



RUSSIAN TECHNOLOGICAL JOURNAL

**РОССИЙСКИЙ
ТЕХНОЛОГИЧЕСКИЙ
ЖУРНАЛ**

*Информационные системы.
Информатика.
Проблемы информационной безопасности*

*Роботизированные комплексы и системы.
Технологии дистанционного зондирования и неразрушающего контроля*

Современные радиотехнические и телекоммуникационные системы

*Микро- и наноэлектроника.
Физика конденсированного состояния*

Аналитическое приборостроение и технологии

Математическое моделирование

*Экономика наукоемких и высокотехнологичных
предприятий и производств. Управление в организационных системах*

Управление качеством продукции. Стандартизация

Мировоззренческие основы технологии и общества



RUSSIAN TECHNOLOGICAL JOURNAL

РОССИЙСКИЙ ТЕХНОЛОГИЧЕСКИЙ ЖУРНАЛ

- Информационные системы. Информатика. Проблемы информационной безопасности
- Роботизированные комплексы и системы. Технологии дистанционного зондирования и неразрушающего контроля
- Современные радиотехнические и телекоммуникационные системы
- Микро- и нанoeлектроника. Физика конденсированного состояния
- Аналитическое приборостроение и технологии
- Математическое моделирование
- Экономика наукоемких и высокотехнологичных предприятий и производств. Управление в организационных системах
- Управление качеством продукции. Стандартизация
- Мировоззренческие основы технологии и общества
- Information systems. Computer sciences. Issues of information security
- Multiple robots (robotic centers) and systems. Remote sensing and nondestructive testing
- Modern radio engineering and telecommunication systems
- Micro- and nanoelectronics. Condensed matter physics
- Analytical instrument engineering and technology
- Mathematical modeling
- Economics of knowledge-intensive and high-tech enterprises and industries. Management in organizational systems
- Product quality management. Standardization
- Philosophical foundations of technology and society

Russian Technological Journal
2025, том 13, № 3

Russian Technological Journal
2025, Vol. 13, No. 3

Russian Technological Journal 2025, том 13, № 3

Дата опубликования 30 мая 2025 г.

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

Периодичность: один раз в два месяца.

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

Учредитель и издатель:

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

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

Главный редактор:

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

Редакция:

Зав. редакцией	к.т.н. Г.Д. Середина
Научный редактор	д.т.н., проф. Г.В. Куликов
Выпускающий редактор	А.С. Алексеенко
Технический редактор	Д.В. Трофимова

119571, г. Москва, пр-т Вернадского, 86, оф. Р-108.
Тел.: +7 (499) 600-80-80 (#31288).
E-mail: seredina@mirea.ru.

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

Индекс по объединенному каталогу «Пресса России» 79641.

<https://www.rtj-mirea.ru>

Russian Technological Journal 2025, Vol. 13, No. 3

Publication date May 30, 2025.

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

Periodicity: bimonthly.

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

Founder and Publisher:

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

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

Editor-in-Chief:

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

Editorial staff:

Managing Editor	Cand. Sci. (Eng.) Galina D. Seredina
Scientific Editor	Dr. Sci. (Eng.), Prof. Gennady V. Kulikov
Executive Editor	Anna S. Alekseenko
Technical Editor	Darya V. Trofimova

86, Vernadskogo pr., Moscow, 119571 Russia.
Phone: +7 (499) 600-80-80 (#31288).
E-mail: seredina@mirea.ru.

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

The subscription index of *Pressa Rossii*: 79641.

Редакционная коллегия

Кудж Станислав Алексеевич	д.т.н., профессор, ректор РТУ МИРЭА, Москва, Россия. Scopus Author ID 56521711400, ResearcherID AAG-1319-2019, https://orcid.org/0000-0003-1407-2788 , rector@mirea.ru
Банис Юрас Йонович	хабилированный доктор наук, профессор, проректор Вильнюсского университета, Вильнюс, Литва. Scopus Author ID 7003687871, juras.banyas@ff.vu.lt
Бетелин Владимир Борисович	академик Российской академии наук (РАН), д.ф.-м.н., профессор, научный руководитель Федерального научного центра «Научно-исследовательский институт системных исследований» РАН, Москва, Россия. Scopus Author ID 6504159562, ResearcherID J-7375-2017, betelin@niisi.msk.ru
Боков Алексей Алексеевич	д.ф.-м.н., старший научный сотрудник, химический факультет и 4D LABS, Университет Саймона Фрейзера, Ванкувер, Британская Колумбия, Канада. Scopus Author ID 35564490800, ResearcherID C-6924-2008, http://orcid.org/0000-0003-1126-3378 , abokov@sfu.ca
Вахрушев Сергей Борисович	д.ф.-м.н., профессор, заведующий лабораторией нейтронных исследований Физико-технического института им. А.Ф. Иоффе РАН, профессор кафедры Физической электроники СПбГПУ, Санкт-Петербург, Россия. Scopus Author ID 7004228594, ResearcherID A-9855-2011, http://orcid.org/0000-0003-4867-1404 , s.vakhrushev@mail.ioffe.ru
Гуляев Юрий Васильевич	академик РАН, д.ф.-м.н., профессор, научный руководитель Института радиотехники и электроники им. В.А. Котельникова РАН, Москва, Россия. Scopus Author ID 35562581800, gulyaev@cplire.ru
Жуков Дмитрий Олегович	д.т.н., профессор кафедры телекоммуникаций Института радиоэлектроники и информатики РТУ МИРЭА, Москва, Россия. Scopus Author ID 57189660218, zhukov_do@mirea.ru
Кимель Алексей Вольдемарович	к.ф.-м.н., профессор, Университет Радбауд, г. Наймеген, Нидерланды. Scopus Author ID 6602091848, ResearcherID D-5112-2012, a.kimel@science.ru.nl
Крамаров Сергей Олегович	д.ф.-м.н., профессор, Сургутский государственный университет, Сургут, Россия. Scopus Author ID 56638328000, ResearcherID E-9333-2016, https://orcid.org/0000-0003-3743-6513 , mavoo@yandex.ru
Новиков Дмитрий Александрович	академик РАН, д.т.н., директор Института проблем управления им. В.А. Трапезникова РАН, Москва, Россия. Scopus Author ID 7102213403, ResearcherID Q-9677-2019, https://orcid.org/0000-0002-9314-3304 , novikov@ipu.ru
Перно Филипп	Dr. Sci. (Electronics), профессор, Центральная Школа г. Лилль, Франция. Scopus Author ID 7003429648, philippe.pernod@ec-lille.fr
Романов Михаил Петрович	д.т.н., профессор, научный руководитель Института искусственного интеллекта РТУ МИРЭА, Москва, Россия. Scopus Author ID 14046079000, https://orcid.org/0000-0003-3353-9945 , m_romanov@mirea.ru
Савиных Виктор Петрович	академик РАН, Дважды Герой Советского Союза, д.т.н., профессор, президент Московского государственного университета геодезии и картографии, Москва, Россия. Scopus Author ID 5641283700, vp@miigaik.ru
Соболевский Андрей Николаевич	д.ф.-м.н., директор Института проблем передачи информации им. А.А. Харкевича, Москва, Россия. Scopus Author ID 7004013625, ResearcherID D-9361-2012, http://orcid.org/0000-0002-3082-5113 , sobolevski@iitp.ru
Сюй Ли Да	академик Европейской академии наук, Российской инженерной академии и Инженерной академии Армении, Dr. Sci. (Systems Science), профессор, Университет Олд Доминион, Норфолк, Соединенные Штаты Америки. Scopus Author ID 13408889400, https://orcid.org/0000-0002-5954-5115 , lxu@odu.edu
Харин Юрий Семенович	академик Национальной академии наук Беларуси, д.ф.-м.н., профессор, директор НИИ прикладных проблем математики и информатики Белорусского государственного университета, Минск, Беларусь. Scopus Author ID 6603832008, http://orcid.org/0000-0003-4226-2546 , kharin@bsu.by
Чаплыгин Юрий Александрович	академик РАН, д.т.н., профессор, член Отделения нанотехнологий и информационных технологий РАН, президент Института микроприборов и систем управления им. Л.Н. Преснухина НИУ «МИЭТ», Москва, Россия. Scopus Author ID 6603797878, ResearcherID B-3188-2016, president@miet.ru
Шпак Василий Викторович	к.э.н., зам. министра промышленности и торговли Российской Федерации, Министерство промышленности и торговли РФ, Москва, Россия; доцент, Институт микроприборов и систем управления им. Л.Н. Преснухина НИУ «МИЭТ», Москва, Россия, mishinevaiv@minprom.gov.ru

Editorial Board

Stanislav A. Kudzh	Dr. Sci. (Eng.), Professor, Rector of RTU MIREA, Moscow, Russia. Scopus Author ID 56521711400, ResearcherID AAG-1319-2019, https://orcid.org/0000-0003-1407-2788 , rector@mirea.ru
Juras Banys	Habilitated Doctor of Sciences, Professor, Vice-Rector of Vilnius University, Vilnius, Lithuania. Scopus Author ID 7003687871, juras.banys@ff.vu.lt
Vladimir B. Betelin	Academician at the Russian Academy of Sciences (RAS), Dr. Sci. (Phys.-Math.), Professor, Supervisor of Scientific Research Institute for System Analysis, RAS, Moscow, Russia. Scopus Author ID 6504159562, ResearcherID J-7375-2017, betelin@niisi.msk.ru
Alexei A. Bokov	Dr. Sci. (Phys.-Math.), Senior Research Fellow, Department of Chemistry and 4D LABS, Simon Fraser University, Vancouver, British Columbia, Canada. Scopus Author ID 35564490800, ResearcherID C-6924-2008, http://orcid.org/0000-0003-1126-3378 , abokov@sfu.ca
Sergey B. Vakhrushev	Dr. Sci. (Phys.-Math.), Professor, Head of the Laboratory of Neutron Research, A.F. Ioffe Physico-Technical Institute of the RAS, Department of Physical Electronics of St. Petersburg Polytechnic University, St. Petersburg, Russia. Scopus Author ID 7004228594, ResearcherID A-9855-2011, http://orcid.org/0000-0003-4867-1404 , s.vakhrushev@mail.ioffe.ru
Yury V. Gulyaev	Academician at the RAS, Dr. Sci. (Phys.-Math.), Professor, Academic Supervisor of V.A. Kotelnikov Institute of Radio Engineering and Electronics of the RAS, Moscow, Russia. Scopus Author ID 35562581800, gulyaev@cplire.ru
Dmitry O. Zhukov	Dr. Sci. (Eng.), Professor of the Department of Telecommunications, Institute of Radio Electronics and Informatics, RTU MIREA, Moscow, Russia. Scopus Author ID 57189660218, zhukov_do@mirea.ru
Alexey V. Kimel	PhD (Phys.-Math.), Professor, Radboud University, Nijmegen, Netherlands, Scopus Author ID 6602091848, ResearcherID D-5112-2012, a.kimel@science.ru.nl
Sergey O. Kramarov	Dr. Sci. (Phys.-Math.), Professor, Surgut State University, Surgut, Russia. Scopus Author ID 56638328000, ResearcherID E-9333-2016, https://orcid.org/0000-0003-3743-6513 , mavoo@yandex.ru
Dmitry A. Novikov	Academician at the RAS, Dr. Sci. (Eng.), Director of V.A. Trapeznikov Institute of Control Sciences, Moscow, Russia. Scopus Author ID 7102213403, ResearcherID Q-9677-2019, https://orcid.org/0000-0002-9314-3304 , novikov@ipu.ru
Philippe Pernod	Dr. Sci. (Electronics), Professor, Dean of Research of Centrale Lille, Villeneuve-d'Ascq, France. Scopus Author ID 7003429648, philippe.pernod@ec-lille.fr
Mikhail P. Romanov	Dr. Sci. (Eng.), Professor, Academic Supervisor of the Institute of Artificial Intelligence, RTU MIREA, Moscow, Russia. Scopus Author ID 14046079000, https://orcid.org/0000-0003-3353-9945 , m_romanov@mirea.ru
Viktor P. Savinykh	Academician at the RAS, Dr. Sci. (Eng.), Professor, President of Moscow State University of Geodesy and Cartography, Moscow, Russia. Scopus Author ID 56412838700, vp@miigaik.ru
Andrei N. Sobolevski	Professor, Dr. Sci. (Phys.-Math.), Director of Institute for Information Transmission Problems (Kharkevich Institute), Moscow, Russia. Scopus Author ID 7004013625, ResearcherID D-9361-2012, http://orcid.org/0000-0002-3082-5113 , sobolevski@iitp.ru
Li Da Xu	Academician at the European Academy of Sciences, Russian Academy of Engineering (formerly, USSR Academy of Engineering), and Armenian Academy of Engineering, Dr. Sci. (Systems Science), Professor and Eminent Scholar in Information Technology and Decision Sciences, Old Dominion University, Norfolk, VA, the United States of America. Scopus Author ID 13408889400, https://orcid.org/0000-0002-5954-5115 , lxu@odu.edu
Yury S. Kharin	Academician at the National Academy of Sciences of Belarus, Dr. Sci. (Phys.-Math.), Professor, Director of the Institute of Applied Problems of Mathematics and Informatics of the Belarusian State University, Minsk, Belarus. Scopus Author ID 6603832008, http://orcid.org/0000-0003-4226-2546 , kharin@bsu.by
Yuri A. Chaplygin	Academician at the RAS, Dr. Sci. (Eng.), Professor, Member of the Departments of Nanotechnology and Information Technology of the RAS, President of the National Research University of Electronic Technology (MIET), Moscow, Russia. Scopus Author ID 6603797878, ResearcherID B-3188-2016, president@miet.ru
Vasily V. Shpak	Cand. Sci. (Econ.), Deputy Minister of Industry and Trade of the Russian Federation, Ministry of Industry and Trade of the Russian Federation, Moscow, Russia; Associate Professor, National Research University of Electronic Technology (MIET), Moscow, Russia, mishinevaiv@minprom.gov.ru

Содержание

Информационные системы. Информатика. Проблемы информационной безопасности

- 7** *В.А. Нечаев, С.В. Косяков*
Метод конвертации акцента с клонированием голоса в реальном времени на основе неавторегрессионной нейросетевой модели
- 21** *Д.В. Радюш*
Методы интеграции знаний для разработки вопросно-ответных систем
- 44** *И.Е. Тарасов, П.Н. Советов, Д.В. Люлява, Н.А. Дуксин*
Управление топологическими ограничениями при реализации конвейерных вычислительных структур на базе программируемых логических интегральных схем

Роботизированные комплексы и системы. Технологии дистанционного зондирования неразрушающего контроля

- 54** *Ю.А. Быковцев, В.М. Лохин*
Анализ и синтез интеллектуальных систем автоматического управления с нечетким регулятором I рода
- 63** *А.А. Коник, Р.Е. Афонин, И.М. Климов, Д.С. Зинченко*
Идентификационный комплекс обнаружения и подавления беспилотных воздушных судов на основе видеоперехвата для использования в условиях боевых действий

Современные радиотехнические и телекоммуникационные системы

- 73** *Б.М. Вовшин, А.А. Пушков, Е.М. Халтурина*
Об эквивалентности характеристик и «зеркальности» построения традиционных и ММО радиолокационных станций при параллельном обзоре пространства на основе антенных решеток
- 84** *Е.А. Рябов, А.А. Андреев, С.А. Сергеев, А.И. Михайлов*
Управление амплитудно-частотной характеристикой узкополосного фильтра для X-диапазона частот на основе фотонного кристалла с подвижным цилиндрическим дефектом
- 92** *А.Д. Ярлыков, О.А. Демин*
Разработка сверхвысокочастотного фильтра нижних частот на основе проекционной модели микрополосковой линии

Микро- и нанoeлектроника. Физика конденсированного состояния

- 103** *П.С. Кузнецов*
Микроэлектромеханические системы: путь к совершенствованию гироскопов
- 122** *А.Н. Юрасов, Р. Кулгунина, М.М. Яшин, М.А. Симдянова*
Особенности магниторефрактивного эффекта в нанокompозитах Co–Si

Contents

Information systems. Computer sciences. Issues of information security

- 7** *Vladimir A. Nechaev, Sergey V. Kosyakov*
Accent conversion method with real-time voice cloning based on a non-autoregressive neural network model
- 21** *Daniil V. Radyush*
Knowledge injection methods in question answering
- 44** *Ilya E. Tarasov, Peter N. Sovietov, Daniil V. Lulyava, Nikita A. Duksin*
Method for designing specialized computing systems on the basis of hardware and software cooptimization

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

- 54** *Yuri A. Bykovtsev, Valery M. Lokhin*
Analysis and synthesis of intelligent automatic control systems with type-1 fuzzy regulator
- 63** *Alexey A. Konik, Roman E. Afonin, Ivan M. Klimov, Dmitry S. Zinchenko*
Identification system for detecting and suppressing unmanned aerial vehicles based on video interception for use in combat conditions

Modern radio engineering and telecommunication systems

- 73** *Boris M. Vovshin, Alexander A. Pushkov, Elizaveta M. Khalturina*
On the equivalence of characteristics and specularity in the construction of traditional and MIMO radars with a parallel view of space based on antenna arrays
- 84** *Evgeny A. Ryabov, Anton A. Andreev, Sergey A. Sergeev, Alexander I. Mikhailov*
Control of the frequency response of a narrow-band filter for the X-band frequency based on a photonic crystal with a movable cylindrical defect
- 92** *Alexey D. Yarlykov, Oleg A. Demin*
Development of a microwave low-pass filter based on a microstrip line projection model

Micro- and nanoelectronics. Condensed matter physics

- 103** *Pavel S. Kuznetsov*
Microelectromechanical systems for improved gyroscope design
- 122** *Alexey N. Yurasov, Regina Kulgunina, Maxim M. Yashin, Marina A. Simdyanova*
Features of the magnetorefractive effect in Co–Si nanocomposites

УДК 004.522

<https://doi.org/10.32362/2500-316X-2025-13-3-7-20>

EDN PVYBDD



НАУЧНАЯ СТАТЬЯ

Метод конвертации акцента с клонированием голоса в реальном времени на основе неавторегрессионной нейросетевой модели

В.А. Нечаев[@],
С.В. Косяков

*Ивановский государственный энергетический университет имени В.И. Ленина, Иваново,
153003 Россия*

[@] Автор для переписки, e-mail: nechaev@gapps.ispu.ru

• Поступила: 23.07.2024 • Доработана: 03.02.2025 • Принята к опубликованию: 26.03.2025

Резюме

Цели. В настоящее время при разработке моделей для преобразования речи с акцентом в речь без акцента используются архитектуры глубоких нейросетей, а также ансамбли предобученных нейросетей для распознавания и генерации речи. При этом доступ к реализациям таких моделей является ограниченным, что затрудняет их применение, изучение и дальнейшее развитие. Также использование данных моделей ограничено особенностями архитектуры, которая не позволяет гибко менять тембр генерируемой речи и требует накопления контекста, что ведет к увеличению задержки при генерации и делает данные системы непригодными для использования в сценариях коммуникации двух и более людей в реальном времени. В связи с этим актуальной задачей и целью настоящей работы является разработка метода, позволяющего на основе входной речи с акцентом генерировать речь без акцента с минимальными задержками с возможностью сохранения, клонирования и модификации тембра говорящего, что позволит преодолеть ограничения текущих моделей.

Методы. Применены методы модификации, обучения и объединения глубоких нейросетей в единую сквозную архитектуру для прямого преобразования речи в речь. Для обучения использованы оригинальные и модифицированные наборы данных из открытых источников.

Результаты. Разработан метод конвертации акцента с клонированием голоса в реальном времени на основе неавторегрессионной нейросетевой модели, которая состоит из модулей определения акцента и пола, идентификации говорящего, преобразования речи в фонетическое представление, генерации спектрограммы и декодирования полученной спектрограммы в аудиосигнал. Метод демонстрирует высокое качество конвертации акцента с сохранением оригинального тембра, а также низкие задержки при генерации, приемлемые для использования в сценариях реального времени.

Выводы. Апробация разработанного метода подтвердила эффективность предложенной неавторегрессионной нейросетевой архитектуры. Разработанная прикладная нейросетевая модель продемонстрировала возможность работы в информационных системах на английском языке в режиме реального времени.

Ключевые слова: конвертация акцента, генерация речи, распознавание речи, конвертация голоса, машинное обучение, нейронная сеть

Для цитирования: Нечаев В.А., Косяков С.В. Метод конвертации акцента с клонированием голоса в реальном времени на основе неавторегрессионной нейросетевой модели. *Russian Technological Journal*. 2025;13(3):7–20. <https://doi.org/10.32362/2500-316X-2025-13-3-7-20>, <https://www.elibrary.ru/PVYBDD>

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

Авторы заявляют об отсутствии конфликта интересов.

RESEARCH ARTICLE

Accent conversion method with real-time voice cloning based on a non-autoregressive neural network model

Vladimir A. Nechaev[@],
Sergey V. Kosyakov

Ivanovo State Power Engineering University, Ivanovo, 153003 Russia

[@] Corresponding author, e-mail: nechaev@gapps.ispu.ru

• Submitted: 23.07.2024 • Revised: 03.02.2025 • Accepted: 26.03.2025

Abstract

Objectives. The development of contemporary models for the conversion of accents in foreign languages utilizes deep neural network architectures, as well as ensembles of neural networks for speech recognition and generation. However, restricted access to implementations of such models limits their application, study, and further development. Moreover, the use of these models is limited by their architectural features, which prevents flexible changes from being carried out in the timbre of the generated speech and requires the accumulation of context, leading to increased delays in generation, making these systems unsuitable for use in real-time multi-user communication scenarios. Therefore, the relevant task and aim of this work is the development of a method that generates native-sounding speech based on input accented speech material with minimal delays and the capability to preserve, clone, and modify the timbre of the speaker's voice.

Methods. Methods for modifying, training, and combining deep neural networks into a single end-to-end architecture for direct speech-to-speech conversion are applied. For training, original and modified open-source datasets were used.

Results. The work resulted in the development of a real-time accent conversion method with voice cloning based on a non-autoregressive neural network. The model comprises modules for accent and gender detection, speaker identification, speech conversion, spectrogram generation, and decoding the resulting spectrogram into an audio signal. As well as demonstrating high accent conversion quality while maintaining the original timbre, the short generation times of the applied method make it acceptable for use in real-time scenarios.

Conclusions. Testing of the developed method confirmed the effectiveness of the proposed non-autoregressive neural network architecture. The developed model demonstrated the ability to work in real-time information systems in English.

Keywords: accent conversion, speech synthesis, text-to-speech, voice conversion, machine learning, neural network

For citation: Nechaev V.A., Kosyakov S.V. Accent conversion method with real-time voice cloning based on a non-autoregressive neural network model. *Russian Technological Journal*. 2025;13(3):7–20. <https://doi.org/10.32362/2500-316X-2025-13-3-7-20>, <https://www.elibrary.ru/PVYBDD>

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

The authors declare no conflicts of interest.

ВВЕДЕНИЕ

Одним из важнейших каналов взаимодействия бизнеса со своими клиентами являются коммуникации посредством голосового общения. Это подтверждается развитием и широким применением колл-центров, которые в настоящее время могут создаваться и работать на межнациональном и межгосударственном уровнях. Для преодоления межязыкового барьера часто используется английский язык, при том, что операторы и клиенты колл-центров могут не являться носителями этого языка. В результате в практике работы таких колл-центров возникают ситуации, когда клиент отказывается от общения из-за сложности взаимопонимания с оператором, что приводит к экономическим потерям бизнеса. С развитием систем искусственного интеллекта для решения данной проблемы стали применяться программные системы конвертации акцента, которые позволяют в определенной степени уменьшать акцент говорящего [1]. Такие системы могут также применяться в процессе обучения иностранному языку [2, 3], для переозвучивания и улучшения качества ранее записанной речи [4], для улучшения качества распознавания речи существующих систем [5]. Несмотря на то, что в этой области исследований выполнено достаточно большое количество разработок, проблема совершенствования и улучшения качества работы таких систем остается актуальной.

Акцент в речи является неотъемлемой особенностью произношения, его можно разделить на родной, зависящий от множества региональных и культурных факторов, и иностранный [6]. При этом иностранный акцент от родного отличается на сегментном (фонемы) и суперсегментном (интонации, ударения, ритмика) уровнях [7]. Иностранный акцент проявляется, когда носитель одного языка (L1-речь) говорит на другом – неродном или втором для него языке (L2-речь) [8]. L2-речь может быть менее разборчивой для носителей, чем аналогичная по содержанию L1-речь [9], что может привести к снижению уровня понимания и доверия к сказанному, негативному отношению к говорящему и его дискриминации со стороны носителей [10–12].

