отражающие специфику темы, объекты и результаты исследования; перечень ключевых слов должен содержать 5–10 слов.

4. В разделе Благодарности (Acknowledgements) указываются все источники финансирования исследования, а также благодарности людям, которые участвовали в работе над статьей.

5. Автор обязан уведомить редакцию о реальном или потенциальном конфликте интересов, включив информацию в раздел Конфликт интересов (Conflict of interest).

6. Формулы, рисунки, таблицы в статье нумеруются в соответствии с порядком их упоминания в тексте. Ссылки на рисунки и таблицы в тексте обязательны. Рисунки должны быть выполнены с хорошим разрешением в масштабе, позволяющем четко различать надписи и обозначения. Цветные иллюстрации печатаются только в том случае, когда это необходимо для понимания излагаемого материала. Подрисуночные подписи с расшифровкой всех позиций, представленных на рисунке, и названия таблиц набираются шрифтом гарнитуры основного текста размером 9 пт. Перевод подрисуночной подписи и пояснений к рисунку, а также перевод названия таблицы, заголовки строк или столбцов располагаются курсивом после русскоязычной версии.

7. Набор формул выполняется в формульном редакторе Microsoft Equation или Math Type. Прямым шрифтом набираются: греческие и русские буквы; математические символы (\sin , \lg , ∞); символы химических элементов (C, Cl, CHCl₃); цифры (римские и арабские); индексы (верхние и нижние), являющиеся сокращениями слов. Курсивом набираются латинские буквы, символы физических величин (в том числе и в индексе).

8. Список использованной литературы оформляется в соответствии с требованиями Высшей аттестационной комиссии Республики Беларусь (ГОСТ 7.5–2008). Номер литературной ссылки в тексте дается порядковым номером в квадратных скобках. Ссылаясь на неопубликованные работы не допускается.

9. Отдельно оформляется References со следующей структурой: авторы (транслитерация), транслитерированное название монографии, *Перевод названия монографии на английский язык*. Выходные данные с обозначениями на английском языке. От транслитераций названий статей можно отказаться.

10. Ссылки на учебно-методическую литературу, ГОСТы, авторефераты, статистические отчеты в список не включаются, а оформляются в виде сносок (с подробными рекомендациями можно ознакомиться на сайте журнала в разделе Правила для авторов).

11. В разделе Информация об авторах (Information about the authors) приводятся ФИО авторов полностью, ученая степень, звание, должность, название организации, ORCID (при наличии).

IV. Все поступающие в редакцию рукописи проходят предварительную проверку на соответствие Правилам для авторов. Статья может быть возвращена автору на доработку с просьбой устранить недостатки или дополнить информацию. После проверки на соответствие правилам статья направляется рецензенту с указанием сроков рецензирования.

V. При наличии замечаний рецензента автору предоставляется определенное время на доработку рукописи. Статьи, направляемые на доработку, должны быть возвращены в исправленном виде с ответами на все замечания. Окончательное решение о публикации или отклонении рукописи принимается редакцией журнала. При положительном заключении рецензента статья передается редактору для подготовки к печати. Редакция оставляет за собой право на редакционные изменения, не искажающие основное содержание статьи.

VI. Редакция журнала предоставляет возможность первоочередного опубликования статей, представленных лицами, которые осуществляют послевузовское обучение (аспирантура, докторантуре, соискательство) в год завершения обучения.

VII. Авторы несут ответственность за направление в редакцию статей, уже опубликованных ранее или принятых к публикации другими изданиями.

ИНДЕКСЫ

00827

для индивидуальных
подписчиков

008272

для предприятий
и организаций

4 | 2025

том ■ vol. 22

информатика ■ INFORMATICS

ISSN 1816-0301 (Print)



9 771816 030000