Ранние методы конвертации акцента на этапе генерации требуют наличия эталонных L1-примеров,

соответствующих L2-речи [13–16]. Для каждой L2-фразы необходима соответствующая L1-фраза. Практическое применение таких моделей ограничено ввиду недостаточности данных для покрытия всех возможных речевых вариаций. Такие подходы требуют значительных ресурсов для сбора и обработки данных, что увеличивает время и стоимость разработки таких систем. Кроме того, строгая привязка к парным примерам может снижать универсальность и масштабируемость технологии, ограничивая ее способность адаптироваться к новым акцентам или речевым стилям, которые не были включены в изначальный набор данных.

В следующих разработках это ограничение преодолено – эталонные примеры на этапе вывода не требуются [17–22]. Однако для тренировки моделей [17, 20] используются параллельные наборы данных, содержащие аналогичные L2 и L1-фразы. Получение достаточного объема таких данных является затруднительным и дорогостоящим процессом. Также методы [17, 20] используют авторегрессионные рекуррентные нейросети, что усложняет процесс их тренировки. Метод [19] использует предварительно обученные нейросети для преобразования текста в речь. Для сохранения индивидуальных голосовых характеристик потребуется тренировать отдельную модель для каждого целевого говорящего, что делает затруднительным многопользовательское использование. Методы [21, 22] предсказывают длительность каждой генерируемой фонемы, что в итоге меняет исходную длительность L2-речи и идентичность говорящего, а также требует накопления контекста, что увеличивает время генерации и усложняет использование в реальном времени. Метод [18] лишен недостатков, перечисленных выше, однако реализация модели ограничена конечным набором акцентов, при этом идентификатор акцента должен быть определен на этапе тренировки модели. Это усложняет процесс подготовки данных и применение модели с акцентами, которые не были в них представлены ранее.

Целью данного исследования является разработка метода конвертации акцента, в котором будут преодолены вышеперечисленные проблемы и недостатки. Он должен обеспечивать преобразование любой речи из L2 в L1 без использования эталонных примеров и параллельных данных на этапах обучения

и генерации, что значительно упрощает, удешевляет и ускоряет процесс адаптации системы к новым акцентам.

1. ЗАДАЧИ ИССЛЕДОВАНИЯ

Носители языка воспринимают, с одной стороны, грамотность и беглость речи, с другой – акцент, как отдельные сущности, улучшение одной из них приводит к улучшению общего восприятия L2-речи носителями [9, 23]. При этом абсолютно точное воспроизведение L1-речи носителем языка на практике сложно достижимо из-за различий в фонетической интерференции и восприятии речи носителями языка [24]. При решении задачи конвертации акцента важно сохранять индивидуальные голосовые особенности говорящего (тембр, высота, громкость), т.е. необходимо производить клонирование голоса [25] при модификации связанных с иностранным акцентом и произношением сегментных и суперсегментных характеристик [13, 17, 18]. Это особенно важно в ситуациях, когда необходимо сохранить эмоциональную окраску, выразительность и индивидуальные особенности речи, особенности голоса, относящиеся к полу говорящего.

Для выполнения этих условий необходимо использовать несколько взаимосвязанных, объединенных в единую сквозную архитектуру моделей для определения акцента и пола, идентификации говорящего (ИГ), преобразования речи в фонетическое представление (РВФ), генерации спектрограммы и декодирования полученной спектрограммы в аудиосигнал.

Также, в сценариях реального времени, например, при голосовом общении людей с использованием высокоскоростных каналов связи, необходимо обеспечить минимальные задержки на генерацию и передачу обработанной речи [26–28]. Перевод акцента должен выполняться в режиме реального времени без использования рекуррентных сетей [29], чтобы избежать эффекта накопления ошибок, связанного с последовательной генерацией вывода.

На этапе тренировки должны использоваться общедоступные открытые данные для обучения систем распознавания и генерации речи. Также метод должен быть независим от наличия эталонных примеров и параллельных данных на этапах обучения и вывода.

2. МЕТОДЫ ИССЛЕДОВАНИЯ

2.1. Архитектура разработанной системы

Разработанный метод конвертации акцента включает в себя несколько взаимосвязанных, объединенных в единую сквозную архитектуру моделей для определения акцента и пола, ИГ,

преобразования РВФ, генерации спектрограммы и декодирования полученной спектрограммы в аудиосигнал. На рис. 1 представлена общая схема взаимодействия указанных моделей на этапе вывода (генерации выходного L1-аудио).

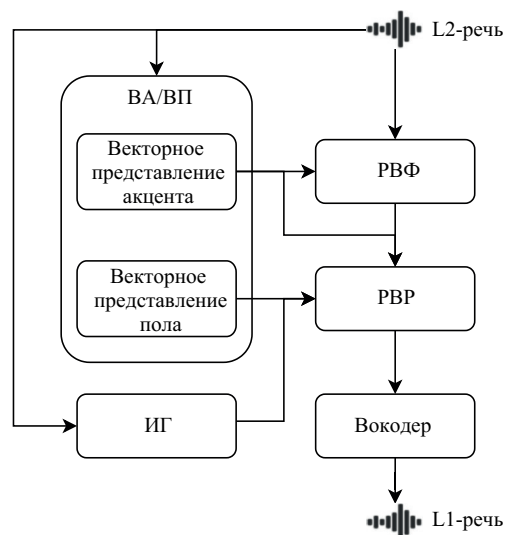


Рис. 1. Общая схема вывода метода конвертации акцента с клонированием голоса

Аудиосигнал L2-речи подается на вход модели РВФ, на вход модели определения и векторизации акцента и пола (ВА/ВП) и на вход модели ИГ. Векторное представление акцента влияет на генерацию фонетического представления, которое в векторной форме подается на вход модели преобразования речи в речь (РВП) и генерации мел-спектрограммы. Также на вход РВП-модели подаются векторные представления акцента (ВА) и пола (ВП), а также выход ИГ-модели, представляющий собой векторное представление индивидуальных голосовых характеристик (тембра). Полученная спектрограмма преобразуется в аудиосигнал L1-речи с помощью декодирующей модели вокодера.

Общий конвейер генерации L1-речи из исходной L2-речи может быть упрощенно представлен в виде формулы:

$$a_{L1} = F_V(F_{RVP}(F_{RVF}(a_{L2}, F_{BA}(a_{L2})), F_{BA}(a_{L2}), F_{ВП}(a_{L2}), F_{ИГ}(a_{L2}))), \quad (1)$$

где a_{L1} – сгенерированный аудиосигнал L1-речи; a_{L2} – входной аудиосигнал L2-речи; F_V – модель вокодера; F_{RVP} – РВП-модель; F_{RVF} – РВФ-модель; F_{BA} – ВА/ВП-модель, ВА; $F_{ВП}$ – ВА/ВП-модель, ВП; $F_{ИГ}$ – ИГ-модель, векторное представление индивидуальных голосовых характеристик.

Для получения единой сквозной модели конвертации акцента необходимо последовательно выполнять процесс тренировки каждой модели.

Так, ВА/ВП- и ИГ-модели не зависят от других моделей и их тренировку можно проводить в любом порядке. На этапе получения РВФ-модели потребуется вывод готовой ВА/ВП-модели. Для получения РВР-модели необходимы все предыдущие модели (ВА/ВП, ИГ, РВФ). Для тренировки модели вокодера требуется вывод РВР-модели.

2.2. ВА/ВП-модель

Для получения векторов фиксированной длины, представляющих свойства акцента и пола говорящего, модель сначала обучалась для решения задачи классификации. В такой конфигурации в процессе обучения используются метки классов, которые модель возвращает на выходе последнего слоя, а векторные представления, используемые в качестве голосовых характеристик, берутся со специального промежуточного слоя.

В данной и остальных моделях применяется пре-процессор на основе быстрого преобразования Фурье, который переводит входящий аудиосигнал (временная область) в мел-спектрограмму (частотная область), показывающую частотное содержание аудиосигнала на перцептивной мел-шкале, которая аппроксимирует нелинейную частотную характеристику человеческого уха. Частота дискретизации (семплирования) – 22050 Гц, ширина окна – 1024 звуковых фрагмента (семпла), шаг окна – 256 семплов, количество генерируемых мел-диапазонов – 80.

На рис. 2 представлена схема тренировки модели определения акцента и пола. Она содержит блоки сверточной сети архитектуры Jasper конфигурации 3×3 [30]. Декодер акцента и декодер пола имеют одинаковую архитектуру и состоят из слоя объединения (attention pooling layer) [31], слоя нормализации, сверточного слоя для получения ВА и ВП размерности 192, а также линейного слоя для получения (предсказания) класса акцента (КА) и пола (КП).

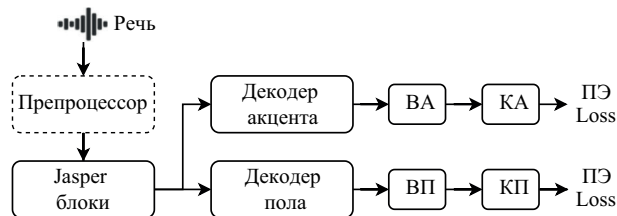


Рис. 2. Схема тренировки ВА/ВП-модели.
ПЭ – перекрестные энтропии

Аудиосигнал поступает в препроцессор, далее мел-спектрограмма поступает в Jasper блоки и, параллельно, в декодер акцента и декодер пола, а также соответствующие полносвязные слои на выходе для получения векторов предсказаний акцента и пола. В процессе обучения модели минимизируется сумма ПЭ:

$$L_{\text{ВА, ВП}}(x_a, y_a, x_g, y_g) = - \sum_{i=1}^A y_{a_i} \ln \left(\frac{\exp x_{a_i}}{\sum_{k=1}^A \exp x_{a_k}} \right) - \sum_{j=1}^G y_{g_j} \ln \left(\frac{\exp x_{g_j}}{\sum_{l=1}^G \exp x_{g_l}} \right), \quad (2)$$

где $L_{\text{ВА, ВП}}$ – общая функция потерь ВА/ВП-модели, A – количество КА (40), G – количество КП (2), x_a – предсказания акцента, x_g – предсказания пола, y_a – реальные метки акцента, y_g – реальные метки пола.

2.3. Модель ИГ

На рис. 3 представлена схема тренировки ИГ-модели и векторного представления тембра. Она содержит входящую сверточную нейросеть архитектуры SincNet [32], слои модели X-Vectors DNN¹ [33], слой для получения векторных представлений размерности 512.

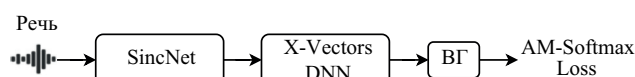


Рис. 3. Схема тренировки ИГ-модели.
ВГ – векторное представление индивидуальных голосовых характеристик говорящего

В отличие от ВА/ВП-модели, предварительное преобразование аудиосигнала в мел-спектрограмму не производится, т.е. оцифрованный звуковой сигнал во временной области с частотой дискретизации 16000 Гц поступает на полосовые фильтры архитектуры SincNet, далее – на сверточные слои X-Vectors DNN и выходной полносвязный слой, дающий на выходе векторное представление индивидуальных голосовых характеристик (тембра). В процессе тренировки модели решается задача обучения представлениям (representation learning) и минимизируется функция потерь Additive Angular Margin (AAM Loss) [34].

2.4. Модель преобразования РВФ

Следующим этапом является распознавание речи с учетом акцента говорящего. Для этого необходимо получить модель преобразования речи в фонетическое или текстовое представление. Схема тренировки РВФ-модели показана на рис. 4. На ней пунктирной линией обозначены блоки, которые фиксируются в процессе обратного распространения ошибки, т.е. веса в данных блоках не обновляются, а используются их заранее полученные состояния.

¹ Deep neural network.

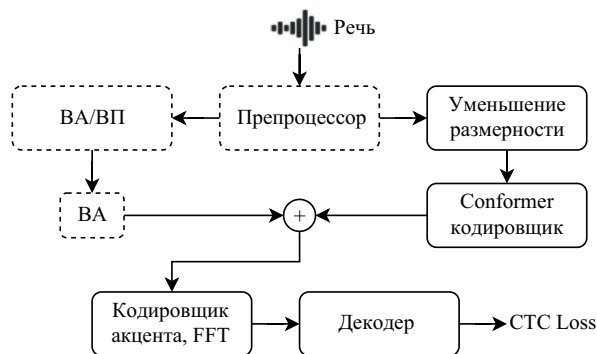


Рис. 4. Схема тренировки РВФ-модели

Речевой аудиосигнал поступает на вход препроцессора, описанного ранее, и далее параллельно в ВА/ВП-модель для получения ВА и в блок сверточного уменьшения размерности (Subsampler) с фактором 4 с дальнейшим преобразованием с помощью Conformer кодировщика, который представляет собой 12 модулей архитектуры Conformer [35] с внутренней размерностью 512, состоящих из полносвязных [36], сверточных [37] и трансформерных слоев с механизмом внимания [38]. Затем ВА нормализуется, приводится к размерности 512, суммируется с выходным сигналом Conformer кодировщика и поступает на вход кодировщика акцента, который имеет архитектуру одного стека трансформера прямой связи (feed-forward transformer, FFT) [39]. Выходной сигнал кодировщика акцента в дальнейшем используется в РВР-модели, как распределение фонетических токенов. В заключении выходной сигнал кодировщика акцента поступает на декодер, имеющий однослойную сверточную архитектуру с многопеременной логистической функцией активации (Softmax), формирующий на выходе вектор предсказаний текстовых токенов размерности, равной размеру словаря токенизатора (128) плюс один (для пустого «blank» токена). В процессе обучения модели минимизируется функция, которая вычисляет потери между непрерывным (несегментированным) временным рядом и целевой последовательностью (CTC Loss) [40]:

$$L_{\text{РВФ}}(x, y) = -\ln \left(\sum_{\rho \in A_{x,y}} \prod_{t=1}^T x_{\rho_t} \right), \quad (3)$$

где $L_{\text{РВФ}}$ – функция потерь РВФ-модели (Connectionist Temporal Classification (CTC) Loss), x – предсказанные моделью вероятности текстовых токенов, y – последовательность текстовых токенов из целевого текста, ρ – путь выравнивания x предсказаний для сокращения до y последовательности путем удаления всех пустых («blank») токенов и слияния повторяющихся токенов, $A_{x,y}$ – множество всех возможных путей выравнивания, T – количество

предсказанных токенов в x , x_{ρ_t} – вероятность конкретного предсказанного токена на шаге t при выбранном пути выравнивания ρ .

2.5. Модель преобразования РВР и генерации спектрограммы

Предыдущие модели объединяются в единую архитектуру для преобразования РВР и генерации спектрограммы. На рис. 5 представлена схема ее тренировки. В состав РВР-модели входят рассмотренные ранее блоки препроцессора, определения акцента, пола (ВА/ВП-модель) и тембра говорящего (ИГ-модель) с соответствующими модулями векторных представлений (ВА, ВП, ВГ), а также блок преобразования РВФ (РВФ-модель). Все эти блоки обозначены пунктирной линией – их обучение было выполнено ранее и не проводится на этапе тренировки РВР-модели. Также в архитектуру добавлен необучаемый блок на основе нормализованной взаимно корреляционной функции (normalized cross correlation function) и медианного сглаживания для извлечения основной или самой низкой частоты периодического звукового сигнала (F0), которая воспринимается человеческим ухом как высота тона [41, 42].

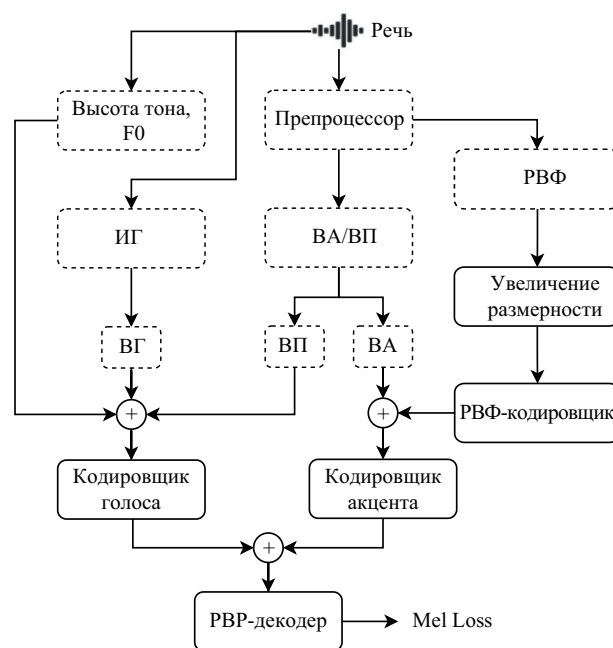


Рис. 5. Схема тренировки РВР-модели

Речевой аудиосигнал поступает на препроцессор, блок высоты тона и вход ИГ-модели. Мел-спектрограмма с препроцессора подается на входы ВА/ВП- и РВФ-моделей. Фонетическое представление из РВФ-модели поступает в блок увеличения размерности (upsampler) с фактором 4, чтобы выровнять исходную и генерируемые спектрограммы, состоящий из двух сверточных

1D-транспонированных слоев и двух функций активации выпрямителя (rectified linear unit, ReLU), располагающихся после каждого сверточного слоя. После блока увеличения размерности фонетическое представление преобразуется с помощью РВФ-кодировщика, который имеет архитектуру шести стеков трансформера прямой связи (FFT) [39], применяющегося в архитектуре Fastpitch в качестве входного блока, работающего в области токенов [43], с внутренней и внешней размерностями, соответственно, 1536 и 384. Векторные представления акцента, пола, тембра говорящего и профиля высоты тона нормализуются и приводятся к размерности 384. Далее векторы акцента и выход РВФ кодировщика суммируются и подаются на вход кодировщика акцента (1 стек FFT). Аналогично суммируются векторы высоты тона, тембра, пола и поступают на вход кодировщика голоса (1 стек FFT). Таким образом, кодировщик голоса агрегирует свойства речи, относящиеся к индивидуальным голосовым характеристикам кроме акцента, за который, в свою очередь, отвечает кодировщик акцента. Сумма выходных векторов кодировщика голоса и кодировщика акцента поступает на вход РВР-декодера, состоящего из 6 стеков FFT архитектуры Fastpitch из выходной мел-области [43]. В завершении, вектор проецируется в размерность 80 для соответствия исходному количеству мел-диапазонов. В процессе обучения минимизируется функция потерь на основе среднеквадратической ошибки:

$$L_{\text{РВР}}(x, y) = \frac{1}{N} \sum_{i=1}^N d_i (y_i - x_i)^2, \quad (4)$$

где $L_{\text{РВР}}$ – функция потерь РВР-модели (Mel Loss), N – количество элементов в мел-спектрограмме, x – предсказанная моделью мел-спектрограмма, y – целевая мел-спектрограмма, d – маска длительности спектрограммы для сбора в партию (batch) фиксированного размера, состоящая из значений 1 («элемент нужно учитывать») и 0 («элемент учитывать не нужно»), полученная из длительности предсказанной спектрограммы.

2.6. Модель генерации звукового сигнала из мел-спектрограммы (вокодер)

Мел-спектрограмма L1-речи в частотной области, полученная с помощью РВР-модели, преобразуется в звуковой сигнал во временной области. Для этого используется модель на основе генеративно-состязательных сетей HiFi-GAN [44]. Аудиосигнал на выходе имеет частоту дискретизации 22050 Гц. Обучение модели проводится следующим образом:

аудиосигнал из обучающего набора данных конвертируется в мел-спектрограмму с помощью РВР-модели, далее полученная спектрограмма передается в вокодер и преобразуется в звуковой сигнал. С помощью полученного и исходного аудиосигналов рассчитываются функции потерь для генератора и дискриминатора, описанные в [44].

2.7. Упрощенная модель преобразования РВР и генерации спектрограммы (Ablation)

С целью проведения сравнительных экспериментов также была разработана упрощенная версия модели конвертации акцента, схема которой представлена на рис. 6.

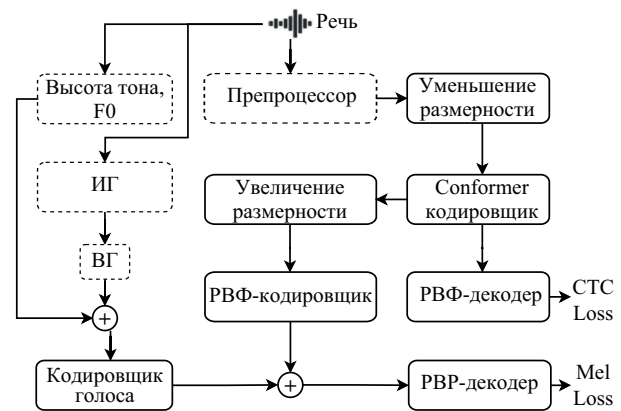


Рис. 6. Схема упрощенной модели конвертации акцента

Из данной упрощенной модели исключена ВА/ВП-модель, а также все связанные с ней кодировщики в РВФ- и РВР-моделях. Таким образом, в полученной упрощенной модели выходной сигнал не обуславливается свойствами акцента и пола. Кроме этого, обучение РВФ-модели проводилось не отдельно, а одновременно с РВР-моделью без фиксации весов РВФ с минимизацией суммы функций потерь CTC Loss и Mel Loss.

3. ПРАКТИЧЕСКОЕ ПРИМЕНЕНИЕ МЕТОДА

3.1. Обучение моделей

Обучение ВА/ВП-модели проводилось на следующих наборах данных: CMU-ARCTIC [45], L2 ARCTIC [46], Speech Accent Archive [47], Common Voice 16.1 [48]. Все они представляют собой аудиозаписи речи на английском языке, соответствующие им текстовые транскрипции, а также содержат дополнительную метаинформацию об акценте, поле и, в некоторых случаях, о родном языке, месте проживания и возрасте говорящего. Используя эту информацию, аудиофайлы были сгруппированы

в 40 классов, обозначающих родной или иностранный английский акцент, например, британский, американский, русский, индийский и южноазиатский, канадский, немецкий, австралийский, африканский, японский, восточноевропейский и т.д. Также выделен пол говорящего. Общая длительность размеченных таким образом аудиофайлов составила для тренировочной выборки – 1087.6 ч, для валидационной и тестовой – 7.6 ч.

Для обучения ИГ-модели использованы коллекции данных VoxCeleb1 [49] и VoxCeleb2 [50] общей продолжительностью 2794 ч. Эти наборы представляют собой сгруппированные аудиозаписи речи 7363 людей. Аудиозаписи, относящиеся к одному человеку, подаются при обучении как положительные примеры и, наоборот, относящиеся к разным людям, как негативные примеры.

Модель РВФ тренировалась на данных CMU-ARCTIC [45], L2 ARCTIC [46], Common Voice 16.1 [48], LibriSpeech [51], NPTEL2020², VCTK [52], GigaSpeech [53]. Указанные наборы состоят из аудиозаписей англоязычной речи с разными акцентами и соответствующих текстовых транскрипций. Общая длительность объединенной тренировочной выборки – 6107 ч, валидационной – 48 ч. Текстовые транскрипции были нормализованы, т.е. преобразованы из канонической письменной формы в устную [54], что особенно важно для чисел и аббревиатур, а также были приведены к единому виду: переведены в нижний регистр в ASCII-формат, удалена пунктуация, специальные символы и дополнительные отступы. На тренировочной части текстов был обучен токенизатор SentencePiece [55] с размером словаря 128, с помощью которого в процессе тренировки и оценки модели обрабатываются все тексты.

Обучение РВР-модели и вокодера производилось с использованием следующих наборов данных: CMU-ARCTIC [45], L2 ARCTIC [46], VCTK [52], LibriTTS-R [56], LJ Speech³. При разделении данных на тренировочную и валидационную выборки их длительность составила 681 ч и 17.6 ч соответственно. В процессе тренировки используется только аудиоинформация без текстовой разметки.

Обучение упрощенной модели (Ablation) проводилось на данных для РВФ- и РВР-моделей.

Для обучения, оценки и использования описанных моделей разработан код с использованием библиотек с открытым исходным кодом Pytorch [57]

² NPTEL2020 – Indian English Speech Dataset. <https://github.com/AI4Bharat/NPTEL2020-Indian-English-Speech-Dataset>. Дата обращения 01.05.2024. / Accessed May 01, 2024.

³ Ito K., Johnson L. *The LJ Speech Dataset*. <https://keithito.com/LJ-Speech-Dataset/>. Дата обращения 01.05.2024. / Accessed May 01, 2024.

и NVIDIA NeMo [58]. Реализация и веса модели векторного представления тембра говорящего (ИГ-модель) взяты из библиотеки Pyannote [59]. Обучение проводилось на сервере с 8 графическими ускорителями (graphics processing unit, GPU) NVIDIA Tesla V100.

Обучение ВА/ВП-модели проводилось с использованием оптимизатора SGD со скоростью обучения (learning rate) $1 \cdot 10^{-3}$, коэффициентом регуляризации (weight decay) $2 \cdot 10^{-4}$, коэффициентом инерции (momentum) 0.9 и планировщиком Cosine Annealing в течение 200 эпох. Для тренировки моделей РВФ и РВР применялся оптимизатор AdamW при скорости обучения $1 \cdot 10^{-3}$, коэффициенте регуляризации 0.001 и аналогичном планировщике, что и для ВА/ВП-модели, в течение 50 эпох для каждой модели. Тонкая настройка модели вокодера HiFi-GAN осуществлялась с инициализации весов модели, полученных из открытых источников [44], с использованием оптимизатора AdamW и скорости обучения $1 \cdot 10^{-6}$ в течение 40 эпох. Обучение упрощенной модели (Ablation) проводилось с аналогичными параметрами, применяемым в РВР-модели.

В табл. 1 представлено количество тренируемых параметров моделей, оптимизируемых в процессе обучения. Всего рассмотренная архитектура преобразования акцента (полная РВР), состоящая из нескольких взаимосвязанных моделей, имеет 164 млн параметров.

Таблица 1. Количество тренируемых параметров

Модель	Количество параметров, млн
ВА/ВП	24.9
ИГ	4.3
РВФ	82.1
РВР	52.7
Полная РВР	164
Вокодер	84.7
Всего	248.7

3.2. Оценка производительности

Оценка производительности модели осуществлялась на сервере под управлением операционной системы Linux с одним графическим вычислителем (GPU) NVIDIA Tesla T4, виртуальным процессором (virtual central processing unit, vCPU) с 8 ядрами и 16 ГБ оперативной памяти (random access memory, RAM). Для этого модель сначала была экспортирована в открытый формат ONNX и затем развернута с использованием программного обеспечения с открытым исходным кодом NVIDIA Triton.

Используя программный интерфейс развернутой в NVIDIA Triton-модели и тестовый аудиофайл продолжительностью 5 с, содержащий англоязычную L2-речь, были выполнены замеры задержки генерации ответа при 200 итерациях. В итоге средняя задержка при генерации (latency) составила 52 мс, пропускная способность – 96 RTFX.

Результаты оценки производительности модели конвертации акцента показывают низкие задержки при генерации. Вместе с особенностями архитектуры, которая не требует накопления длинного контекста, а может работать с отрезками длительностью менее 0.25 с, это дает возможность применения предлагаемой модели в реальном времени в режиме диалога, когда задержки с ответом влияют на коммуникацию [26–28].

3.3. Объективная оценка качества

Для проведения объективной оценки качества были использованы данные из открытых источников, а также предварительно обученные модели распознавания речи. С помощью предложенного метода конвертации акцента для каждого примера из тестового набора был сгенерирован аудиофайл. Далее рассчитывались метрики качества для оригинального и исправленного аудиофайлов. В табл. 2. представлены результаты объективной оценки качества.

В качестве тестовых наборов данных использованы подвыборки общей продолжительностью 26.9 ч, которые не участвовали в процессе тренировки модели конвертации акцента и ее составляющих. Все они включают в себя текстовые транскрипции и аудиофайлы с речью на английском языке с разными родными и неродными акцентами из открытых источников:

- 3.2 ч из CMU-ARCTIC [45], L2 ARCTIC [46] (ARCTIC), 10 акцентов: американский, английский, китайский, индийский, корейский, вьетнамский, испанский, арабский, голландский, немецкий;
- 3.1 ч из Common Voice [48], 12 акцентов: американский, английский, индийский, австралийский, африканский, китайский, филиппинский, малазийский, немецкий, русский, французский, восточноевропейский;
- 15.2 ч из NPTEL2020, индийский акцент;
- 5.4 ч из Afrispeech-200 [60], африканский акцент (йоруба, суахили, игбо, зулу, тсвана, идома, африкаанс).

Были использованы модели распознавания речи, полученные из открытых источников: Conformer [35], Citrinet [61], Whisper [62]. При этом модель Whisper взята в двух вариантах: «большая» мультиязычная (L. Mult.) и «средняя» англоязычная (M. En.). Распознавание проводилось на аудиофайлах без

обработки и на аудиофайлах после конвертации акцента. Затем распознанные и истинные транскрипции приводились к единому виду с помощью нормализации [54], после чего сравнивались и считались метрики качества: частота ошибок в словах (word error rate, WER), частота ошибок в символах (character error rate, CER). В таблице жирным шрифтом выделены наилучшие результаты для каждой пары: тестовый набор данных и модель распознавания речи.

Таблица 2. Результаты оценки модели конвертации акцента с помощью моделей распознавания речи. Данные после конвертации отмечены, как «конв.»

Тестовый набор данных	Модель распознавания речи			
	Conformer	Citrinet	Whisper L. Mult.	Whisper M. En.
WER, %				
ARCTIC	9.57	11.73	16.23	8.91
ARCTIC конв.	8.78	11.55	12.69	8.68
Common Voice	9.07	25.80	36.89	11.26
Common Voice конв.	9.12	23.38	22.71	10.62
NPTEL2020	29.18	29.88	16.41	15.18
NPTEL2020 конв.	25.26	29.41	13.87	11.64
Afrispeech-200	43.2	46.24	37.91	33.61
Afrispeech-200 конв.	35.19	39.49	35.56	29.96
CER, %				
ARCTIC	3.73	4.85	10.30	3.98
ARCTIC конв.	3.52	4.68	6.06	3.92
Common Voice	3.75	8.74	21.41	5.66
Common Voice конв.	3.77	8.29	13.63	5.22
NPTEL2020	16.87	17.70	11.94	10.67
NPTEL2020 конв.	14.79	17.01	10.10	9.44
Afrispeech-200	31.52	34.79	24.30	20.04
Afrispeech-200 конв.	27.86	28.92	23.15	18.88

Как видно из результатов, практически во всех случаях применение метода конвертации акцента улучшает распознавание предварительно обученных моделей, на что указывают пониженные значения частот ошибок в словах и символах. Модель конвертации акцента улучшает качество речи, делая ее более распознаваемой.

3.4. Субъективная оценка качества

Были проведены тесты на звуковое восприятие, основанные на мнении группы людей, в которых приняли участие 53 человека из разных стран

с уровнем владения английским языком не ниже B2 согласно шкале CEFR⁴. Для этого каждому из участников были даны инструкции, где в рамках каждого эксперимента предлагалось прослушать 1 или 2 аудиофайла и дать свою оценку соответствия критерию качества по пятибалльной шкале, где «1» – точно не соответствует, «2» – скорее не соответствует, «3» – компромисс, «4» – скорее соответствует, «5» – точно соответствует. Далее полученные оценки были использованы для расчета средней экспертной оценки (mean opinion score, MOS) для каждого эксперимента. Результаты представлены в табл. 3.

В качестве звуковых образцов были случайно отобраны 20 пар аудиофайлов из тестовых подвыборок наборов данных L2 ARCTIC [46] и NPTTEL2020 с неродным английским акцентом (Original): индийский, китайский, корейский, вьетнамский, испанский, арабский, немецкий. Каждая оригинальная аудиопара представляет собой запись одного и того же говорящего. Для каждого отобранного аудиофайла (всего 40) были сгенерированы варианты с использованием упрощенной модели конвертации акцента (Ablation) и с помощью предлагаемой модели (Proposed). Всего проведено 3 эксперимента для оценки натуральности голоса, сходства говорящих и отсутствия иностранного акцента. Во всех экспериментах предлагалось сделать не менее 3 оценок для каждого типа звукового образца. При этом сами тестовые образцы менялись в экспериментах, исключая повторение. Перед выставлением оценки можно было прослушать образец неограниченное число раз. Таким образом, каждый опрошенный сделал в общей сложности от 9 до 12 оценок.

При оценке натуральности голоса участникам предлагалось определить по пятибалльной шкале, насколько «натурально» звучит речь в аудиопримере, т.е. создается ли впечатление у слушателя, что это настоящий живой человеческий голос, а не сгенерированная или роботизированная речь. Оценка «1» означает, что голос точно искусственный, синтезированный с использованием методов компьютерной генерации, а «5» – что в примере звучит речь, полученная с использованием способов аналоговой или цифровой звукозаписи голоса реального человека. Также опрошенным были даны рекомендации не обращать внимание на наличие или отсутствие фонового шума в записи, чтобы сконцентрироваться именно на оценке речи.

⁴ Шкала CEFR (Common European Framework of Reference) – система уровней владения иностранным языком, используемая в Европе. <https://www.coe.int/en/web/common-european-framework-reference-languages>. Дата обращения 01.05.2024. [CEFR (Common European Framework of Reference) is the system of foreign language proficiency levels used in Europe. <https://www.coe.int/en/web/common-european-framework-reference-languages>. Accessed May 01, 2024.]

Таблица 3. Результаты субъективной оценки качества (MOS с 95% доверительным интервалом)

Примеры	Натуральность голоса	Сходство говорящих	Отсутствие иностранного акцента
Original	4.83 ± 0.10	4.91 ± 0.08	2.06 ± 0.18
Ablation	3.38 ± 0.13	3.92 ± 0.15	3.58 ± 0.17
Proposed	4.04 ± 0.16	4.30 ± 0.18	4.11 ± 0.14

С целью проведения эксперимента по оценке сходства говорящих были подготовлены пары аудиозаписей: Original – Original, Original – Ablation и Original – Proposed. При этом в первую пару входят записи только из оригинальных данных, представляющих собой записи одного и того же говорящего, но произносящего разные фразы. В другие пары входит оригинальная запись одной фразы и сгенерированный вариант другой фразы того же самого говорящего. Участникам предлагалось прослушать такие пары аудиозаписей и решить, произнесены ли они одним и тем же человеком, т.е. насколько тембр в одном файле похож на тембр в другом файле. Оценка «1» – речь в аудиозаписях точно принадлежит разным людям, «5» – тембр говорящих в аудиозаписях идентичный, принадлежащий одному человеку. Опрошенным рекомендовано во время оценки не обращать внимание на свойства L1 и L2-акцента, чтобы сосредоточиться на сравнении обертоновой окраски голоса.

Для оценки отсутствия иностранного акцента участникам предлагалось прослушать англоязычный аудиофайл и решить, на сколько, по их мнению, в данной записи выражен иностранный акцент. Английский и американский акценты приняли считать родными L1, а все остальные – неродными L2. Оценка «1» означает, что в речи ярко выражен иностранный L2-акцент, «5» – речь точно является англоязычной L1 без иностранного акцента.

Анализ таблицы показывает, что наивысшие оценки натуральности голоса и сходства говорящих показывают оригинальные примеры, что очевидно, т.к. они получены без использования методов синтеза речи, а вместе с наименьшей оценкой отсутствия иностранного акцента демонстрирует калибровку мнений участников эксперимента на реальных данных. Добавление ВА/ВП-модели в общую схему модели конвертации акцента существенно повышает качество генерации, это демонстрируют повышенные результаты по сравнению с упрощенной Ablation-моделью. Во всех субъективных экспериментах предложенная модель показывает оценку выше «4», означающую, что, по мнению опрошенных, модель скорее соответствует обозначенным критериям качества.

ЗАКЛЮЧЕНИЕ

В работе представлен метод конвертации акцента, который позволяет преобразовывать любую L2-речь с выраженным иностранным акцентом в L1-речь, не зависит от наличия эталонных примеров и параллельных данных на этапах обучения и генерации, что значительно упрощает, удешевляет и ускоряет процесс адаптации системы к новым акцентам.

Предложенная модель является неавторегрессионной и не использует в своей архитектуре рекуррентные сети, что позволяет ускорить процесс тренировки, выполнять перевод акцента в режиме реального времени, избежать эффекта накопления ошибок, связанного с последовательной генерацией вывода.

Рассматриваемый метод также включает алгоритм клонирования речевых характеристик голоса, благодаря которому сохраняется идентичность говорящего даже после преобразования акцента. Это особенно важно в ситуациях, когда необходимо сохранить эмоциональную окраску, выразительность и индивидуальные особенности речи. Кроме того, метод позволяет в реальном времени в процессе генерации модифицировать характеристики голоса, такие как акцент, тембр и особенности голоса, относящиеся к полу говорящего, путем копирования соответствующих характеристик из аудиообразца,

что делает его применимым в более широком круге сценариев, чем прежние разработки.

Модель демонстрирует высокое качество конвертации акцента с сохранением оригинального тембра, а также низкие задержки при генерации, приемлемые для использования в сценариях реального времени.

Применение метода позволяет:

- 1) преобразовывать англоязычную речь с иностранным L2-акцентом в L1-речь без иностранного акцента;
- 2) улучшать качество речи и, как следствие, повышать качество распознавания существующих систем;
- 3) копировать и менять голосовые характеристики говорящего в реальном времени;
- 4) применять конвертацию акцента в реальном времени в режиме диалога.

Разработанная прикладная нейросетевая модель продемонстрировала возможность работы в информационных системах на английском языке в режиме реального времени. Результаты исследования могут применяться при разработке систем модификации голоса, а также распознавания и генерации речи.

Вклад авторов

Все авторы в равной степени внесли свой вклад в исследовательскую работу.

Authors' contribution

All authors equally contributed to the research work.

СПИСОК ЛИТЕРАТУРЫ / REFERENCES

1. McMillin D.C. Outsourcing identities: Call centres and cultural transformation in India. *Economic and Political Weekly*. 2006;41(3):235–241.
2. Felps D., Bortfeld H., Gutierrez-Osuna R. Foreign accent conversion in computer assisted pronunciation training. *Speech Communication*. 2009;51(10):920–932. <https://doi.org/10.1016/j.specom.2008.11.004>
3. Probst K., Ke Y., Eskenazi M. Enhancing foreign language tutors–In search of the golden speaker. *Speech Communication*. 2002;37(3–4):161–173. [https://doi.org/10.1016/S0167-6393\(01\)00009-7](https://doi.org/10.1016/S0167-6393(01)00009-7)
4. Türk O., Arslan L. M. Subband based voice conversion. In: *7th International Conference on Spoken Language Processing, ICSLP2002 – INTERSPEECH 2002*. *Interspeech*. 2002. P. 289–292.
5. Biadsy F., Weis, R.J., Moreno P.J., Kanevsky D., Jia Y. Parrottron: An end-to-end speech-to-speech conversion model and its applications to hearing-impaired speech and speech separation. *Interspeech*. 2019. P. 4115–4119. <http://doi.org/10.21437/Interspeech.2019-1789>
6. Birner B. *Why Do Some People Have an Accent?* Linguistic Society of America. Washington, DC. 1999. 6 p.
7. Baese-Berk M.M., Morrill T.H. Speaking rate consistency in native and non-native speakers of English. *J. Acoust. Soc. Am.* 2015;138(3):EL223–EL228. <https://doi.org/10.1121/1.4929622>
8. Piske T., MacKay I.R.A., Flege J.E. Factors affecting degree of foreign accent in an L2: A review. *J. Phonetics*. 2001;29(2):191–215. <https://doi.org/10.1006/jpho.2001.0134>
9. Munro M.J., Derwing T.M. Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Language Learning*. 1995;45(1):73–97. <https://doi.org/10.1111/j.1467-1770.1995.tb00963.x>
10. Lev-Ari S., Keysar B. Why don't we believe non-native speakers? The influence of accent on credibility. *J. Exp. Soc. Psychol.* 2010;46(6):1093–1096. <https://doi.org/10.1016/j.jesp.2010.05.025>
11. Rubin D.L., Smith K.A. Effects of accent, ethnicity, and lecture topic on undergraduates' perceptions of nonnative English-speaking teaching assistants. *Int. J. Intercult. Relat.* 1990;14(3):337–353. [https://doi.org/10.1016/0147-1767\(90\)90019-S](https://doi.org/10.1016/0147-1767(90)90019-S)

12. Nelson Jr. L.R., Signorella M.L., Botti K.G. Accent, gender, and perceived competence. *Hispanic J. Behavior. Sci.* 2016;38(2):166–185. <https://doi.org/10.1177/0739986316632319>
13. Zhao G., Gutierrez-Osuna R. Using phonetic posteriorgram based frame pairing for segmental accent conversion. In: *IEEE/ACM Transactions on Audio, Speech, and Language Processing*. 2019;27(10):1649–1660. <https://doi.org/10.1109/TASLP.2019.2926754>
14. Zhao G., Sonsaat S., Levis J., Chukharev-Hudilainen E., Gutierrez-Osuna R. Accent conversion using phonetic posteriorgrams. In: *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE; 2018. P. 5314–5318. <https://doi.org/10.1109/ICASSP.2018.8462258>
15. Aryal S., Gutierrez-Osuna R. Can voice conversion be used to reduce non-native accents? In: *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE; 2014. P. 7879–7883. <https://doi.org/10.1109/ICASSP.2014.6855134>
16. Ding S., Zhao G., Gutierrez-Osuna R. Accentron: Foreign accent conversion to arbitrary non-native speakers using zero-shot learning. *Computer Speech & Language*. 2022;72:101302. <https://doi.org/10.1016/j.csl.2021.101302>
17. Quamer W., Das A., Levis J., Chukharev-Hudilainen E., Gutierrez-Osuna R. Zero-shot foreign accent conversion without a native reference. *Proc. Interspeech*. 2022. <http://doi.org/10.21437/Interspeech.2022-10664>
18. Jin M., Serai P., Wu J., Tjandra A., Manohar V., He Q. Voice-preserving zero-shot multiple accent conversion. In: *ICASSP 2023–2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE; 2023. P. 1–5. <https://doi.org/10.1109/ICASSP49357.2023.10094737>
19. Zhou Y., Wu Z., Zhang M., Tian X., Li H. TTS-guided training for accent conversion without parallel data. *IEEE Signal Proc. Lett.* 2023;30:533–537. <https://doi.org/10.1109/lsp.2023.3270079>
20. Zhao G., Ding S., Gutierrez-Osuna R. Converting foreign accent speech without a reference. In: *IEEE/ACM Transactions on Audio, Speech, and Language Processing*. 2021;29:2367–2381. <https://doi.org/10.1109/TASLP.2021.3060813>
21. Liu S., Wang D., Cao Y., Sun L., Wu X., Kang S., Wu Z., Liu X., Su D., Yu D., Meng H. End-to-end accent conversion without using native utterances. In: *ICASSP 2020–2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE; 2020. P. 6289–6293.
22. Zhou X., Zhang M., Zhou Y., Wu Z., Li H. Accented text-to-speech synthesis with limited data. In: *IEEE/ACM Transactions on Audio, Speech, and Language Processing*. 2024;32:1699–1711. <https://doi.org/10.1109/TASLP.2024.3363414>
23. Pinget A.F., Bosker H.R., Quené H., De Jong, N.H. Native speakers' perceptions of fluency and accent in L2 speech. *Language Testing*. 2014;31(3):349–365. <https://doi.org/10.1177/0265532214526177>
24. Бархударова Е.Л. Методологические проблемы анализа иностранного акцента в русской речи. *Вестник Московского университета. Серия 9. Филология*. 2012;6:57–70.
[Barkhudarova E.L. Methodological Problems in Analyzing Foreign Accents in Russian Speech. *Vestnik Moskovskogo universiteta. Seriya 9. Filologiya* = *Lomonosov Philology J.* 2012;6:57–70 (in Russ.).]
25. Arik S., Chen J., Peng K., Ping W., Zhou Y. Neural voice cloning with a few samples. *Advances in Neural Information Processing Systems (NeurIPS 2018)*. 2018;31.
26. Cohen D. Issues in transnet packetized voice communication. In: *Proceedings of the fifth Symposium on Data Communications (SIGCOMM'77)*. 1977. P. 6.10–6.13. <https://doi.org/10.1145/800103.803349>
27. Liang Y.J., Farber N., Girod B. Adaptive playout scheduling and loss concealment for voice communication over IP networks. *IEEE Trans. Multimedia*. 2003;5(4):532–543. <https://doi.org/10.1109/TMM.2003.819095>
28. Matzinger T., Pleyer M., Żywicznyński P. Pause Length and Differences in Cognitive State Attribution in Native and Non-Native Speakers. *Languages*. 2023;8(1):26. <http://doi.org/10.3390/languages8010026>
29. Medsker L.R., Jain L. (Eds.). *Recurrent Neural Networks. Design and Applications*. Boca Raton: CRC Press; 2001. 416 p.
30. Li J., Lavrukhin V., Ginsburg B., Leary R., Kuchaiev O., Cohen J.M., Nguyen H., Gadde R.T. Jasper: An End-to-End Convolutional Neural Acoustic Model. *Interspeech 2019*. 2019. <https://doi.org/10.21437/interspeech.2019-1819>
31. Dawalatabad N., Ravanelli M., Grondin F., Thienpondt J., Desplanques B., Na H. *ECAPA-TDNN Embeddings for Speaker Diarization*. arXiv preprint arXiv:2104.01466. 2021. <https://doi.org/10.48550/arXiv.2104.01466>
32. Ravanelli M., Bengio Y. Speaker recognition from raw waveform with SincNet. In: *2018 IEEE Spoken Language Technology Workshop (SLT)*. IEEE; 2018. P. 1021–1028. <https://doi.org/10.1109/SLT.2018.8639585>
33. Snyder D., Garcia-Romero D., Sell G., Povey D., Khudanpur S. X-vectors: Robust DNN embeddings for speaker recognition. In: *2018 IEEE international Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE; 2018. P. 5329–5333. <http://doi.org/10.1109/ICASSP.2018.8461375>
34. Deng J., Guo J., Xue N., Zafeiriou S. Arcface: Additive angular margin loss for deep face recognition. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. IEEE; 2019. P. 4690–4699. <https://doi.org/10.1109/CVPR.2019.00482>
35. Gulati A., Qin J., Chiu C., Parmar N., Zhang Y., Yu J., Han W., Wang S., Zhang Z., Wu Y., Pang R. Conformer: Convolution-augmented trans-former for speech recognition. *Proc. Interspeech 2020*. 2020. P. 5036–5040. <https://doi.org/10.21437/interspeech.2020-3015>
36. Glorot X., Bengio Y. Understanding the difficulty of training deep feedforward neural networks. In: *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics. Proceedings of Machine Learning Research*. 2010. P. 249–256. URL: <http://proceedings.mlr.press/v9/glorot10a.html>

37. Gu J., Wang Z., Kuen J., Ma L., Shahroudy A., Shuai B., Liu T., Wang X., Wang G., Cai J., Chen T. Recent advances in convolutional neural networks. *Pattern Recognition*. 2018;77:354–377. <https://doi.org/10.1016/j.patcog.2017.10.013>
38. Vaswani A., Shazeer N., Parmar N., Uszkoreit J., Jones L., Gomez A.N., Kaiser Ł., Polosukhin I. Attention is all you need. *Advances in Neural Information Processing Systems*. 2017;30:5999–6009. <https://doi.org/10.48550/arXiv.1706.03762>
39. Ren Y., Ruan Y., Tan X., Qin T., Zhao S., Zhao Z., Liu T.Y. FastSpeech: Fast, robust and controllable text to speech. *Advances in Neural Information Processing Systems*. 2019;32. <https://doi.org/10.48550/arXiv.1905.09263>
40. Graves A., Fernández S., Gomez F., Schmidhuber J. Connectionist temporal classification: Labelling unsegmented sequence data with recurrent neural networks. In: *Proceedings of the 23rd International Conference on Machine Learning*. 2006. P. 369–376. <https://doi.org/10.1145/1143844.1143891>
41. Ghahremani P., BabaAli B., Povey D., Riedhammer K., Trmal J., Khudanpur S. A pitch extraction algorithm tuned for automatic speech recognition. In: *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE; 2014. P. 2494–2498. <http://doi.org/10.1109/ICASSP.2014.6854049>
42. Gerhard D. *Pitch Extraction and Fundamental Frequency: History and Current Techniques*. Masters Thesis. Regina, SK, Canada: Department of Computer Science, University of Regina; 2003. 23 p.
43. Łańcucki A. Fastpitch: Parallel text-to-speech with pitch prediction. In: *ICASSP 2021–2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE; 2021. P. 6588–6592. <https://doi.org/10.1109/ICASSP39728.2021.9413889>
44. Kong J., Kim J., Bae J. HiFi-GAN: Generative adversarial networks for efficient and high fidelity speech synthesis. *Advances in Neural Information Processing Systems*. 2020;33:17022–17033. <http://doi.org/10.48550/arXiv.2010.05646>
45. Kominek J., Black A.W. The CMU Arctic speech databases. In: *Fifth ISCA Workshop on Speech Synthesis*. 2004. P. 223–224.
46. Zhao G., Somaat S., Silpachai A., Lucic I., Chukharev-Hudilainen E., Levis J., Gutierrez-Osuna R. L2-ARCTIC: A Non-native English Speech Corpus. *Interspeech 2018*. 2018. P. 2783–2787. <http://doi.org/10.21437/Interspeech.2018-1110>
47. Weinberger S.H., Kunath S.A. The Speech Accent Archive: towards a typology of English accents. In: *Corpus-based Studies in Language Use, Language Learning, and Language Documentation*. Brill; 2011. P. 265–281. https://doi.org/10.1163/9789401206884_014
48. Ardila R., Branson M., Davis K., Kohler M., Meyer J., Henretty M., Henretty M., Morais R., Saunders L., Tyers F., Weber G. Common Voice: A Massively-Multilingual Speech Corpus. In: *Proceedings of the Twelfth Language Resources and Evaluation Conference*. 2020. P. 4218–4222. <https://doi.org/10.48550/arXiv.1912.06670>
49. Nagrani A., Chung J.S., Zisserman A. Voxceleb: a large-scale speaker identification dataset. *Interspeech 2017*. 2017. <http://doi.org/10.21437/Interspeech.2017-950>
50. Chung J., Nagrani A., Zisserman A. VoxCeleb2: Deep speaker recognition. *Interspeech 2018*. 2018. <http://doi.org/10.21437/Interspeech.2018-1929>
51. Panayotov V., Chen G., Povey D., Khudanpur S. Librispeech: an ASR corpus based on public domain audio books. In: *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE; 2015. P. 5206–5210. <http://doi.org/10.1109/ICASSP.2015.7178964>
52. Veaux C., Yamagishi J., MacDonald K. CSTR VCTK Corpus: English Multi-speaker Corpus for CSTR Voice Cloning Toolkit. *University of Edinburgh. The Center for Speech Technology Research (CSTR)*. 2017. <https://doi.org/10.7488/ds/2645>
53. Chen G., Chai S., Wang G., Du J., Zhang W., Weng C., Su D., Povey D., Trmal J., Zhang J., Jin M., Khudanpur S., Watanabe S., Zhao S., Zou W., Li X., Yao X., Wang Y., Wang Y., You Z., Yan Z. GigaSpeech: An evolving, multi-domain ASR corpus with 10,000 hours of transcribed audio. In: *22nd Annual Conference of the International Speech Communication Association, Interspeech 2021*. International Speech Communication Association; 2021. P. 4376–4380. <https://doi.org/10.21437/Interspeech.2021-1965>
54. Bakhturina E., Zhang Y., Ginsburg B. Shallow Fusion of Weighted Finite-State Transducer and Language Model for Text Normalization. *Proc. Interspeech 2022*. 2022. <http://doi.org/10.48550/arXiv.2203.15917>
55. Kudo T., Richardson J. SentencePiece: A simple and language independent subword tokenizer and detokenizer for Neural Text Processing. In: *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*. 2018. P. 66–71. <https://doi.org/10.48550/arXiv.1808.06226>
56. Koizumi Y., Zen H., Karita S., Ding Y., Yatabe K., Morioka N., Bacchiani M., Zhang Y., Han W., Bapna A. Libritts-r: A Restored Multi-Speaker Text-to-Speech Corpus. *arXiv preprint arXiv:2305.18802*. 2023. <https://doi.org/10.48550/arXiv.2305.18802>
57. Paszke A., Gross S., Massa F., Lerer A., Bradbury J., Chanan G., Killeen T., Lin Z., Gimelshein N., Antiga L., Desmaison A., Kopf A., Yang E., DeVito Z., Raison M., Tejani A., Chilamkurthy S., Steiner B., Fang L., Bai J., Chintala S. PyTorch: An imperative style, high-performance deep learning library. *Advances in Neural Information Processing Systems*. 2019;32:8024–8035.
58. Kuchaiev O., Li J., Nguyen H., Hrinchuk O., Leary R., Ginsburg B., Krizan S., Beliaev S., Lavrukhin V., Cook J., Castonguay P., Popova M., Huang J., Cohen J. Nemo: a toolkit for building ai applications using neural modules. *arXiv preprint arXiv:1909.09577*. 2019. <https://doi.org/10.48550/arXiv.1909.09577>
59. Bredin H., Yin R., Coria J.M., Gelly G., Korshunov P., Lavechin M., Fustes D., Titeux H., Bouaziz W., Gill M.P. Pyannote. Audio: neural building blocks for speaker diarization. In: *ICASSP 2020–2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE; 2020. P. 7124–7128. <https://doi.org/10.1109/ICASSP40776.2020.9052974>

60. Olatunji T., Afonja T., Yadavalli A., Emezue C.C., Singh S., Dossou B.F., Osuchukwu J., Osei S., Tonja A.L., Etori N., Mbataku C. AfriSpeech-200: Pan-African Accented Speech Dataset for Clinical and General Domain ASR. *Transactions of the Association for Computational Linguistics*. 2023;11:1669–1685. https://doi.org/10.1162/tac1_a_00627
61. Majumdar S., Balam J., Hrinchuk O., Lavrukhin V., Noroozi V., Ginsburg B. Citrinet: Closing the gap between non-autoregressive and autoregressive end-to-end models for automatic speech recognition. *arXiv preprint arXiv:2104.01721*. 2021. <http://doi.org/10.48550/arXiv.2104.01721>
62. Radford A., Kim J.W., Xu T., Brockman G., McLeavey C., Sutskever I. Robust speech recognition via large-scale weak supervision. In: *Proceedings of the 40th International Conference on Machine Learning*. PMLR 202. 2023. P. 28492–28518. <http://doi.org/10.48550/arXiv.2212.04356>

Об авторах

Нечаев Владимир Алексеевич, преподаватель-исследователь, ФГБОУ ВО «Ивановский государственный энергетический университет имени В.И. Ленина» (153003, Россия, Иваново, ул. Рабфаковская, д. 34). E-mail: nechaev@gapps.ispu.ru. SPIN-код РИНЦ 7002-3878, <https://orcid.org/0009-0007-1449-3968>

Косяков Сергей Витальевич, д.т.н., профессор, заведующий кафедрой программного обеспечения компьютерных систем, ФГБОУ ВО «Ивановский государственный энергетический университет имени В.И. Ленина» (153003, Россия, Иваново, ул. Рабфаковская, д. 34). E-mail: ksv@ispu.ru. Scopus Author ID 6507182528, ResearcherID H-5686-2018, SPIN-код РИНЦ 1371-9929, <https://orcid.org/0000-0003-0231-0750>

About the Authors

Vladimir A. Nechaev, Teacher-Researcher, Ivanovo State Power Engineering University (34, Rabfakovskaya ul., Ivanovo, 153003 Russia). E-mail: nechaev@gapps.ispu.ru. RSCI SPIN-code 7002-3878, <https://orcid.org/0009-0007-1449-3968>

Sergey V. Kosyakov, Dr. Sci. (Eng.), Professor, Head of the Department of Computer Systems Software, Ivanovo State Power Engineering University (34, Rabfakovskaya ul., Ivanovo, 153003 Russia). E-mail: ksv@ispu.ru. Scopus Author ID 6507182528, ResearcherID H-5686-2018, RSCI SPIN-code 1371-9929, <https://orcid.org/0000-0003-0231-0750>

УДК 004.8

<https://doi.org/10.32362/2500-316X-2025-13-3-21-43>

EDN QKUGFZ



ОБЗОР

Методы интеграции знаний для разработки вопросно-ответных систем

Д.В. Радюш[@]

Национальный исследовательский университет ИТМО, Санкт-Петербург, 197101 Россия

[@] Автор для переписки, e-mail: daniil.radyush@gmail.com

• Поступила: 26.06.2024 • Доработана: 13.02.2025 • Принята к опубликованию: 27.03.2025

Резюме

Цели. Несмотря на наблюдаемые в последние несколько лет успехи больших языковых моделей, которые способны решать широкий перечень задач, ряд практических проблем остается не до конца решенным. В контексте построения вопросно-ответных систем к таким проблемам можно отнести использование общих знаний и учет причинно-следственных связей. Целью статьи является рассмотрение методов интеграции знаний, которые способны усовершенствовать функционирование больших языковых моделей путем предоставления необходимых сведений и закономерностей из внешних источников.

Методы. В работе осуществляются классификация, анализ и сопоставление методов интеграции знаний, используемых в актуальных реализациях вопросно-ответных систем. В частности, рассматривается вовлечение вспомогательных сведений через самообучение, дообучение, механизм внимания и использование токенов взаимодействия, а также описываются соответствующие вспомогательные подходы для акцентирования наиболее релевантных сведений.

Результаты. Рассмотренные в обзоре вопросно-ответные системы непосредственно демонстрируют возрастание точности относительно базового решения на основе предобученной языковой модели за счет использования методов интеграции знаний на примере бенчмарка CommonsenseQA. При этом в целом более высокие результаты показывают методы интеграции знаний, основанные на использовании языковых моделей и механизма внимания.

Выводы. Представленный систематический обзор существующих методов интеграции знаний из внешних источников в работу вопросно-ответных систем фактически подтверждает эффективность и перспективность этого направления исследований. Данные методы демонстрируют не только возможность увеличить точность вопросно-ответных систем, но и в некоторой степени сгладить проблемы, связанные с интерпретируемостью результатов и устареванием знаний в предобученных моделях. Последующие изыскания способны как улучшить и оптимизировать отдельные аспекты существующих подходов, так и выработать концептуально новые.

Ключевые слова: глубокое обучение, обработка естественного языка, вопросно-ответная система, база знаний, графовые нейронные сети, интеграция знаний

Для цитирования: Радюш Д.В. Методы интеграции знаний для разработки вопросно-ответных систем. *Russian Technological Journal*. 2025;13(3):21–43. <https://doi.org/10.32362/2500-316X-2025-13-3-21-43>, <https://www.elibrary.ru/QKUGFZ>

Прозрачность финансовой деятельности: Автор не имеет финансовой заинтересованности в представленных материалах или методах.

Автор заявляет об отсутствии конфликта интересов.

REVIEW ARTICLE

Knowledge injection methods in question answering

Daniil V. Radyush [@]

ITMO University, Saint Petersburg, 197101 Russia

[@] Corresponding author, e-mail: daniil.radyush@gmail.com

• Submitted: 26.06.2024 • Revised: 13.02.2025 • Accepted: 27.03.2025

Abstract

Objectives. Despite the recent success of large language models, which are now capable of solving a wide range of tasks, a number of practical issues remain unsolved. For example, users of systems providing question answering (QA) services may experience a lack of commonsense knowledge and reasoning proficiency. The present work considers knowledge injection methods as a means of providing functional enhancements to large language models by providing necessary facts and patterns from external sources.

Methods. Knowledge injection methods leveraged in relevant QA systems are classified, analyzed, and compared. Self-supervised learning, fine-tuning, attention mechanism and interaction tokens for supporting information injection are considered along with auxiliary approaches for emphasizing the most relevant facts.

Results. The reviewed QA systems explicitly show the accuracy increase on the CommonsenseQA benchmark compared to pretrained language model baseline due to knowledge injection methods exploitation. At the same time, in general the higher results are related to knowledge injection methods based on language models and attention mechanism.

Conclusions. The presented systematic review of existing external knowledge injection methods for QA systems confirms the continuing validity of this research direction. Such methods are not only capable of increasing the accuracy of QA systems but also mitigating issues with interpretability and factual obsolescence in pretrained models. Further investigations will be carried out to improve and optimize different aspects of the current approaches and develop conceptually novel ideas.

Keywords: deep learning, nature language processing, question answering system, knowledge base, graph neural networks, knowledge injection

For citation: Radyush D.V. Knowledge injection methods in question answering. *Russian Technological Journal*. 2025;13(3):21–43. <https://doi.org/10.32362/2500-316X-2025-13-3-21-43>, <https://www.elibrary.ru/QKUGFZ>

Financial disclosure: The author has no financial or proprietary interest in any material or method mentioned.

The author declares no conflicts of interest.

ВВЕДЕНИЕ

Развитие вопросно-ответных систем в последние годы в существенной степени было обусловлено появлением и последующим совершенствованием предобученных (Pretrained) языковых моделей [1]. Эффективность таких моделей основывается на обработке большого корпуса текстов, содержащего разнородную информацию, что позволяет фиксировать в весах модели как определенные языковые закономерности, так и конкретные факты [2]. Тем не менее, в силу особенностей естественных языков значительное количество сведений об окружающем мире не всегда представлено в тексте в явном виде, что затрудняет их выявление на этапе обучения языковыми моделями. В первую очередь такие сведения касаются разного рода социальных взаимодействий, психологических аспектов и базовых физических законов, которые еще в раннем возрасте усваиваются человеком. В качестве примеров здесь можно привести понимание необходимости смотреть по сторонам при переходе через дорогу или остужать слишком горячую еду перед ее употреблением.

Для того чтобы нивелировать этот недостаток, могут использоваться разного рода источники сведений, в которых такие данные будут зафиксированы в однозначной форме. Одним из первых примеров базы знаний, разработка которой ставила своей целью сбор в т.ч. и общих представлений об окружающем мире, можно считать Cys¹. В ней подобные сведения фиксировались в виде логических правил, что соответствовало основному направлению развития приложений из области искусственного интеллекта того времени и должно было обеспечить их дальнейшее совершенствование. К настоящему времени получен целый ряд подобных источников, хотя и с несколько иными подходами к описанию знаний, таких как ATOMIC [3] и ConceptNet².

С формальной точки зрения можно выделить несколько доводов в пользу привлечения внешних источников знаний при разработке вопросно-ответных систем. Во-первых, вероятно, главным мотивом является непосредственно получение более точных и удовлетворяющих пользователя результатов. Предполагается, что за счет использования дополнительного контекста к запросу модель сможет ответить на ряд вопросов, для которых внутренних представлений предобученных языковых моделей может быть недостаточно. К подобным вопросам,

с одной стороны, можно отнести такие, где опущены отдельные причинно-следственные связи, а, с другой стороны, существует неопределенность в плане идентификации семантики некоторых слов из-за их многозначности и недостаточности контекста. Во многом это определяется ограничениями, выявленными в рамках анализа применения моделей-трансформеров: они в основном используют только поверхностный, статистически наиболее популярный смысл отдельных слов [4], а при логическом выводе в значительной степени опираются на эвристики, усвоенные из обучающей выборки [5].

При этом даже языковые модели с большим количеством весов, демонстрирующие высокие результаты на множестве бенчмарков, могут не просто ошибаться, а еще и выдавать ответы, не имеющие отношения к действительности, – галлюцинации (Hallucinations) [6]. В связи с этим даже сформировалось отдельное направление исследований, посвященное методам извлечения релевантной для запроса информации и включению ее во входные данные для улучшения качества ответов [7] и уменьшения числа галлюцинаций – Retrieval Augmented Generation (расширенная поисковая генерация) [8].

Во-вторых, использование внешних источников знаний способно снизить требования к необходимым вычислительным ресурсам для использования предобученных языковых моделей. В частности, целенаправленное вовлечение дополнительных сведений может позволить применять модели с меньшим числом весов при сохранении точности системы на сопоставимом уровне [9]. Это может упростить работу с вопросно-ответными системами на практике, а также компенсировать издержки на извлечение и обработку вспомогательных данных.

В-третьих, не менее важным является использование баз знаний с позиции интерпретируемого искусственного интеллекта (Explainable Artificial Intelligence). Так, благодаря структурированности баз знаний, извлеченные из них сведения способны воспроизводить своего рода логические цепочки, которые можно предоставлять пользователю в качестве обоснования результата работы системы. Подобное свойство может быть крайне важным с точки зрения практического применения, т.к. нередко именно отсутствие интерпретируемости у результатов, полученных с помощью нейронных сетей, сдерживает их применение в областях, где риск ошибки и соответствующий потенциальный ущерб для общества достаточно велик. Помимо этого, и в целом для контроля за адекватностью функционирования системы желательно иметь наиболее полное представление о ее работе.

Наконец, существенной является проблема обновления фактов, зафиксированных в весах языковых

¹ Cyscorp. <https://cysc.com>. Дата обращения 01.12.2024. / Accessed December 01, 2024.

² ConceptNet. An open, multilingual knowledge graph. <https://conceptnet.io>. Дата обращения 01.12.2024. / Accessed December 01, 2024.

моделей. Обучение таких моделей осуществляется на конкретных наборах данных и обычно занимает довольно длительный промежуток времени. В то же время каждый день происходит огромное количество событий, что приводит к изменению части знаний и появлению новых фактов. Одним из способов решения данной проблемы как раз и может быть извлечение такой информации из внешних баз данных.

В связи с этим достаточно актуальным становится рассмотрение способов использования вспомогательных общих сведений для решения конкретных задач, таких как разработка вопросно-ответных систем. В частности, для успешной работы системы требуется, чтобы полученная информация была достаточна, но при этом не избыточна, т.к. в противном случае это, наоборот, может затруднять ее функционирование и ухудшать соответствующие результаты. Также не меньшее значение имеет то, каким образом обрабатываются дополнительные знания, поскольку это будет во многом определять эффект от их использования в системе. Таким образом, т.к. на процедуру интеграции знаний может влиять существенное количество аспектов, в данной работе представлена попытка систематически проанализировать и сопоставить существующие подходы, чтобы составить полную картину соответствующих идей.

БАЗЫ ЗНАНИЙ

В целом в области разработки вопросно-ответных систем можно выделить несколько направлений в зависимости от особенностей привлечения дополнительных данных. Первое направление подразумевает отсутствие специализированной базы знаний и ориентировано на использование сведений из источников общего назначения – Open Domain Question Answering (поиск ответов на вопросы общего характера). Чаще всего в качестве такого источника может выступать Wikipedia³, вследствие чего, в силу ее значительного объема и структурной неоднородности содержания, акцент существенно смещается в сторону методов поиска релевантной информации.

Также существует направление Closed Domain Question Answering (поиск ответов на специализированные вопросы), которое имеет дело с более узконаправленными запросами, в т.ч. усложненными необходимостью осуществления логического вывода и учета специфической информации. В связи с этим в качестве внешнего источника знаний могут выступать специализированные базы структурированных знаний, например, графы знаний, что

в определенной степени упрощает поиск информации, а также осуществление логического вывода по данным запроса и вспомогательных операций. В качестве соответствующего примера можно привести граф знаний DBpedia⁴.

В контексте разработки подходов к интеграции знаний в области вопросно-ответных систем интерес представляют в первую очередь базы структурированных знаний. В какой-то степени это можно обосновать текущим положением вещей в данной области. В частности, появление и последующее развитие предобученных языковых моделей позволило существенно снизить требования к предоставляемому к запросу контексту. В результате в настоящее время на ряде датасетов с учетом дообучения (Fine-tuning) некоторые модели способны показывать результаты, сопоставимые с результатами человека. Вследствие этого основной интерес приходится именно на анализ случаев, в которых человек превосходит существующие вопросно-ответные системы. И как раз основную часть таких случаев, как правило, и составляют запросы, требующие внеконтекстных общих представлений об устройстве окружающего мира, а также анализа причинно-следственных связей между отдельными фактами.

Именно поэтому особенно полезными в таких условиях могут считаться базы структурированных общих знаний. Во-первых, они непосредственно предоставляют системе отсутствующие факты. Во-вторых, эти факты могут быть извлечены с учетом существующих связей между собой и вместе с другой сопутствующей информацией. В-третьих, структурированность знаний значительно упрощает их машинную обработку и, следовательно, использование на практике. Таким образом, представляется возможным в определенной степени одновременно решить проблемы, связанные с частью запросов, которые могут считаться сложными для существующих вопросно-ответных систем.

Как уже отмечалось ранее, в качестве внешнего источника дополнительных данных может использоваться Wikipedia, что сохраняет определенную актуальность. Тем не менее, в силу отсутствия систематизации и большой избыточности информации как альтернатива начала также применяться структурированная база знаний, основанная на сведениях из Wikipedia – Wikidata⁵. Граф Wikidata состоит из более чем 100 млн записей, некоторым образом описывающих элементы человеческого знания. Каждому элементу графа соответствует определенный набор свойств, характеризующих его и устанавливающих

⁴ The DBpedia Knowledge Base. <https://www.dbpedia.org>. Дата обращения 01.12.2024. / Accessed December 01, 2024.

⁵ Wikidata. The free knowledge base. <https://www.wikidata.org>. Дата обращения 01.12.2024. / Accessed December 01, 2024.

³ <https://www.wikipedia.org/>. Дата обращения 01.12.2024. / Accessed December 01, 2024.

его взаимосвязи с другими элементами. Таким образом, как и для других графов знаний, содержание Wikidata можно представить в виде набора так называемых триплетов «субъект-предикат-объект», где объект в данном случае является набором конкретных значений свойств, либо ссылкой на другую сущность.

Также к наиболее часто используемым источникам знаний в контексте построения вопросно-ответных систем относится база знаний ConceptNet. Данная база знаний, помимо уникальных сведений общего характера, частично включает в себя информацию и из других относительно часто используемых источников, таких как упомянутые ранее Сус и DBpedia. В рамках ConceptNet слова и словосочетания группируются на основании нескольких десятков отношений. По сравнению с рассмотренной ранее Wikidata, ConceptNet содержит более 30 млн записей, хотя при этом надо учитывать, что значительная часть этой величины обусловлена наличием фактически дублирующих записей из-за существования аналогов на другом языке, однокоренных слов и симметричных взаимосвязей. Кроме того, чуть больший акцент в ConceptNet приходится на лингвистические свойства, например, за счет фиксирования для слова синонимов, антонимов и этимологически связанных слов. Наконец, особенностью ConceptNet является существование весов у каждого отношения между элементами, что эвристически отражает степень вероятности или важности данного отношения.

Из относительно недавно появившихся баз общих знаний можно также выделить ATOMIC, содержащую более 1 млн элементов. Особенностью ATOMIC является отражение сведений в виде абстрактных событий и их результатов, что позволяет делать акцент на комплексные причинно-следственные связи, существующие в окружающем мире. В частности, например, исходя из некоторого события в ATOMIC, возможно выявить какие-либо его последствия, а также намерение, желание или характеристику одного из участников, что способно предоставить модели потенциально отсутствующие знания о социальных взаимодействиях.

В табл. 1 представлены примеры информации, которая может быть извлечена из рассмотренных выше баз знаний, что в целом говорит об их определенной схожести, за исключением обладающей более специфичными целями базы ATOMIC.

МЕТОДЫ ИНТЕГРАЦИИ ЗНАНИЙ

Основываясь на анализе актуальных исследований по теме, автор разработал классификацию методов интеграции знаний, представленную на рис. 1. Согласно ей, возможно рассматривать основные

Таблица 1. Примеры извлекаемой из баз знаний информации

База знания	Пример данных
DBpedia	DBpedia subject SemanticWeb (DBpedia субъект Семантической паутины)
Wikidata	Wikidata uses semantic technology (Wikidata использует семантические технологии)
ConceptNet	ConceptNet motivated by goal let computers understand what people already know (ConceptNet ставит задачей позволить компьютерам понимать то, что люди уже знают)
ATOMIC	Person X pays Person Y a compliment. Person X wanted to be nice (Персона X делает комплимент Персоне Y. Персона X хочет казаться приятной)



Рис. 1. Классификация методов интеграции знаний

идеи методов интеграции знаний в контексте разработки вопросно-ответных систем с приведением соответствующих примеров и с учетом особенностей конкретных выделенных классов методов. Основное место в данной классификации занимает разделение методов интеграции знаний по использованию баз знаний, при этом под использованием понимается вовлечение сведений непосредственно при получении ответов на запросы, что исключает случаи привлечения баз знаний в процессе предобучения моделей. В свою очередь, для извлечения признаков из данных баз знаний могут использоваться как языковые, так и графовые модели.

МЕТОДОЛОГИЧЕСКАЯ ОСНОВА

Несмотря на различия в использованных подходах к интеграции знаний, можно выделить также и общую методологическую основу в рассмотренных далее вопросно-ответных системах. В частности, это

касается как самой постановки задачи, так и применяемых вспомогательных методов.

Использование в работе системы дополнительного контекста создает определенную специфику с точки зрения функционирования. В связи с этим также начинают приобретать значительную важность такие вспомогательные этапы, как извлечение релевантных данных к запросу. В общем виде этот этап подразумевает определение в поступившем запросе некоторого количества сущностей n : $(q_1, \dots, q_n)$. Для этого в настоящее время на практике преимущественно используются классические методы из области обработки естественного языка, такие как лемматизация и частеречная разметка. Последующая же часть процесса может варьироваться в зависимости от конкретной задачи.

Во многих работах, посвященных интеграции знаний в вопросно-ответных системах, подразумевается наличие у вопроса вариантов ответа. Соответственно, целью системы становится оценка вероятности каждого ответа и выбор наиболее вероятного. Это позволяет существенно упростить и унифицировать построение и оценивание систем. Поэтому в таких случаях будем считать, что выявленным n сущностям соответствуют m аналогично извлеченных сущностей из вариантов ответа: $(a_1, \dots, a_m)$.

На следующем шаге в работу включается некоторая база знаний, которую можно формализовать как $G = (V, E)$, где V – набор сущностей в базе знаний, а $E \subseteq V \times R \times V$ – множество триплетов базы знаний вида «сущность-отношение-сущность». На практике устоявшейся формой представления подобных баз знаний является граф. На основании этого между сущностями, определенными в контексте используемой базы знаний, представляется возможным построить множество путей вида:

$$p = ((q_i, r_p, v_l), (v_l, r_{l+1}, v_{l+1}), \dots, (v_{k-1}, r_k, a_j)),$$

где $i \in (1, \dots, n)$, $j \in (1, \dots, m)$, k – длина пути в графе, $l \in (1, \dots, k)$, q_i – i -я сущность из запроса, a_j – j -я сущность из ответа, а v_l и r_l – соответственно l -я по счету в пути сущность и отношение в графе.

Впоследствии подграф базы знаний или совокупность путей используются в качестве дополнительного контекста для определения наиболее вероятного ответа.

Одной из основных проблем при такой постановке является определение наиболее релевантной по отношению к запросу информации. Возможным инструментом для ее решения может служить механизм внимания (Attention) [10], который позволяет вычислять так называемые веса внимания (Attention Weights), количественно оценивающие степень значимости той или иной информации из контекста. Формально механизм внимания определяется выражением:

$$\begin{aligned} \text{Attention}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) &= \text{softmax} \left(\frac{\mathbf{Q} \times \mathbf{K}^T}{\sqrt{d_{\text{model}}}} \right) \cdot \mathbf{V} = \\ &= \text{Attention weights} \cdot \mathbf{V}, \end{aligned} \quad (1)$$

где $\mathbf{Q} = \text{Query} = \mathbf{X} \times \mathbf{W}_q$, $\mathbf{K} = \text{Key} = \mathbf{X} \times \mathbf{W}_k$, $\mathbf{V} = \text{Value} = \mathbf{X} \times \mathbf{W}_v$, $\mathbf{X} \in \mathbb{R}^{N \times d_{\text{model}}}$ – векторное представление входных данных, $\mathbf{W}_q \in \mathbb{R}^{d_{\text{model}} \times d_k}$, $\mathbf{W}_k \in \mathbb{R}^{d_{\text{model}} \times d_k}$, $\mathbf{W}_v \in \mathbb{R}^{d_{\text{model}} \times d_v}$, N – количество векторов во входных данных, d_{model} , d_k , d_v – размерности векторных представлений в модели и матрицах $\mathbf{K}$ и $\mathbf{V}$, а $\text{softmax}(X_i) = \frac{\exp(X_i)}{\sum_{j=1}^N \exp(X_j)}$.

Таким образом, веса внимания при умножении на векторное представление контекста способны скорректировать нужным образом влияние отдельных его элементов на результат. На практике чаще всего для учета разных аспектов данных в рамках механизма применяется несколько групп различных матриц (так называемых *голов*), а результат их применения конкатенируется и проецируется в нужную размерность с помощью еще одной матрицы, что получило название Multi-Head Attention (многоголовое внимание):

$$\begin{aligned} \text{Multi-Head Attention} &= \\ &= \text{Concatenation}(\text{Attention}_1, \dots, \text{Attention}_z) \times \mathbf{W}_o, \end{aligned} \quad (2)$$

где Attention_i – i -й результат блока Attention, $\mathbf{W}_o \in \mathbb{R}^{z d_v \times d_{\text{model}}}$, z – количество голов внимания.

Механизм внимания играет большую роль во многих моделях глубокого обучения и широко используется при разработке подходов для интеграции знаний. Также в связке с Multi-Head Attention часто используется Feed Forward Neural Network (нейронная сеть прямого распространения), что в совокупности составляет основную часть модели, называемой трансформером. В качестве конкретной реализации Feed Forward Neural Network может выступать, например, многослойный перцептрон, а с практической точки зрения роль этой составляющей трансформера рассматривается в контексте хранения и извлечения усвоенных в процессе обучения закономерностей.

МЕТОДЫ ИНТЕГРАЦИИ ЗНАНИЙ БЕЗ ИСПОЛЬЗОВАНИЯ БАЗ ЗНАНИЙ

В общем случае вовлечение вспомогательных знаний не обязательно подразумевает использование определенных баз знаний. Так, в качестве информации, способствующей получению более точного ответа на запрос, могут выступать похожие примеры с указанием правильного ответа. Например, в работах [11]

и [12] демонстрируется положительный эффект от добавления во входные данные запросов из обучающей выборки на основе близости их векторных представлений эмбедингу⁶ исходного запроса.

Другой тип подходов отталкивается от идеи непосредственно обращаться к усвоенным в процессе предобучения модели сведениям: извлекать их в зависимости от запроса и применять в качестве дополнительных входных данных. В [13] для этого предлагается задавать предобученной модели уточняющие вопросы с помощью шаблонов, а ответы на них использовать в качестве полезного контекста. Аналогичный подход в [14] также подразумевает вовлечение в работу вопросно-ответной системы сгенерированных по запросу вспомогательных данных. При этом для генерации знаний могут применяться и специально обученные модели, как это происходит в исследовании [15], в котором подобная модель создает структурированную информацию в формате путей по базе знаний. Таким образом, рассмотренные выше подходы отталкиваются от идеи предоставления на вход дополнительных сведений, для получения которых могут использоваться предобученные и другие вспомогательные модели, тогда как для работы вопросно-ответной системы в целом может не требоваться дообучение языковых моделей.

Количественно самая обширная группа подходов исходит из концепции предварительного обучения моделей. Множество экспериментальных результатов подтверждают идею, что модели с большим числом весов, обученные на как можно более значительном объеме разноплановой информации, способны показывать лучшие результаты при их последующей адаптации к конкретным условиям [16]. Данная концепция во многом опирается на методологию самообучения (Self-Supervised Learning), которая позволяет извлекать представления из корпусов текстов без необходимости их предварительной разметки. Для осуществления этого разрабатываются специальные задачи, в соответствии с которыми модель и обучается. В частности, в рамках разработки языковой модели BERT⁷ [1] использовались две такие задачи. Первой задачей является предсказание слов в предложении, маскированных специальным токеном. В рамках этой задачи с вероятностью 15% выбираются некоторые токены из предложения, затем 80% из этих токенов маскируются, 10% – заменяются на случайный токен, а остальные 10% – остаются без изменений. В качестве меры ошибки при предсказании моделью маскированного токена может использоваться кросс-энтропия:

$$\text{Cross-entropy} = -\frac{1}{N} \sum_{i=1}^N \mathbf{y}_i \times \log(\bar{\mathbf{y}}_i), \quad (3)$$

где N – общее число примеров; $\mathbf{y}_i$ – one-hot-вектор⁸, кодирующий правильный ответ для i -го примера; $\bar{\mathbf{y}}_i$ – вектор-предсказание модели для i -го примера, обозначающий вероятность соответствия каждому возможному варианту ответа в рамках задачи.

Вторая задача касается определения правильного порядка двух предложений в тексте. Это реализуется через добавление во входные данные в процессе обучения специального токена [CLS], репрезентирующего информацию из предложения в целом, и сводится к задаче бинарной классификации. Цель данной классификации – определение того, является ли некоторое предложение B продолжением для предложения A на основе их итоговых векторных представлений на выходе модели для токенов [CLS]. При этом в рамках обучения в 50% случаев B является случайным предложением, а в других 50% – правильным продолжением, тогда как мерой ошибки для задачи также может служить кросс-энтропия. Итоговая ошибка модели при обучении рассматривается как сумма ошибок по каждой задаче. Общая схема обучения модели BERT представлена на рис. 2:

На этапе предобучения (Pretraining) часть токенов пары предложений (Unlabeled Sentence A and B Pair) маскируется, после чего эмбединги ($E_{[CLS]}$, E_1 , ..., E_N , $E_{[SEP]}$ и E'_1 , ..., E'_M) токенов ($Tok\ 1$, ..., $Tok\ N$ и $Tok\ 1$, ..., $Tok\ M$) предложений A (Masked Sentence A) и B (Masked Sentence B) с добавлением обобщающего и разделительного токенов ([CLS] и [SEP]) поступают на вход трансформера BERT. Полученные итоговые эмбединги (C , T_1 , ..., T_N , $T_{[SEP]}$, T'_1 , ..., T'_M) используются для предсказания маскированных токенов (Mask LM, Language Modeling) и порядка предложений (NSP, Next Sentence Prediction). На этапе дообучения (Fine-Tuning) в зависимости от задачи (MNLI⁹, NER¹⁰, SQuAD¹¹)

⁸ One-hot вектор – бинарный вектор, в котором только один элемент имеет значение 1, а остальные равны 0. [One-hot vector is a binary vector in which only one element has the value 1, and remaining elements are equal to 0.]

⁹ Multi-Genre Natural Language Inference – датасет для задачи Natural Language Inference – установление логической взаимосвязи между фрагментами текста. [Multi-Genre Natural Language Inference – dataset for the Natural Language Inference task – establishes a logical relationship between the text fragments.]

¹⁰ Named-entity recognition – задача распознавания в тексте именованных сущностей. [Named Entity Recognition is the task of recognizing named entities in the text.]

¹¹ Stanford Question Answering Dataset – вопросно-ответный датасет, подразумевает автоматическое получение ответов на вопросы на естественном языке. [Stanford Question Answering Dataset is a QA dataset, which implies automatic answers to questions in natural language.]

⁶ Embedding – векторное представление. [Embedding means a vector representation.]

⁷ Bidirectional encoder representations from transformers.

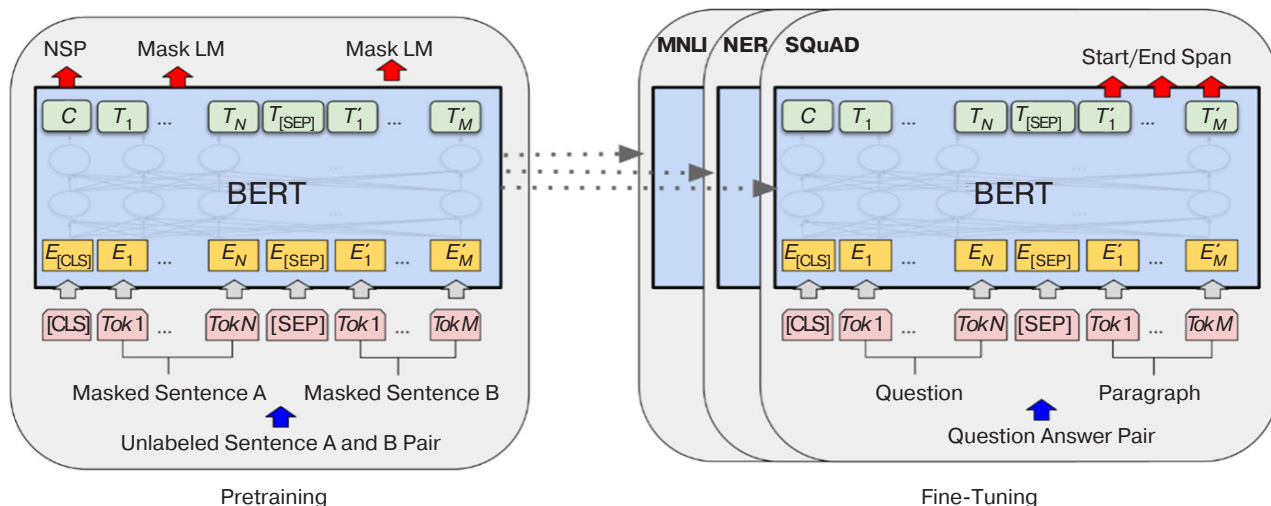


Рис. 2. Схема обучения модели BERT [1]

меняется формат входных данных и предсказываемого. В случае вопросно-ответного датасета SQuAD на вход подается вопрос (Question) и соответствующий контекст (Paragraph), а на выходе предсказывается положение в контексте правильного ответа (Start/End Span).

Впоследствии данная методология видоизменялась и адаптировалась в рамках предобучения и других языковых моделей. В контексте вопросно-ответных систем было разработано множество подходов, основанных на модификации и расширении задач для самообучения BERT или замене их на другие. В общем и целом, при создании такого рода моделей чаще всего модифицируется процедура маскирования за счет введения ограничений на то, что должно маскироваться в предложении, а также изменения параметров маскирования при обучении.

Одной из первых и наиболее значимых разработок в этом направлении стала модель Enhanced Language Representation with Informative Entities (ERNIE) [17], схема которой изображена на рис. 3. Основная ее идея заключается в том, что, если также предобучать предсказывать выявленные в тексте на основе базы знаний маскированные именованные сущности (Entities) в рамках дополнительной задачи для самообучения, то это способно улучшить понимание языка моделью, а также контекстуализировать ее определенные знания о мире. В частности, для этой цели в 5% случаев для токенов текста соответствующая им именованная сущность заменяется на случайную, а в 15% случаев – сущность маскируется, и ее требуется предсказать по текстовым токенам. Кроме того, в работе вводится механизм взаимодействия между эмбедами сущностей и соответствующих токенов текста, что позволяет привнести дополнительную информацию в оба векторных представления, увеличив тем самым

точность предсказания правильных токенов. С этой целью вводится промежуточное векторное представление, объединяющее информацию на уровне токенов и именованных сущностей, за счет которого впоследствии обновляются исходные эмбединги токенов и сущностей, что задается выражениями:

$$\mathbf{h}_j = \sigma(\tilde{\mathbf{W}}_t \tilde{\mathbf{w}}_j + \tilde{\mathbf{W}}_e \tilde{\mathbf{e}}_k + \tilde{\mathbf{b}}),$$

$$\mathbf{w}_j = \sigma(\mathbf{W}_t \mathbf{h}_j + \mathbf{b}_t), \quad (4)$$

$$\mathbf{e}_k = \sigma(\mathbf{W}_e \mathbf{h}_j + \mathbf{b}_e),$$

где $\mathbf{h}_j$ – агрегированное векторное представление для токена с номером j , σ – заданная нелинейная функция активации, $\tilde{\mathbf{w}}_j$ и $\mathbf{w}_j$ – векторные представления токена j до и после интеграции знаний, $\tilde{\mathbf{e}}_k$ и $\mathbf{e}_k$ – векторные представления соответствующей токеноу j именованной сущности k до и после интеграции знаний, $\tilde{\mathbf{W}}$, $\mathbf{W}$, $\tilde{\mathbf{b}}$ и $\mathbf{b}$ – параметры модели.

В процессе обучения модели ERNIE поступившие на вход эмбединги токенов текста (Token Input) проходят через N слоев трансформера (T-Encoder), после чего вместе с эмбедами именованных сущностей (Entity Input) они обрабатываются M слоями агрегатора (K-Encoder). На каждом слое i агрегатора эмбединги сущностей ($\mathbf{e}_1$ и $\mathbf{e}_2$) и текста ($\mathbf{w}_1, \dots, \mathbf{w}_n$) проходят через свой блок Multi-Head Attention, а соответствующие обновленные эмбединги сущностей ($\tilde{\mathbf{e}}_1$ и $\tilde{\mathbf{e}}_2$) и текста ($\tilde{\mathbf{w}}_1, \dots, \tilde{\mathbf{w}}_n$) попадают в блок интеграции знаний (Information Fusion), на выходе которого, согласно формулам (4), получают эмбединги сущностей (Entity Output) и текста (Token Output) с учетом интеграции знаний.

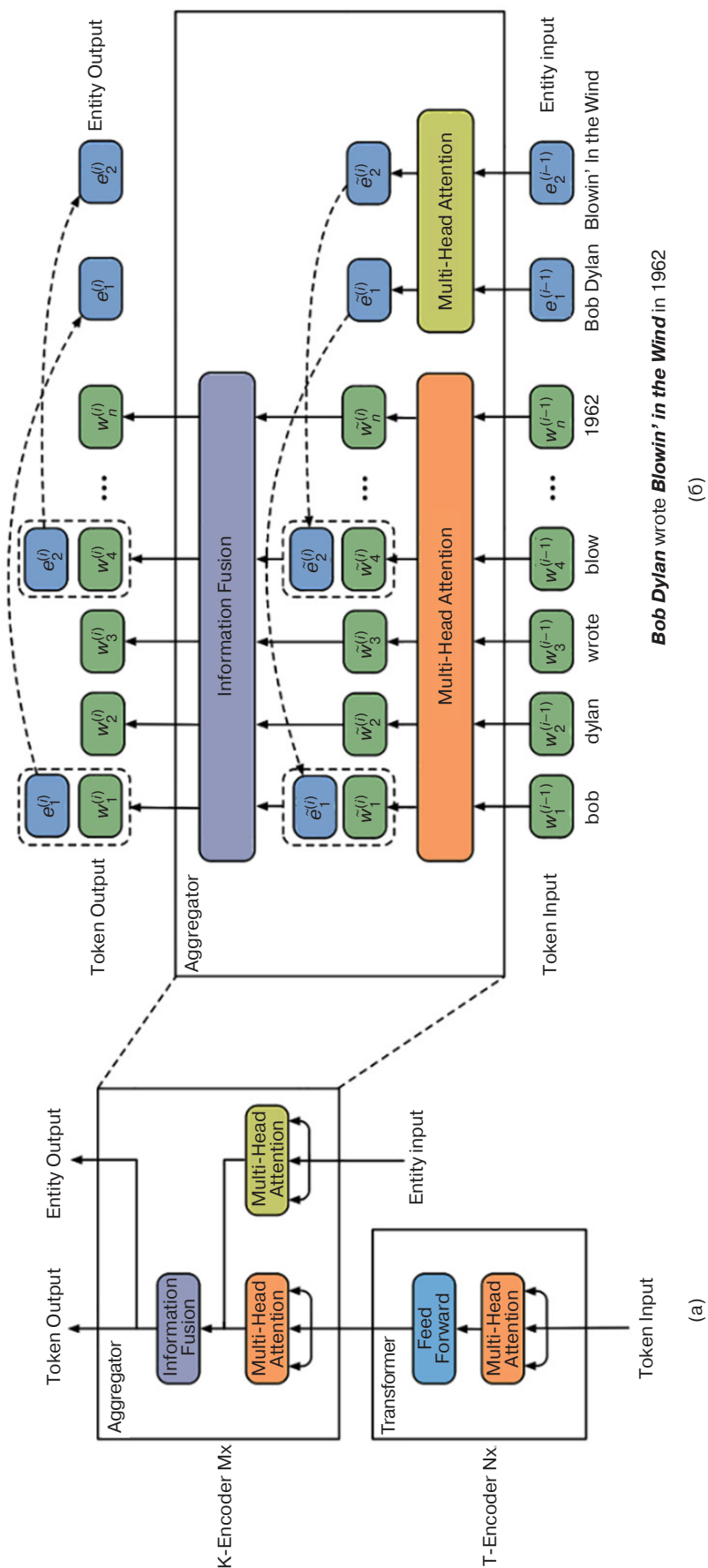


Рис. 3. Схема обучения модели ERNIE [17]: (а) архитектура модели, (б) архитектура агрегатора

Схожий метод лежит в основе модели KnowBERT [18], однако вовлечение внешней информации происходит на уровне векторных представлений сущностей, которые обновляются с помощью механизма внимания и за счет добавления предобученных эмбеддингов сущностей из базы знаний, что впоследствии оказывает влияние и на эмбеддинги всех токенов через механизм внимания, согласно формуле:

$$\mathbf{H}'_i = \text{MLP}(\text{MHA}(\mathbf{H}_i, \mathbf{S}'^e, \mathbf{S}'^e)), \quad (5)$$

где $\mathbf{H}'_i$ – векторное представление токена i после интеграции знаний, MLP – многослойный перцептрон (Multi-layer Perceptron), MHA – Multi-Head Attention, $\mathbf{H}_i$ – векторное представление токена i до интеграции знаний, $\mathbf{S}'^e$ – обновленные векторные представления выявленных именованных сущностей.

Концептуально схожая с ERNIE архитектура предложена в работе [19], основным ее отличием является использование информации об отношениях между сущностями, предсказание которых представлено отдельной задачей для предобучения. В рамках метода Weakly Supervised Knowledge-Pretrained Language Model [20] вместо маскирования сущностей на этапе предобучения модели в качестве дополнительной задачи модель учит предсказывать, были ли сущности во входных данных заменены на другие такого же типа в рамках базы знаний Wikidata. Архитектура модели при этом соответствует BERT, однако маскирование токенов осуществляется только в 5%, а не в 15%, чтобы избежать маскирования слишком большого фрагмента контекста, т.к. сущности могут состоять из нескольких слов. В работе [21] вдобавок к маскированию слов и именованных сущностей применяется и маскирование словосочетаний, что улучшает понимание сочетаемости слов моделью, а интеграция осуществляется поэтапно: на каждом этапе модель, подобная BERT, тренируется только на одном типе маскирования. Использование нескольких режимов обучения, при которых модель переключается с предсказания слов на предсказание фраз в зависимости от того, в каком режиме между двумя последними последовательными итерациями наблюдалось наибольшее снижение ошибки модели относительно общего снижения ошибки за все итерации, является основным новшеством [22].

Авторы исследования [23] предобучают модель на основе BERT, ставя целью обучение предсказанию маскированных сущностей по их описанию, а также сближение векторных представлений описаний синонимичных сущностей и отдаление антонимичных, для чего используется специальная функция потерь:

$$L = -\sum \log \frac{f(\mathbf{h}_{\text{ori}}, \mathbf{h}_{\text{syn}})}{f(\mathbf{h}_{\text{ori}}, \mathbf{h}_{\text{syn}}) + f(\mathbf{h}_{\text{ori}}, \mathbf{h}_{\text{ant}})}, \quad (6)$$

где $f(\mathbf{h}_i, \mathbf{h}_j) = \exp(\mathbf{h}_i \mathbf{h}_j)$, $\mathbf{h}_{\text{ori}}$ – векторное представление описания маскированной сущности, $\mathbf{h}_{\text{syn}}$ – векторное представление описания синонимичной сущности, $\mathbf{h}_{\text{ant}}$ – векторное представление описания антонимичной сущности.

Для решения конкретных задач эта модель используется в паре с BERT в качестве дополнительного источника знаний в форме векторных представлений выявленных сущностей, а их инъекция осуществляется через конкатенацию выходов модели с выходами BERT с опциональным применением механизма внимания для учета важности данных по конкретным сущностям. При этом рассматривается использование механизма внимания как для выходных представлений с последних слоев моделей, так и для выходных представлений по слоям моделей с применением усреднения, а результат работы механизма внимания конкатенируется с выходом модели BERT вместо выхода вспомогательной модели. Лингвистические признаки играют важную роль и в работе [24], в которой дополнительной задачей для самообучения является определение семантической близости пары слов, тогда как обучается модель, чередуя задачи для самообучения BERT с дополнительной. Схожая идея представлена в статье [25], где на основе данных WordNet¹² модель тренируется классифицировать слова по группам со схожим значением.

Логичным развитием подходов с маскированием сущностей и отношений является использование на этапе предобучения моделей задачи предсказания структурированных единиц знания в форме триплетов, что может позволить усвоить более общие принципы и взаимосвязи, подобные содержащимся в базах знаний. Так, в модели Knowledge Embedding and Pretrained Language Representation [26] векторные представления элементов триплета рассматриваются как векторные представления их описаний из базы знаний, полученные с помощью той же модели, что используется для генерации векторных представлений токенов текста в задаче предсказания маскированных токенов. При этом для расчета ошибки в задаче предсказания триплетов применяется скоринговая функция из модели для получения эмбеддингов графов знаний TransE [27]:

¹² Лексическая база данных английского языка, разработанная в Принстонском университете. <https://wordnet.princeton.edu/>. Дата обращения 01.12.2024. [A lexical database of the English language developed at Princeton University. <https://wordnet.princeton.edu/>. Accessed December 01, 2024.]

$$d(\mathbf{h}, \mathbf{t}) = \|\mathbf{h} + \mathbf{r} - \mathbf{t}\|, \quad (7)$$

где $\mathbf{h}$ – векторное представление субъекта в триплете, $\mathbf{r}$ – векторное представление отношения в триплете, $\mathbf{t}$ – векторное представление объекта в триплете.

В работе [28] на вход модели подается набор триплетов из одного подграфа, и поэтому в рамках механизма внимания дополнительно используется матрица смежности для учета существующих взаимосвязей, а обучение проводится в формате восстановления триплетов, что подразумевает составление триплетов из обновленных представлений вершин и использование скоринговой функции (7). Исследование [29] содержит идею предобучения трех функций на основе модели-энкодера для предсказания каждого элемента триплета по двум другим, что должно способствовать выучиванию возможных сочетаний. В этой постановке оценка ответа является произведением оценок схожести значений трех предобученных функций и соответствующих реальных элементов триплета, у которого субъект представляет собой контекст вопроса, отношение – сам вопрос, а объект – конкретный вариант ответа. Предобучение функции, которая будет участвовать в нахождении ответов на запросы, определяя наиболее вероятные взаимосвязи со вспомогательными данными из базы знаний, предлагается в статье [30]. С помощью этой функции извлеченные вспомогательные факты к каждому варианту ответа сопоставляются по степени схожести с фактами для вопроса, и более вероятным ответом считается тот, для которого эта схожесть фактов в среднем окажется выше.

В то же время триплеты из релевантного подграфа к запросу могут непосредственно подаваться на вход модели в рамках предобучения наравне с обычными текстовыми токенами с применением специальных эмбеддингов для указания типа токена, как это показано в [31]. В связи с этим при реализации механизма внимания в модели используется матрица-маска, ограничивающая взаимодействие несвязанных в подграфе вершин. В работе [32] было предложено усовершенствовать подход модели ERNIE путем модифицирования представления сущностей за счет учета их связей в соответствующем подграфе, а также использования механизма внимания для отфильтровывания потенциально нерелевантного контекста для запроса.

Другим способом использования баз знаний на этапе предварительного обучения модели может быть построение на их основе новых вопросно-ответных датасетов, с помощью которых система также определенным образом улучшает свои возможности находить верные ответы. Такой подход используется в [33], а в [34] его развивают с помощью концептуализации: конкретные факты рассматриваются

в более общем ключе, благодаря чему можно охватить большее количество ситуаций и усовершенствовать способность различать схожие варианты. Например, с помощью базы ATOMIC «игра в футбол» может быть представлена как «утомительное событие».

Концепция модели Self-supervised Bidirectional Encoder Representation Learning of Commonsense [35] в большей степени акцентирована на количественном расширении числа задач для самообучения. Для того чтобы улучшить возможности системы при обработке трудных запросов, к задачам для самообучения BERT в ней было добавлено еще 3: первая нацелена на различение контекста с противоположным смыслом; вторая требует расстановки по порядку нескольких перемешанных предложений, взятых из одного абзаца; третья расширяет усвоение контекстных взаимосвязей через маскирование сущностей. По задумке авторов такой подход также позволит системе лучше улавливать языковые закономерности и обеспечит появление более универсальных реализаций.

Другая концепция интеграции знаний подразумевает в качестве дополнительного шага дообучение на основе соответствующих практической задаче существующих датасетов. Например, определенное распространение в связи с этим получило использование датасета SQuAD [36] из области построения вопросно-ответных систем. К его ключевым особенностям можно отнести относительно существенный размер (более 100000 запросов), при том, что к каждому запросу прилагается соответствующий контекст, взятый из Wikipedia. Таким образом, в результате обучения на данном датасете модель лучше адаптируется к постановке и формату задачи, а также в дополнение обрабатывает достаточно значительный объем данных, увеличивая тем самым количество усвоенных фактических сведений.

В качестве актуального и характерного примера в связи с этим можно упомянуть модель UnifiedQA [37], разработка которой строилась через дообучение языковой модели на 8 вопросно-ответных датасетах разного типа, что позволяет адаптироваться к существующим форматам бенчмарков и обеспечивает прирост точности модели на неиспользованных в процессе обучения вопросах, открывая также и новые возможности для последующего ее дообучения. Целесообразность подобного подхода была подтверждена и для модели Unicorn из [38], однако в данном случае область исследования была ограничена исключительно датасетами, подразумевающими использование общих знаний (Commonsense Question Answering).

Рассмотренные выше методы можно отнести к классу подходов без явного привлечения баз

знаний, т.к. при использовании соответствующих систем не подразумевается непосредственное извлечение контекста к запросу именно из баз знаний, а акцент создается на знаниях, которые были получены в процессе обучения. По сути, разработанные в последние несколько лет большие языковые модели в существенной степени основываются на схожей идее: обучение на большом количестве качественных данных, учитывающих различную специфику и человеческие предпочтения, способно повышать универсальность систем и оценку получаемых с их помощью результатов при достаточном масштабировании этого процесса.

Среди преимуществ данного класса можно назвать в некотором смысле большую универсальность благодаря независимости от использования баз знаний. Кроме того, существенная опора в архитектуре вопросно-ответной системы на предобученные и дообученные модели позволяет несколько упростить ее разработку, последующее использование и адаптацию к конкретным задачам.

В то же время этот класс подходов может считаться в определенной степени ограниченным в своих возможностях для дальнейшего развития. Дело в том, что в основном прирост эффективности моделей здесь связан с целенаправленным и точечным улучшением, расширением объема усвоенных сведений, тогда как никаких принципиально новых механизмов, улучшающих способности системы к рассуждению, не вводится. Кроме того, в рамках данного направления практически никак не решается проблема отсутствия интерпретируемости у полученных ответов и их обоснования системой, а также устаревания знаний.

МЕТОДЫ ИНТЕГРАЦИИ ЗНАНИЙ С ИСПОЛЬЗОВАНИЕМ БАЗ ЗНАНИЙ

Основной единицей информации в базах знаний можно считать триплеты «сущность-отношение-сущность», тогда как более сложные взаимосвязи могут передаваться набором триплетов – путем. Помимо путей или вместо них могут использоваться также подграфы базы знаний, которые можно рассматривать как совокупность путей, имеющих общие элементы. В результате с точки зрения обработки информации в наличии у вопросно-ответной системы может оказаться 3 типа признаков в том или ином сочетании:

- 1) признаки, полученные обработкой контекста запроса языковой моделью;
- 2) признаки, основанные на извлеченных путях;
- 3) признаки, связанные с подграфами базы знаний.

Таким образом, одним из направлений для исследований являются способы эффективной обработки

этих разных типов признаков и методы объединения соответствующих результатов.

Один из подходов к интеграции признаков в работу системы исходит из того, что триплеты и их совокупности зачастую достаточно просто могут быть транслированы в предложения на естественном языке, и в такой форме их представляется возможным подать на вход языковой модели в качестве вспомогательного контекста. При этом нужно отметить, что в такой форме они могут служить также и в качестве обоснования для получаемого ответа. Примером реализации подобного подхода может служить модель Knowledge-Augmented language model PromptING (KAPING) [39], а в работе [40] его эффективность исследуется в условиях использования предобученных на вспомогательных датасетах языковых моделей. В модели DEscriptive Knowledge for COmmonsense question answERING (DEKCOR) [41], помимо извлеченных из ConceptNet триплетов, на вход подаются и словарные определения соответствующих сущностей, тогда как в Knowledgeable External Attention for commonsense Reasoning (KEAR) [42] (рис. 4) контекст к запросу расширяется за счет привлечения дополнительной информации из ряда вопросно-ответных датасетов, что позволяет использовать более специфические сведения.

В частности, в рамках архитектуры KEAR к конкатенации вопроса с одним из вариантов ответов (Question & Candidate) на вход модели добавляются соответствующие извлеченные вспомогательные данные (Knowledge Retrieval) из базы знаний ConceptNet, словаря Wiktionary (Definition) и дополнительных датасетов (Training Data). На вход трансформера подаются эмбединги токенов запроса ($E_{[CLS]}, E_0, \dots, E_N$), к которым добавляется индикатор сегмента (S_0) и вспомогательного контекста ($E_0^k, \dots, E_{N_k}^k$), к которым добавляется индикатор сегмента (S_1). Вероятность ответа (Score Prediction) определяется на основании итогового эмбединга вспомогательного токена $E_{[CLS]}$, полученного с помощью механизма внимания (Self-Attention / External Attention).

Преимуществом интеграции знаний через языковые модели является возможность в значительной степени опираться на эффективность предобученных моделей, при этом в некоторых случаях даже избегая необходимости изменять их веса (например, модель KAPING). Также вычислительные затраты на получение дополнительных признаков могут считаться относительно небольшими. В то же время, т.к. большинство предобученных моделей могут эффективно использовать только фиксированное количество информации из входных данных, возникает необходимость ограничивать количество менее релевантных сведений.

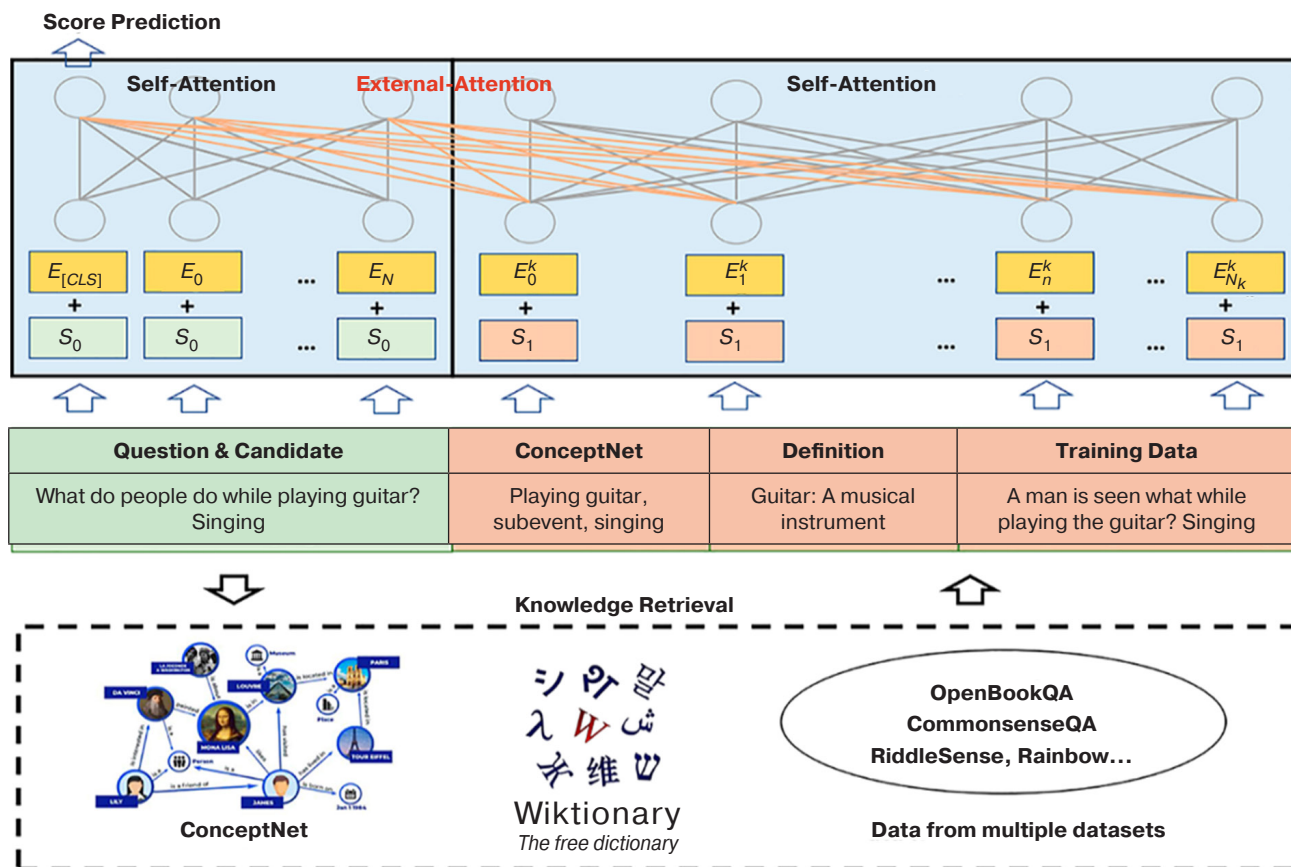


Рис. 4. Архитектура модели KEAR [42]

В простейшем случае для этого может использоваться ограничение на число триплетов (например, не более 3 последовательных) или набор некоторых эвристик, учитывающих особенности конкретной базы знаний. В [43] с целью вовлечения в работу модели только потенциально релевантных метаданных из Wikidata оценивается увеличение вероятности правильного ответа с учетом соответствующего прироста информации (Information Gain):

$$P(y | k_m) = 2^{\text{pmi}(y, k_m)} P(y), \quad (8)$$

где y – правильный ответ; k_m – определенный паттерн, содержащий метаданные m ;

$$\text{pmi} = \log \left(\frac{P(y, k_m)}{P(y)P(k_m)} \right).$$

Исследование [44] показывает целесообразность ранжирования извлеченных дополнительных сведений на основе определения их важности через вспомогательную предобученную модель, а в [45] в дополнение к ранжированию показан положительный эффект от использования взвешенного суммирования эмбедингов знаний при их интеграции, что позволяет сделать акцент на более существенных фактах.

Также, для того чтобы интегрировать разнородные данные, может применяться механизм внимания,

который позволяет агрегировать признаки с учетом их важности для задачи. Иначе говоря, более релевантными будут считаться извлеченные знания, которые имеют больше семантической связи с запросом, что определяется на основании операций над векторными представлениями. Чаще всего на практике на основе действий над эмбедингами получают веса внимания, которые позволяют обновить соответствующие векторные представления с учетом конкретной задачи и ее контекста. В частности, подобный подход представлен в работах [46–48], а в статье [49] вспомогательные знания также фильтруются на основании частоты встречаемости сущностей и соответствующих путей, тогда как для интеграции знаний дополнительно используется сигмоидальная функция, которая регулирует то, как сильно они повлияют на обновление контекста для запроса в целом. В исследовании [50] помимо механизма внимания для отфильтровывания нерелевантных данных предложено использовать графовые подходы для определения важности отдельных вершин в извлеченном подграфе: расчет степени близости вершин, PageRank¹³

¹³ Алгоритм ранжирования, который оценивает количество и качество ссылок, ведущих на веб-страницы. [A ranking algorithm that evaluates the number and quality of links leading to web pages.]

и его модификация, что позволяет учитывать только наиболее информативные пути.

Кроме того, недостатком интеграции знаний через языковые модели является ограниченное использование структурированности баз знаний, что может снижать потенциальную эффективность итоговой реализации. Для сохранения эффекта от учета связей при переводе триплетов в текст и предотвращения смешения информации в модели K-BERT [51] соответствующие позиционные данные включаются на этапе получения эмбеддингов, а в последующих вычислениях прибегают к специально вводимой в статью матрице видимости (Visible Matrix), элементы которой фиксируют то, с какими токенами должен взаимодействовать в заданном контексте конкретный токен.

В связи с этим необходимо упомянуть один из главных инструментов для обработки структурированных знаний – графовые нейронные сети. Этот инструмент позволяет получать и обновлять векторные представления вершин графов с помощью концепции передачи сообщений (Message Passing):

$$\mathbf{h}_u = \phi(x_u, \bigoplus_{v \in N_u} \psi(x_u, x_v, e_{uv})), \quad (9)$$

где $\mathbf{h}_u$ – векторное представление вершины u ; x_u и x_v – признаки вершин u и v ; e_{uv} – признаки ребра между вершинами u и v ; ϕ и ψ – заданные дифференцируемые функции; $\bigoplus_{v \in N_u}$ – инвариантный к перестановкам оператор агрегирования, действующий на соседей вершины u .

Благодаря этому на практике вектор представления каждой сущности может использовать различные полученные из базы знаний контекстные данные, учитывая определенным образом информацию от своих соседей по графу. Примером используемой в контексте инъекции знаний модели такого рода является графовая сверточная нейронная сеть (Graph Convolutional Network) [52].

Полученные с помощью графовых нейронных сетей признаки впоследствии также могут интегрироваться в работу системы с помощью механизма внимания. Среди реализаций такого рода можно назвать архитектуру модели [53], изображенную на рис. 5. Здесь векторное представление каждой вершины вспомогательного подграфа перед его непосредственным использованием для получения ответа корректируется с учетом релевантности относительно имеющегося векторного представления запроса:

$$\alpha_i = \frac{\mathbf{h}^c \sigma(\mathbf{W} \mathbf{h}_i)}{\sum_{j \in N} \mathbf{h}^c \sigma(\mathbf{W} \mathbf{h}_j)}, \quad (10)$$

где α_i – степень релевантности вершины i ; $\mathbf{h}^c$ – векторное представление контекста запроса; $\mathbf{W}$ – матрица

весов; $\mathbf{h}_i$ – векторное представление вершины i ; N – множество индексов вершин, соседних по отношению к вершине i .

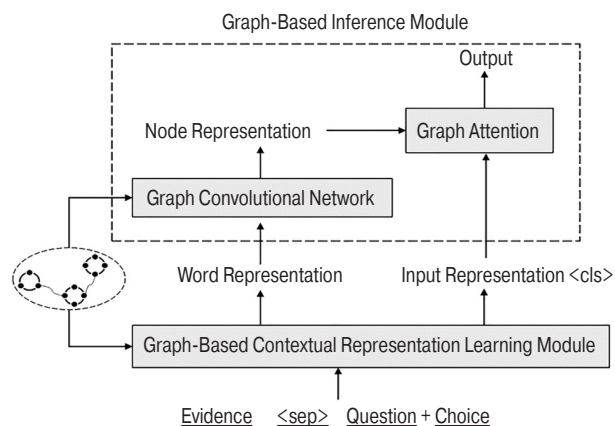


Рис. 5. Архитектура модели из [53]

В целом же схема работы модели устроена следующим образом: сначала по имеющемуся запросу, состоящего из вопроса (Question) и одного из вариантов ответа (Choice) извлекается подграф со вспомогательными сведениями. Эти сведения в форме текста (Evidence) присоединяются к запросу и подаются на вход языковой модели (Graph-Based Contextual Representation Learning Module). Полученные на выходе модели эмбеддинги токенов текста (Word Representation) попадают в модуль интеграции знаний (Graph-Based Inference Module), где используются для инициализации соответствующих вершин вспомогательного графа, которые впоследствии обновляются с помощью графовой сверточной сети (Graph Convolutional Network). Итоговые эмбеддинги вершин графа (Node Representation) затем агрегируются с помощью механизма внимания (Graph Attention) с учетом важности относительно эмбеддинга текстового контекста (Input Representation), а результирующий эмбеддинг графа вместе с текстовым эмбеддингом непосредственно используются для предсказания вероятности ответа с помощью многослойного перцептрона.

Схожим образом устроена модель Multi-hop Graph Relation Networks (MHGRN) [54], но ее ключевым отличием является рассмотрение вспомогательного подграфа в виде совокупности связывающих вершины путей, в соответствии с чем векторное представление каждой вершины обновляется на основании заданной длины путей из нее. Для агрегирования информации по путям вводятся специальные веса внимания, которые определяются как условная вероятность заданной последовательности триплетов с учетом имеющего контекста для запроса, тогда как для расчета вероятности конкретного ответа итоговые эмбеддинги для соответствующих ответу

сущностей агрегируются с помощью механизма внимания и вместе с эмбедингом контекста запроса обрабатываются многослойным перцептроном. Таким образом, в рамках данного подхода учитывается и степень важности отношений между сущностями. В модели Joint reasoning with Language models and Knowledge graphs (JointLK) [55] используется отсечение наименее релевантных относительно запроса вершин вспомогательного подграфа, а также дополнительно вводится новое представление контекста запроса, которое учитывает степень его важности относительно подграфа и является третьей компонентой для получения оценки ответа наравне с исходным представлением контекста и эмбедингом подграфа. В исследовании [56] в рамках механизма обмена сообщениями реализуется последовательное обновление и эмбедингов сущностей, и отношений, которые в данном случае также используются для оценки вероятности ответа. При этом для формализации релевантности отношений между вершинами при заданном контексте запроса используется модифицированная матрица смежности, элементами которой являются соответствующие веса внимания. В модели Knowledge-aware graph Network (KagNet) [57] обновленные с помощью механизма обмена сообщениями векторные представления вершин вспомогательного графа рассматриваются в качестве элементов путей, соединяющих сущности из вопроса и одного из вариантов ответа. В итоге по каждой такой паре сущностей генерируется вектор структурированных признаков, как усреднение эмбедингов соединяющих их путей, и вектор текстовых признаков, получаемый как результат применения многослойного перцептрона к конкатенации эмбедингов запроса и каждой сущности из пары. Для оценки вероятности конкретного ответа реализуется усреднение по всем парам сущностей из запроса и ответа. Кроме того, в дополнение вместо усреднений авторами также предлагается использовать механизм внимания при агрегировании признаков.

Одним из определенных недостатков использования графовых нейронных сетей является увеличение числа параметров в модели и, соответственно, ресурсов на ее обучение и использование. В связи с этим в работе [58] предлагается упрощенный алгоритм получения эмбедингов триплетов на основании one-hot-векторов, указывающих тип сущности в графе и определенное отношение в рамках базы ConceptNet. Для расчета итоговой вероятности ответа в модели используются две оценки: по текстовым и графовым признакам. Первая основывается на обработке многослойным перцептроном векторного представления запроса, а вторая – взвешенной суммы эмбедингов путей, учитывающей частоту их встречаемости.

Несколько более сложным может представляться процесс интеграции знаний в случае наличия нескольких источников сведений и обучения на разных типах задач. В таких условиях приходится решать проблемы, связанные с необходимостью переучивать веса модели и вытеснением выученных фактов новыми, что способно вести к нестабильности результатов. Одним из возможных решений является использование адаптеров [59] – специальных модулей, ориентированных под конкретный источник данных или задачу, что позволяет не изменять веса основной модели и обучать только относительно небольшое число весов адаптера, а также тем самым избежать смешения знаний. При этом на практике обычно независимо обучается несколько различных адаптеров, которые затем уже используются совместно для решения определенной задачи. Так, в модели [60] применяется два вида адаптеров: первый ориентирован на усвоение общих фактов из баз знаний, тогда как второй – на лингвистические сведения. В рамках архитектуры каждый выход со слоя модели-трансформера подается на вход соответствующего слоя адаптера, в результате чего на последнем слое адаптера формируются определенные вспомогательные признаки, которые могут использоваться для предсказания ответа вместе с выходами последнего слоя трансформера. В работе [61] (рис. 6) реализован несколько иной подход, при котором веса предобученных на данных из баз знаний ATOMIC, ConceptNet, WikiData и WordNet адаптеров также не меняются при тренировке модели на конкретной задаче, а вместо этого интеграция знаний осуществляется за счет механизма внимания (формула (1)), где адаптеры формируют Value и Key, а предобученный трансформер – Query:

На каждом слое модели входные данные проходят через слой трансформера и попадают в блок интеграции знаний (Zero-shot Fusion) как прямую (круг 4), так и после взаимодействия с моделями-адаптерами (круги 1, 2 и 3). В этом блоке происходит взаимодействие эмбедингов в рамках механизма внимания (формула (1)): выходное представление из трансформера используются как query, тогда как выходы с адаптеров выступают в качестве Value и Key. Впоследствии результат блока интеграции знаний суммируется с выходом из блока Multi-Head Attention трансформера и нормализуется (Add & Norm). Целью обучения модели в рамках данной архитектуры является способность обращаться к более релевантному адаптеру, что в определенной степени напоминает концепцию смеси экспертов [62].

Отдельно в контексте инъекции знаний можно выделить группу подходов, опирающихся на использование так называемых токенов

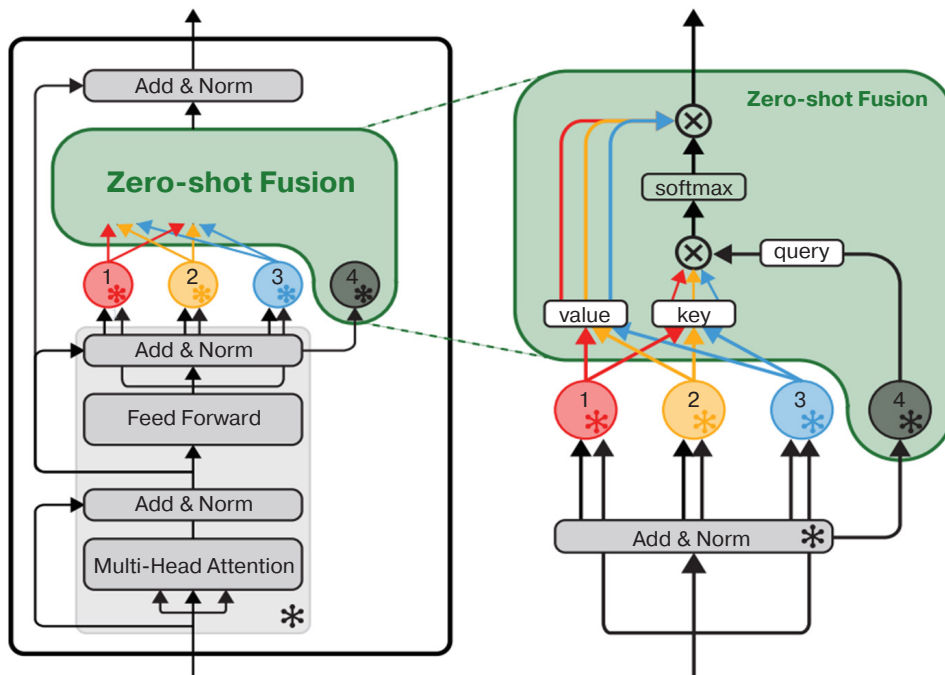


Рис. 6. Схема интеграции знаний с помощью адаптеров из [61]

взаимодействия (Interaction Token). Концепция токенов взаимодействия в значительной степени сходна с идеей применения специального токена [CLS] в языковых моделях, который может служить для классификации целого фрагмента текста. Аналогично токены взаимодействия в случае текстовой информации или вершины взаимодействия (Interaction Node) в случае графов способны выступить промежуточным вместилищем необходимых сведений для объединения разнородных данных. Пример соответствующей вопросно-ответной архитектуры можно увидеть на рис. 7: в рамках модели Graph REASoning Enhanced Language Model (GreaseLM) [63] текстовая и структурированная информация обрабатываются независимо, а их интеграция осуществляется за счет обновления векторных представлений токена взаимодействия и вершины взаимодействия посредством применения двуслойного перцептрона к их конкатенации:

$$[\mathbf{h}_{\text{int}}; \mathbf{e}_{\text{int}}] = \text{MInt}([\tilde{\mathbf{h}}_{\text{int}}; \tilde{\mathbf{e}}_{\text{int}}]) = \text{MLP}([\tilde{\mathbf{h}}_{\text{int}}; \tilde{\mathbf{e}}_{\text{int}}]), \quad (11)$$

где $\tilde{\mathbf{h}}_{\text{int}}$ – векторное представление токена взаимодействия до интеграции знаний, $\tilde{\mathbf{e}}_{\text{int}}$ – векторное представление вершины взаимодействия до интеграции знаний, MInt (modality interaction layer) – слой взаимодействия модальностей.

При обучении GreaseLM конкатенация вопроса с одним из вариантов ответа обрабатывается с помощью N слоев модели-энкодера (Uni-modal Encoder)

и вместе со вспомогательным графом (KG Retrieval) проходит через M слоев (GreaseLM Layer) блока интеграции знаний (Cross-modal Fuser). На каждом слое этого блока эмбединги текстовых токенов ($h_1, \dots, h_T, h_{\text{int}}$) и вершин графа ($e_{\text{int}}, e_1, \dots, e_J$) обрабатываются соответственно слоями языковой (LM Layer) и графовой моделей (GNN Layer), а сам процесс интеграции осуществляется через взаимодействие (MInt, формула (11)) эмбедингов специальных токенов ($\tilde{h}_{\text{int}}$ и $\tilde{e}_{\text{int}}$). После завершения процесса интеграции знаний эмбединги специальных токенов с агрегированным с помощью механизма внимания эмбедингом графа (Pooling) используются для выбора ответа (Answer Selection) посредством перцептрона (MLP).

В рамках модели DRAGON¹⁴ [64] архитектура GreaseLM рассматривалась в контексте самообучения моделей: после слоя интеграции знаний полученные текстовые признаки служат для предсказания маскированных токенов текста, а графовые – для задачи предсказания связи (Link Prediction), которая подразумевает установление степени вероятности наличия связи между вершинами в графе с применением скоринговых функций, подобных (7).

В модели QA-GNN [65] применяется только вершина взаимодействия, инициализируемая векторным представлением текстового контекста из запроса, на основании близости к которой, определяемой с помощью предобученной модели, оценивается

¹⁴ DRAGON — Deep Bidirectional Language-Knowledge Graph Pretraining.

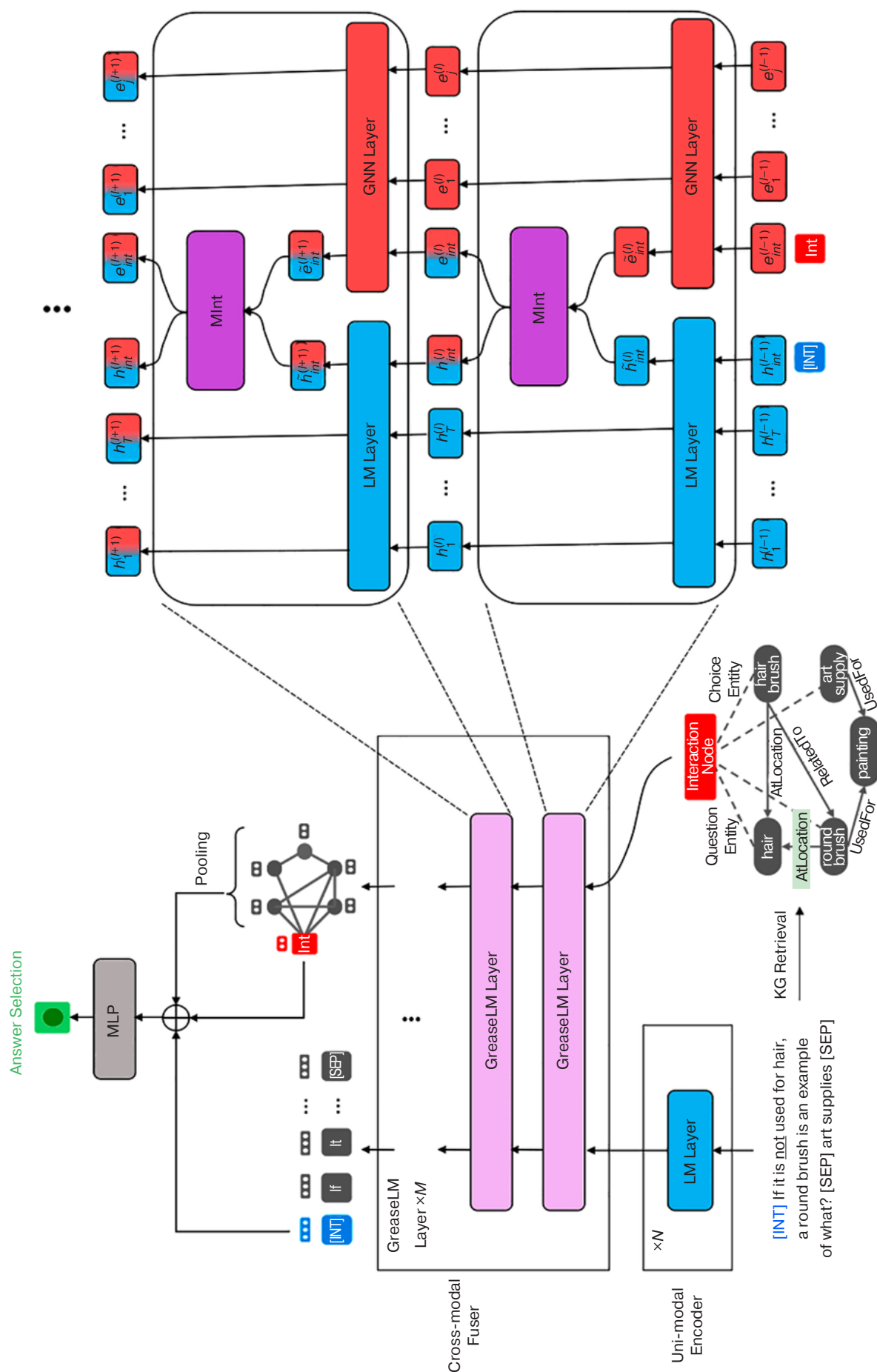


Рис. 7. Архитектура модели GreaseLM [63]

важность других вершин. Данные оценки вместе с признаками, репрезентирующими типы вершин и отношений в виде one-hot-кодирования, используются для расчета весов внимания, с учетом которых реализуется передача сообщений между вершинами и соответствующее обновление их эмбедингов. Процесс выбора ответа при этом также по сути формулируется аналогично модели GreaseLM. В PipeNet [66] по сравнению с QA-GNN вычисление близости вершин вспомогательного графа контексту запроса в каком-то смысле заменяется алгоритмом отсекаания нерелевантных вершин, основанном на определении кратчайшего расстояния между сущностями в рамках соответствующего запросу графа языковых зависимостей:

$$D(c_q) = - \frac{\sum_{i=1}^{|V_a|} \text{Dist}(c_q, c_a)}{|V_a|}, \quad (12)$$

где $D(c_q)$ – релевантность сущности c_q из запроса, $\text{Dist}(c_q, c_a)$ – кратчайшее расстояние между сущностью c_q из запроса и сущностью c_a из соответствующего варианта ответа, V_a – множество сущностей из варианта ответа для запроса.

Остальная же часть архитектуры QA-GNN по сути сохраняется, за исключением использования оценки важности вершин при расчете весов внимания.

Подводя некоторый итог методам интеграции знаний с помощью графовых моделей, можно констатировать, что они характеризуются наибольшим разнообразием используемых идей, которые демонстрируют широкие возможности для учета особенностей структурированных знаний и их включения в работу вопросно-ответных систем. Положительным аспектом можно также считать возможность повысить интерпретируемость модели за счет формирования цепочек фактов с помощью баз знаний, которые при этом могут своевременно отдельно обновляться в зависимости от происходящих событий. В то же время полноценная интеграция графовых признаков приводит к существенному усложнению архитектур моделей, а также, в зависимости от реализации, может требовать определенных дополнительных вычислительных ресурсов, вследствие чего польза от применения интеграции знаний становится менее однозначной.

СПРАВНИТЕЛЬНЫЙ АНАЛИЗ

Описанные выше подходы имеют расхождение также и с точки зрения постановки экспериментальной части. Так, в первую очередь, для тестирования эффективности реализаций могли использоваться разные бенчмарки. В результате сравнительный анализ было решено проводить относительно датасета,

который встречался наиболее часто в рассмотренных работах – CommonsenseQA [67].

CommonsenseQA состоит из 12102 вопросов, предлагающих 5 вариантов ответов, один из которых является верным. Выбор в пользу этого набора данных может обосновываться его более высокой сложностью с точки зрения относительно невысоких результатов вопросно-ответных систем на нем по сравнению с аналогами. Эта сложность в данном случае обусловлена наличием акцента в вопросах на социальные и психологические аспекты и необходимостью установления причинно-следственных связей, а также отсутствием у вопросов какого-либо дополнительного контекста. Это, с одной стороны, усложняет эффективную реализацию предобученных языковых моделей в связи с меньшим числом входных данных. С другой же стороны, подобная постановка задачи как раз благоприятствует формированию такого контекста за счет внешних баз знаний.

В табл. 2 представлены сводные результаты моделей без использования ансамблирования (Ensemble) по тестовой выборке датасета CommonsenseQA. В контексте данного датасета для практического сравнения реализаций в качестве метрики принято использовать точность (Accuracy) – процент вопросов, на которые был получен правильный ответ. Также необходимо указать, что в качестве базового бенчмарка была выбрана одна из наиболее часто используемых языковых моделей-энкодеров – RoBERTa [68].

Приведенные результаты показывают, что любой из рассмотренных подходов позволяет увеличить точность вопросно-ответной системы относительно базового решения с помощью предобученной языковой модели, что подтверждает перспективность данного исследовательского направления. При этом использующие графовые эмбединги модели демонстрируют заметно более низкую точность, а лучший результат на датасете CommonsenseQA получен при использовании модели KEAR с привлечением сведений из базы знаний через текстовые эмбединги.

Тем не менее, с практической точки зрения необходимо принимать во внимание и существование других важных факторов при сравнении подходов. К примеру, основанные на самообучении и дообучении модели, несмотря на более низкую точность, требуют меньше дополнительных вычислений для получения ответа на запрос. В то же время сам процесс предобучения таких моделей подразумевает достаточно существенные затраты вычислительных ресурсов. Помимо этого, не во всех реализациях используются одинаковые языковые модели, что само по себе может давать разницу в итоговой точности. При этом можно учитывать и то, сколько времени в совокупности требуется модели для получения ответа.

Таблица 2. Сравнение эффективности методов интеграции знаний

Модель	Метод интеграции	Точность на тестовой выборке CommonsenseQA, %
RoBERTa [68] (2019)	—	68.7
Модель из [15] (2020)	Самообучение	75.6
Модель из [23] (2022)	Самообучение	78.5
UnifiedQA [37] (2020)	Дообучение	79.1
Модель из [44] (2023)	Текстовые эмбединги и механизм внимания	75.0
Модель из [47] (2020)	Текстовые эмбединги и механизм внимания	80.3
DEKCOR [41] (2021)	Текстовые эмбединги и механизм внимания	80.7
KEAR [42] (2022)	Текстовые эмбединги и механизм внимания	86.1
JointLK [55] (2022)	Графовые эмбединги и механизм внимания	74.4
Модель из [53] (2020)	Графовые эмбединги и механизм внимания	75.3
MHGRN [54] (2020)	Графовые эмбединги и механизм внимания	75.4
QA-GNN [65] (2021)	Токены взаимодействия	73.4
GreaseLM [63] (2022)	Токены взаимодействия	74.2
DRAGON [64] (2022)	Токены взаимодействия	76.0

Если же исходить исключительно из результатов на бенчмарке CommonsenseQA, то можно констатировать, что использование в целом более сложных с точки зрения архитектуры моделей не приносит достаточно существенного эффекта, чтобы конкурировать с более устоявшимися подходами, ориентированными исключительно на использование языковых моделей. Несмотря на это, необходимо отметить и целесообразность продолжения сравнительного анализа с применением других бенчмарков для более полной оценки реального положения вещей.

ЗАКЛЮЧЕНИЕ

Представленный обзор позволяет утверждать об эффективности использования методов интеграции знаний в области разработки вопросно-ответных систем. Уже существующие решения экспериментально подтверждают возможность одновременного достижения нескольких основных целей интеграции знаний в данном контексте.

При этом все еще существует значительное пространство для дальнейшего совершенствования множества аспектов данного процесса. Во-первых, в настоящее время преобладают относительно базовые

и устоявшиеся в области обработки естественного языка методы извлечения данных из баз знаний по запросу. Только в нескольких работах предлагаются способы улучшения этого процесса, такие как перефразирование дополнительных знаний из базы для упрощения их обработки системой. В связи с этим, учитывая потенциальную важность извлечения релевантных сведений с точки зрения дальнейшей имплементации, представляется возможным также отдельно рассмотреть более специфические подходы и их влияние на результат.

Во-вторых, интерес представляет анализ потенциального влияния выбора определенной графовой модели для обработки структурированной информации, т.к. в существующих работах основной акцент смещен, прежде всего, на сравнение по критерию использованной языковой модели. В то же время за последние несколько лет появилось множество новых перспективных моделей векторных представлений графов знаний и графовых нейронных сетей, чьи возможности в рамках практических задач такого рода на данный момент не установлены, но могут существенно влиять на результаты работы системы в целом.

В-третьих, к настоящему моменту отсутствуют систематические исследования, касающиеся сопоставления методов объединения данных разных модальностей в контексте разработки вопросно-ответных систем. Этот вопрос также может считаться актуальным в силу возможности обобщения на более широкий спектр задач.

Наконец, в рамках текущего вектора развития области разработки вопросно-ответных систем, ведущего к преобладанию в приложениях универсальных генеративных языковых моделей-декодеров, таких как ChatGPT, имеет смысл сделать акцент на изучении особенностей способов инъекции знаний в модели такого типа.

СПИСОК ЛИТЕРАТУРЫ / REFERENCES

1. Devlin J., Chang M.-W., Lee K., Toutanova K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In: *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. 2019;1:4171–4186. <https://doi.org/10.18653/v1/N19-1423>
2. Petroni F., Rocktäschel T., Lewis P., et al. Language Models as Knowledge Bases? *Processing (EMNLP-IJCNLP)*. 2019. P. 2463–2473. <https://doi.org/10.18653/v1/D19-1250>
3. Sap M., Le Bras R., Allaway E., et al. ATOMIC: An Atlas of Machine Commonsense for If-Then Reasoning. In: *Proceedings of the AAAI Conference on Artificial Intelligence*. 2019;33(1):3027–3035. <https://doi.org/10.1609/aaai.v33i01.33013027>
4. Niven T., Kao H.-Y. Probing Neural Network Comprehension of Natural Language Arguments. *arXiv preprint arXiv:1907.07355*. 2019. <https://doi.org/10.48550/arXiv.1907.07355>
5. McCoy R. T., Pavlick E., Linzen T. Right for the Wrong Reasons: Diagnosing Syntactic Heuristics in Natural Language Inference. In: *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*. 2019. P. 3428–3448. <http://doi.org/10.18653/v1/P19-1334>
6. Li J., Chen J., Ren R., et al. The Dawn After the Dark: An Empirical Study on Factuality Hallucination in Large Language Models. *arXiv preprint arXiv:2401.03205*. 2024. <https://doi.org/10.48550/arXiv.2401.03205>
7. Wei J., Wang X., Schuurmans D., et al. Chain-of-Thought Prompting Elicits Reasoning in Large Language Models. In: *36th Conference on Neural Information Processing Systems*. 2022;35:24824–24837. <https://doi.org/10.48550/arXiv.2201.11903>
8. Lewis P., Perez E. Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks. *Advances in Neural Information Processing Systems*. 2020;33:9459–9474. <https://doi.org/10.48550/arXiv.2005.11401>
9. Ye Zhi-Xiu, Chen Q., Wang W., Ling Zhen-Hua. Align, Mask and Select: A Simple Method for Incorporating Commonsense Knowledge into Language Representation Models. *arXiv preprint arXiv:1908.06725v5*. 2020. <https://doi.org/10.48550/arXiv.1908.06725>
10. Vaswani A., Shazeer N., Parmar N., et al. Attention Is All You Need. *Advances in Neural Information Processing Systems* 30. 2018. <https://doi.org/10.48550/arXiv.1706.03762>
11. Liu J., Shen D., Zhang Y., et al. What Makes Good In-Context Examples for GPT-3? *arXiv preprint arXiv:2101.06804*. 2021. <https://doi.org/10.48550/arXiv.2101.06804>
12. Gao T., Fisch A., Chen D. Making Pre-trained Language Models Better Few-shot Learners. In: *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*. 2021. P. 3816–3830. <http://doi.org/10.18653/v1/2021.acl-long.295>
13. Shwartz V., West P., Le Bras R., et al. Unsupervised Commonsense Question Answering with Self-Talk. In: *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. 2020. P. 4615–4629. <http://doi.org/10.18653/v1/2020.emnlp-main.373>
14. Wang J., Zhao H. ArT: All-round Thinker for Unsupervised Commonsense Question-Answering. In: *Proceedings of the 29th International Conference on Computational Linguistics*. 2022. P. 1490–1501. <https://doi.org/10.48550/arXiv.2112.13428>
15. Wang P., Peng N., Ilievski F., et al. Connecting the Dots: A Knowledgeable Path Generator for Commonsense Question Answering. *arXiv preprint arXiv:2005.00691*. 2020. <https://doi.org/10.48550/arXiv.2005.00691>
16. Raffel C., Shazeer N., Roberts A., et al. Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer. *Journal of Machine Learning Research*. 2020;21(140):1–67. <https://doi.org/10.48550/arXiv.1910.10683>
17. Zhang Z., Han X., Liu Z., et al. ERNIE: Enhanced Language Representation with Informative Entities. In: *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*. 2019. P. 1441–1451. <https://doi.org/10.18653/v1/P19-1139>
18. Peters M.E., Neumann M., Logan IV R.L., et al. Knowledge enhanced contextual word representations. In: *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*. 2019. P. 43–54. <https://doi.org/10.18653/V1/D19-1005>
19. He L., Zheng S., Yang T., Zhang F. KLMo: Knowledge Graph Enhanced Pretrained Language Model with Fine-Grained Relationships. In: *Findings of the Association for Computational Linguistics: EMNLP*. 2021. P. 4536–4542. <https://doi.org/10.18653/v1/2021.findings-emnlp.384>
20. Xiong W., Du J., Wang W.Y., Stoyanov V. Pretrained Encyclopedia: Weakly Supervised Knowledge-Pretrained Language Model. *arXiv preprint arXiv:1912.09637*. 2019. <https://doi.org/10.48550/arXiv.1912.09637>

21. Sun Y., Wang S., Li Y., et al. ERNIE: Enhanced Representation through Knowledge Integration. *arXiv preprint arXiv:1904.09223*. 2019. <https://doi.org/10.48550/arXiv.1904.09223>
22. Zhang D., Yuan Z., Liu Y., et al. E-BERT: A Phrase and Product Knowledge Enhanced Language Model for E-commerce. *arXiv preprint arXiv:2009.02835*. 2020. <https://doi.org/10.48550/arXiv.2009.02835>
23. Chen Q., Li F.-L., Xu G., et al. DictBERT: Dictionary Description Knowledge Enhanced Language Model Pre-training via Contrastive Learning. *arXiv preprint arXiv:2208.00635*. 2022. <https://doi.org/10.48550/arXiv.2208.00635>
24. Lauscher A., Vulić I., Ponti E.M., et al. Informing Unsupervised Pretraining with External Linguistic Knowledge. *arXiv preprint arXiv:1909.02339v1*. 2019. <https://doi.org/10.48550/arXiv.1909.02339>
25. Levine Y., Lenz B., Dagan O., et al. SenseBERT: Driving Some Sense into BERT. In: *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*. 2020. P. 4656–4667. <https://doi.org/10.18653/v1/2020.acl-main.423>
26. Wang X., Gao T., Zhu Z., et al. KEPLER: A Unified Model for Knowledge Embedding and Pre-trained Language Representation. *Trans. Assoc. Comput. Linguis.* 2021;9:176–194. https://doi.org/10.1162/tac1_a_00360
27. Bordes A., Usunier N., Garcia-Durán A., et al. Translating Embeddings for Modeling Multi-relational Data. *Advances in Neural Information Processing Systems*. 2013. P. 2787–2795.
28. He B., Zhou D., Xiao J., et al. BERT-MK: Integrating Graph Contextualized Knowledge into Pre-trained Language Models. *Findings of the Association for Computational Linguistics: EMNLP*. 2020. P. 2281–2290. <https://doi.org/10.18653/v1/2020.findings-emnlp.207>
29. Banerjee P., Baral C. Self-Supervised Knowledge Triplet Learning for Zero-shot Question Answering. In: *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. 2020. P. 151–162. <https://doi.org/10.18653/v1/2020.emnlp-main.11>
30. Zhong W., Tang D., Duan N., et al. Improving Question Answering by Commonsense-Based Pre-training. In: Tang J., Kan M.Y., Zhao D., Li S., Zan H. (Eds.). *Natural Language Processing and Chinese Computing. NLPCC 2019. Lecture Notes in Computer Science*. Springer; 2019. V. 11838. P. 16–28. https://doi.org/10.1007/978-3-030-32233-5_2
31. Sun T., Shao Y., Qiu X., et al. CoLAKE: Contextualized Language and Knowledge Embedding. *arXiv preprint arXiv:2010.00309v1*. 2020. <https://doi.org/10.48550/arXiv.2010.00309>
32. Su Y., Han X., Zhang Z., et al. CokeBERT: Contextual knowledge selection and embedding towards enhanced pre-trained language models. *AI Open*. 2021;2:127–134. <https://doi.org/10.1016/j.aiopen.2021.06.004>
33. Ma K., Ilievski F., Francis J., et al. Knowledge-driven Data Construction for Zero-shot Evaluation in Commonsense Question Answering. In: *Proceedings of the AAAI Conference on Artificial Intelligence*. 2021;35(15):13507–13515. <https://doi.org/10.1609/aaai.v35i15.17593>
34. Wang W., Fang T., Ding W., et al. CAR: Conceptualization-Augmented Reasoner for Zero-Shot Commonsense Question Answering. *Findings of the Association for Computational Linguistics: EMNLP 2023*. 2023. P. 13520–13545. <https://doi.org/10.18653/v1/2023.findings-emnlp.902>
35. Zhan X., Li Y., Dong X., et al. eBERTo: Self-supervised Commonsense Learning for Question Answering. *arXiv preprint arXiv:2203.09424v1*. 2022. <https://doi.org/10.48550/arXiv.2203.09424>
36. Rajpurkar P., Jia R., Liang P. Know What You Don't Know: Unanswerable Questions for SQuAD. In: *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics*. 2018;2(Short Papers):784–789. <https://doi.org/10.18653/v1/P18-2124>
37. Khashabi D., Min S., Khot T., et al. UnifiedQA: Crossing Format Boundaries with a Single QA System. In: *Findings of the Association for Computational Linguistics*. 2020. P. 1896–1907. <https://doi.org/10.18653/v1/2020.findings-emnlp.171>
38. Lourie N., Le Bras R., Bhagavatula C., Choi Y. UNICORN on RAINBOW: A Universal Commonsense Reasoning Model on a New Multitask Benchmark. *arXiv preprint arXiv:2103.13009v1*. 2021. <https://doi.org/10.48550/arXiv.2103.13009>
39. Baek J., Aji A.F., Saffari A. Knowledge-Augmented Language Model Prompting for Zero-Shot Knowledge Graph Question Answering. In: *Proceedings of the 1st Workshop on Natural Language Reasoning and Structured Explanations (NLRSE)*. 2023. P. 78–106. <https://doi.org/10.18653/v1/2023.nlrse-1.7>
40. Pan X., Sun K., Yu D., et al. Improving Question Answering with External Knowledge. In: *Proceedings of the 2nd Workshop on Machine Reading for Question Answering*. 2019. P. 27–37. <https://doi.org/10.18653/v1/D19-5804>
41. Xu Y., Zhu C., Xu R., et al. Fusing Context Into Knowledge Graph for Commonsense Question Answering. *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*. 2021. P. 1201–1207. <https://doi.org/10.18653/v1/2021.findings-acl.102>
42. Xu Y., Zhu C., Wang S., et al. Human Parity on CommonsenseQA: Augmenting Self-Attention with External Attention. In: *Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence (IJCAI)*. 2022. P. 2762–2768. <https://doi.org/10.24963/ijcai.2022/383>
43. Arora S., Wu S., Liu E., Ré C. Metadata Shaping: A Simple Approach for Knowledge-Enhanced Language Models. In: *Findings of the Association for Computational Linguistics: ACL 2022*. 2022. P. 1733–1745. <https://doi.org/10.18653/v1/2022.findings-acl.137>
44. Li S., Gao Y., Jiang H., et al. Graph Reasoning for Question Answering with Triplet Retrieval. In: *Findings of the Association for Computational Linguistics: ACL 2023*. 2023. P. 3366–3375. <https://doi.org/10.18653/v1/2023.findings-acl.208>
45. Mitra A., Banerjee P., Pal K.K., et al. How Additional Knowledge can Improve Natural Language Commonsense Question Answering? *arXiv preprint arXiv:1909.08855v3*. 2020. <https://doi.org/10.48550/arXiv.1909.08855>
46. Chen Q., Zhu X., Ling Z.-H., et al. Neural Natural Language Inference Models Enhanced with External Knowledge. In: *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics 2018;1(Long Papers)*:2406–2417. <https://doi.org/10.18653/v1/P18-1224>

47. Chen Q., Ji F., Chen H., Zhang Y. Improving Commonsense Question Answering by Graph-based Iterative Retrieval over Multiple Knowledge Sources. In: *Proceedings of the 28th International Conference on Computational Linguistics*. 2020. P. 2583–2594. <https://doi.org/10.18653/v1/2020.coling-main.232>
48. Ma K., Francis J., Lu Q., et al. Towards Generalizable Neuro-Symbolic Systems for Commonsense Question Answering. In: *Proceedings of the First Workshop on Commonsense Inference in Natural Language Processing*. 2019. P. 22–32. <https://doi.org/10.18653/v1/D19-6003>
49. Bauer L., Wang Y., Bansal M. Commonsense for Generative Multi-Hop Question Answering Tasks. In: *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*. 2018. P. 4220–4230. <https://doi.org/10.18653/v1/D18-1454>
50. Paul D., Frank A. Ranking and Selecting Multi-Hop Knowledge Paths to Better Predict Human Needs. In: *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. 2019;1(Long and Short Papers):3671–3681. <https://doi.org/10.18653/v1/N19-1368>
51. Liu W., Zhou P., Zhao Z., et al. K-BERT: Enabling Language Representation with Knowledge Graph. In: *Proceedings of the AAAI Conference on Artificial Intelligence*. 2020;34(03):2901–2908. <https://doi.org/10.1609/aaai.v34i03.5681>
52. Kipf T.N., Welling M. Semi-Supervised Classification with Graph Convolutional Networks. *arXiv preprint arXiv:1609.02907*. 2017. <https://doi.org/10.48550/arXiv.1609.02907>
53. Lv S., Guo D., Xu J., et al. Graph-Based Reasoning over Heterogeneous External Knowledge for Commonsense Question Answering. In: *Proceedings of the AAAI Conference on Artificial Intelligence*. 2020;34(05):8449–8456. <https://doi.org/10.1609/aaai.v34i05.6364>
54. Feng Y., Chen Y., Lin B.Y., et al. Scalable Multi-Hop Relational Reasoning for Knowledge-Aware Question Answering. In: *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. 2020. P. 1295–1309. <https://doi.org/10.18653/v1/2020.emnlp-main.99>
55. Sun Y., Shi Q., Qi L., Zhang Y. JointLK: Joint Reasoning with Language Models and Knowledge Graphs for Commonsense Question Answering. In: *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. 2022. P. 5049–5060. <https://doi.org/10.18653/v1/2022.naacl-main.372>
56. Yan J., Raman M., Chan A., et al. Learning Contextualized Knowledge Structures for Commonsense Reasoning. In: *Findings of the Association for Computational Linguistics: ACL-IJCNLP*. 2021. P. 4038–4051. <https://doi.org/10.18653/v1/2021.findings-acl.354>
57. Lin B.Y., Chen X., Chen J., Ren X. KagNet: Knowledge-Aware Graph Networks for Commonsense Reasoning. In: *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*. 2019. P. 2829–2839. <https://doi.org/10.18653/v1/D19-1282>
58. Jiang J., Zhou K., Zhao W.X., Wen J.-R. Great Truths are Always Simple: A Rather Simple Knowledge Encoder for Enhancing the Commonsense Reasoning Capacity of Pre-Trained Models. In: *North American Chapter of the Association for Computational Linguistics-Findings*. 2022. <https://doi.org/10.48550/arXiv.2205.01841>
59. Houshy N., Giurigu A., Jastrzebski S., et al. Parameter-Efficient Transfer Learning for NLP. In: *Proceedings of Machine Learning Research*. 2019;97:2790–2799. <https://doi.org/10.48550/arXiv.1902.00751>
60. Wang R., Tang D., Duan N., et al. K-Adapter: Infusing Knowledge into Pre-Trained Models with Adapters. In: *Findings of the Association for Computational Linguistics: ACL-IJCNLP*. 2021. P. 1405–1418. <https://doi.org/10.18653/v1/2021.findings-acl.121>
61. Kim Y.J., Kwak B., Kim Y., et al. Modularized Transfer Learning with Multiple Knowledge Graphs for Zero-shot Commonsense Reasoning. In: *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. 2022. P. 2244–2257. <https://doi.org/10.18653/v1/2022.naacl-main.163>
62. Jacobs R., Jordan M., Nowlan S., Hinton G. Adaptive Mixtures of Local Experts. *Neural Computation*. 1991;3(1):79–87. <https://doi.org/10.1162/neco.1991.3.1.79>
63. Zhang X., Bosselut A., Yasunaga M., et al. GreaseLM: Graph REASONing Enhanced Language Models for Question Answering. In: *The International Conference on Learning Representations (ICLR)*. 2022. <https://doi.org/10.48550/arXiv.2201.08860>
64. Yasunaga M., Bosselut A., Ren H., et al. Deep Bidirectional Language-Knowledge Graph Pretraining. In: *36th Conference on Neural Information Processing Systems (NeurIPS)*. 2022. <https://doi.org/10.48550/arXiv.2210.09338>
65. Yasunaga M., Ren H., Bosselut A., et al. QA-GNN: Reasoning with Language Models and Knowledge Graphs for Question Answering. In: *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. 2021. P. 535–546. <https://doi.org/10.18653/v1/2021.naacl-main.45>
66. Su Y., Zhang J., Song Y., Zhang T. PipeNet: Question Answering with Semantic Pruning over Knowledge Graphs. *arXiv preprint arXiv:2401.17536v2*. 2024. <https://doi.org/10.48550/arXiv.2401.17536>
67. Talmor A., Herzig J., Lourie N., Berant J. CommonsenseQA: A Question Answering Challenge Targeting Commonsense Knowledge. In: *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. 2019;1(Long and Short Papers):4149–4158. <https://doi.org/10.18653/v1/N19-1421>

68. Liu Y., Ott M., Goyal N., et al. RoBERTa: A Robustly Optimized BERT Pretraining Approach. *arXiv preprint arXiv:1907.11692*. 2019. <https://doi.org/10.48550/arXiv.1907.11692>
69. Robertson S.E., Walker S. Some Simple Effective Approximations to the 2-Poisson Model for Probabilistic Weighted Retrieval. In: *SIGIR'94: Proceedings of the Seventeenth Annual International ACM-SIGIR Conference on Research and Development in Information Retrieval*. 1994. P. 232–241. https://doi.org/10.1007/978-1-4471-2099-5_24

Об авторе

Радюш Даниил Валентинович, аспирант, факультет программной инженерии и компьютерной техники, ФГАОУ ВО «Национальный исследовательский университет ИТМО» (197101, Россия, Санкт-Петербург, Кронверкский пр., д. 49, лит. А). E-mail: daniil.radyush@gmail.com. Scopus Author ID 58234958500, <https://orcid.org/0000-0001-8823-0609>

About the Author

Daniil V. Radyush, Postgraduate Student, Faculty of Software Engineering and Computer Systems, ITMO University (49-A, Kronverkskii pr., Saint Petersburg, 197101 Russia). E-mail: daniil.radyush@gmail.com. Scopus Author ID 58234958500, <https://orcid.org/0000-0001-8823-0609>

УДК 004.2

<https://doi.org/10.32362/2500-316X-2025-13-3-44-53>

EDN QWXGNC



НАУЧНАЯ СТАТЬЯ

Управление топологическими ограничениями при реализации конвейерных вычислительных структур на базе программируемых логических интегральных схем

И.Е. Тарасов[@],
П.Н. Советов,
Д.В. Люлява,
Н.А. Дуксин

МИРЭА – Российский технологический университет, Москва, 119454 Россия

[@] Автор для переписки, e-mail: tarasov_j@mirea.ru

• Поступила: 03.11.2024 • Доработана: 13.02.2025 • Принята к опубликованию: 20.03.2025

Резюме

Цели. Конвейеризация является эффективным приемом повышения тактовой частоты цифровых схем. При этом балансировка стадий конвейера при синтезе схемы на уровне регистровых передач еще не гарантирует сбалансированную по задержкам распространения сигнала топологическую реализацию такого конвейера в выбранном технологическом базисе. Это обусловлено спецификой алгоритмов размещения и трассировки компонентов цифровых устройств, которые не позволяют получать оптимальные решения в строгом математическом смысле за приемлемое время. В практике разработки цифровых устройств применяются подходы, основанные на комбинации ручного управления топологическими ограничениями, задающими общие правила размещения компонентов, и автоматической оптимизации для локализованных фрагментов схемы, которая в этом случае позволяет получать результаты, близкие к оптимальным. Конвейерные структуры имеют простую схему соединений отдельных стадий, что позволяет продемонстрировать на их примере эффект от применения топологических проектных ограничений. В то же время, на базе конвейерных структур возможна реализация ряда алгоритмов, эффективно дополняющих программируемые процессорные устройства и обеспечивающие аппаратное ускорение некоторых задач. Цель работы – разработка методических рекомендаций по управлению топологическими проектными ограничениями при реализации конвейерных вычислительных структур на базе программируемых логических интегральных схем (ПЛИС) с архитектурой field-programmable gate array (FPGA).

Методы. Использованы методы проектирования и моделирования цифровых систем.

Результаты. На основе проведенного анализа разработаны модификации конвейерного вычислителя 32-разрядного преобразования CORDIC для вычисления трансцендентных функций. Установлено, что добавление проектных ограничений по размещению групп регистров, соответствующих стадиям конвейера, позволяет существенно повысить тактовую частоту по сравнению с автоматическим размещением

и уменьшить время работы алгоритмов трассировки. Полученный эффект систематически воспроизводится в нескольких реализованных вариантах конвейера.

Выводы. Рассмотренные рекомендации позволяют управлять тактовой частотой и количеством стадий конвейерных вычислительных структур при одновременном уменьшении времени одной итерации размещения и трассировки модуля на базе ПЛИС.

Ключевые слова: ПЛИС, конвейер, проектные ограничения, CORDIC

Для цитирования: Тарасов И.Е., Советов П.Н., Люлява Д.В., Дуксин Н.А. Управление топологическими ограничениями при реализации конвейерных вычислительных структур на базе программируемых логических интегральных схем. *Russian Technological Journal*. 2025;13(3):44–53. <https://doi.org/10.32362/2500-316X-2025-13-3-44-53>, <https://www.elibrary.ru/QWXGNC>

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

Авторы заявляют об отсутствии конфликта интересов.

RESEARCH ARTICLE

Method for designing specialized computing systems on the basis of hardware and software cooptimization

Ilya E. Tarasov[@],
Peter N. Sovietov,
Daniil V. Lulyava,
Nikita A. Duksin

MIREA – Russian Technological University, Moscow, 119454 Russia

[@] Corresponding author, e-mail: tarasov_j@mirea.ru

• Submitted: 03.11.2024 • Revised: 13.02.2025 • Accepted: 20.03.2025

Abstract

Objectives. Pipelining is an effective method for increasing the clock frequency of digital circuits. At the same time, balancing the pipeline stages during circuit synthesis at the register transfer level does not yet guarantee a balanced topological implementation of such a pipeline in terms of signal propagation delays according to the selected technological basis. This is due to the specifics of the algorithms for placing and routing components of digital devices, which are not capable of optimizing solutions in a strict mathematical sense in an acceptable time. In practice, approaches for developing digital devices combine manual control of topological constraints that set general rules for placing components with automatic optimization for localized fragments of the circuit are used to obtain results close to optimal. Pipeline circuits are based on a simple connection diagram of individual stages to demonstrate the effect of using topological design constraints on their example. On the basis of pipeline structures, a number of algorithms can be implemented to effectively complement programmable processor devices and provide hardware acceleration of some tasks. The present work develops methodological recommendations for managing topological design constraints in the implementation of pipeline computing structures based on programmable logic devices (PLD) with field-programmable gate array (FPGA) architecture.

Methods. The work is based on accepted methods for designing and modeling digital systems.

Results. Based on the analysis, modifications to a 32-bit CORDIC transcendental function computation pipeline were developed. By adding design constraints on the placement of register groups corresponding to the pipeline stages a significant increase in the clock frequency can be achieved as compared to automatic placement to reduce the running time of the tracing algorithms. The resulting effect is systematically reproduced in several implemented versions of the pipeline.

Conclusions. The presented recommendations can be used to control the clock frequency and number of stages of pipeline computing structures while simultaneously reducing the time of one iteration and routing of a module based on PLD with FPGA architecture.

Keywords: PLD, pipeline, constraints, CORDIC

For citation: Tarasov I.E., Sovietov P.N., Lulyava D.V., Duksin N.A. Method for designing specialized computing systems on the basis of hardware and software cooptimization. *Russian Technological Journal*. 2025;13(3):44–53. <https://doi.org/10.32362/2500-316X-2025-13-3-44-53>, <https://www.elibrary.ru/QWXGNC>

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

The authors declare no conflicts of interest.

ВВЕДЕНИЕ

Высокопроизводительные вычислительные системы разрабатываются как комбинация универсальных и специализированных подсистем. При этом на стадии архитектурного проектирования необходимо выделить задачи, подлежащие решению с помощью специализированных подсистем, добавляемых к процессорам общего назначения. Такие задачи должны быть, с одной стороны, массово востребованными, а с другой – недостаточно эффективно решаться с помощью центрального процессорного устройства или существенно замедлять его работу. В цифровой электронике на базе специализированных вычислителей часто реализуются алгоритмы цифровой обработки сигналов [1], вычисление хеш-функций в подсистемах защиты информации [2], ускорение искусственных нейросетей [3] и т.д. В данной работе рассмотрены подходы к проектированию конвейерного вычислителя для ряда примеров.

При разработке вычислительного устройства, преобразующего входной вектор $\vec{x}$ в выходной вектор $\vec{y}$, производится преобразование функции, заданной на входном языке высокого уровня, в последовательность действий на каждой стадии преобразования. Для этого используется синтезатор, разработанный в лаборатории специализированных вычислительных систем РТУ МИРЭА [4]. Выходом синтезатора является текст на языке описания аппаратуры, формирующий регистры на стадиях конвейера и узлы комбинационной логики между ними, выполняющие преобразования $f_1, f_2, f_3, \dots, f_n$. Аналогичный подход использован в ряде синтезаторов [5]¹, однако для разработанного программного продукта имеется возможность управления синтезом на основе обратной связи, формируемой с помощью анализа результатов размещения и трассировки компонентов. При этом минимальный период

сигнала тактовой частоты определится максимальной величиной задержки распространения сигнала между стадиями конвейера. Для того, чтобы синтезатор имел возможность равномерно распределить задержку распространения сигнала между стадиями конвейера, необходимо определить составляющие этой задержки для используемого топологического базиса.

Задержка распространения сигнала между регистрами программируемой логической интегральной схемы (ПЛИС) с архитектурой FPGA² определяется следующим образом³:

$$t = t_{\text{logic}} + t_{\text{route}}, \quad (1)$$

где t_{logic} – задержка, определяемая комбинационными элементами; t_{route} – задержка, определяемая трассировочными цепями ПЛИС.

Для достижения высокой тактовой частоты и равномерного распределения суммарной задержки сигнала между всеми стадиями конвейера синтезатору требуется оценивать составляющие, определяемые комбинационными элементами и определяемые трассировочными цепями. После распределения преобразований сигналов по узлам комбинационной логики получаемое RTL-представление⁴ конвейера передается в систему автоматизированного проектирования (САПР) ПЛИС, выполняющую размещение компонентов схемы и трассировку соединений. При этом неоптимальное размещение компонентов приведет к появлению дополнительной задержки распространения сигналов, которая нарушит равномерность распределения задержки по стадиям конвейера.

Стадии разработки конвейерного вычислительного устройства проиллюстрированы на рис. 1.

² Field programmable gate array – программируемая пользователем вентильная матрица.

³ <https://docs.xilinx.com/r/en-US/ug906-vivado-design-analysis/Timing-Analysis>. Дата обращения 10.10.2024. / Accessed October 10, 2024.

⁴ Register transfer level – уровень регистровых передач.

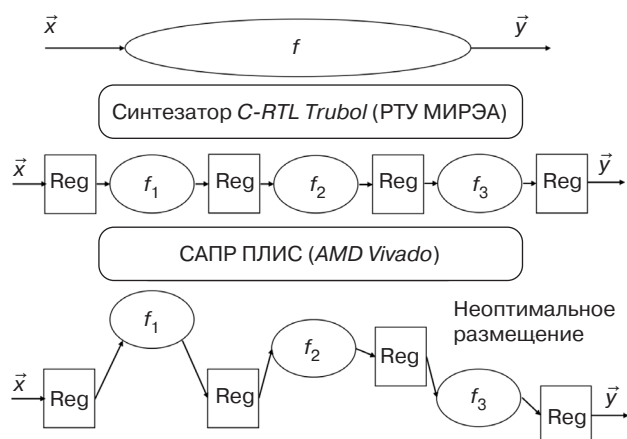


Рис. 1. Стадии разработки конвейерного вычислительного устройства.

Синтезатор *C-RTL Trubol* – программное обеспечение, разработанное коллективом авторов. Синтезатор является инструментом формирования описания в формате электронных САПР

Таким образом, для достижения высокой тактовой частоты работы конвейера необходимо обеспечить оценку составляющих задержек распространения сигнала и устранить негативные эффекты неоптимального взаимного размещения соединяемых компонентов на кристалле ПЛИС. Следует отметить, что задача оптимизации размещения в строгом математическом смысле, хотя и может быть сформулирована, не имеет практического решения за приемлемое время ввиду взрывного роста сложности алгоритмов оптимизации для общей формулировки. В практике проектирования на базе ПЛИС применяются проектные ограничения, устанавливающие области размещения групп компонентов – т.н. топологические проектные ограничения (*area constraints*). Для установленных таким образом групп компонентов (обозначаемых *P-blocks*) удастся выполнить оптимизацию алгоритмами САПР за приемлемое время с достижением субоптимальных результатов. Технические способы управления проектными ограничениями изложены в методическом руководстве *AMD UltraFast™ Design Methodology Guide for FPGAs and Systems-on-a-Chip*⁵. Исследования, связанные с использованием топологических проектных ограничений, отражены в ряде публикаций и датируются, начиная с 2011 г. [6], что связано с появлением таких инструментов проектирования, как *Xilinx PlanAhead* [7, 8]. В настоящее время применение проектных ограничений продолжает находить применение для обработки сетевых пакетов [9] и задач цифровой обработки сигналов [10].

⁵ <https://docs.amd.com/r/en-US/ug949-vivado-design-methodology>. Дата обращения 10.10.2024. / Accessed October 10, 2024.

УПРАВЛЕНИЕ ОПЕРАЦИЯМИ СИНТЕЗА ФУНКЦИОНАЛЬНЫХ УЗЛОВ КОНВЕЙЕРА

В работе рассмотрена следующая последовательность проектирования конвейерной вычислительной структуры. Для оценки задержки, определяемой комбинационной логикой, создается набор тестовых конвейерных цепочек, содержащих узлы с соответствующей логической функцией. Для этих узлов выполняется синтез и размещение конвейера, а экспериментально полученная оценка записывается в структуру данных, передаваемую синтезатору *C-RTL*. Поскольку синтезатор является оригинальной разработкой, в него были введены соответствующие модификации, позволяющие учитывать задержки, передаваемые из внешних источников.

Следует отметить, что при использовании ПЛИС не требуется оценивать широкий спектр возможных арифметических и логических операций. Поскольку поразрядные операции выполняются на базе таблиц истинности, а операции сложения и вычитания – с помощью специальных узлов «цепь ускоренного переноса» (*fast carry chain*), достаточно оценить задержку, определяемую этими двумя классами операций.

При наличии архитектурного шаблона и определенных параметров задержек может быть принята последовательность проектирования конвейерной вычислительной структуры, представленная на рис. 2.

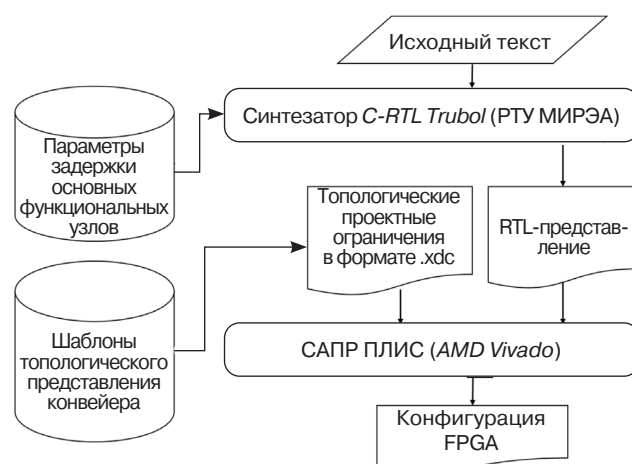


Рис. 2. Последовательность автоматизированного проектирования конвейерной вычислительной структуры

Для разработанной последовательности подразумевается, что на входе имеются исходные тексты программы на проблемно-ориентированном языке высокого уровня, а также проектные ограничения, сформированные на основе исследования характеристик аппаратной платформы. При синтезе используются параметры задержки основных

функциональных узлов, которые предварительно оцениваются в процессе синтеза конвейеров с заранее заданной схемой, явно выделяющей те или иные схемотехнические фрагменты, подлежащие оценке. Специализированный синтезатор формирует RTL-представление модуля на языке описания аппаратуры. Это представление дополняется файлом проектных ограничений в форматах .xdc или .sdc (в зависимости от используемой САПР), который создается путем параметризации одного из разработанных шаблонов топологического представления конвейера.

ФОРМИРОВАНИЕ ТОПОЛОГИЧЕСКИХ ПРОЕКТНЫХ ОГРАНИЧЕНИЙ ДЛЯ КОНВЕЙЕРНОЙ ВЫЧИСЛИТЕЛЬНОЙ СТРУКТУРЫ

При размещении конвейера в ПЛИС необходимо указать правила размещения его отдельных компонентов. Распространенной практикой является наличие в САПР ПЛИС режима иерархического (модульного) размещения, когда алгоритмы размещения используют модули проекта на уровне RTL в качестве локализованных единиц проекта, размещая их компоненты по возможности компактно. В то же время, разработчик имеет возможность сопроводить проект специальными файлами проектных ограничений (constraints), управляющих процессами группы Implementation – временным анализом и размещением. Соответствующие группы команд языка описания проектных ограничений затрагивают временные ограничения (timing constraints) и топологические ограничения (area constraints).

Команда для описания проектных ограничений в рамках одной стадии имеет вид:

```
create_pb <название pblock>
<координата pblock по оси X>
<координата pblock по оси Y>
<ширина pblock> <высота pblock>
<список элементов, связанных с pblock>
```

В этом примере для триггеров, имя которых соответствует шаблону, устанавливаются границы на кристалле ПЛИС с заданными координатами (рис. 3). Размеры границ на кристалле в точности, как и координаты, сопоставляются с общей топологией кристалла, количеством регистров и логических элементов, задействованных в каждой конкретной стадии. Шаблон имени подбирается, исходя из формата описания на RTL-уровне.

В процессе выполнения работы было подтверждено предположение о целесообразности описания топологических проектных ограничений только для регистров конвейера. Это обусловлено тем, что комбинационная логика между группами регистров,

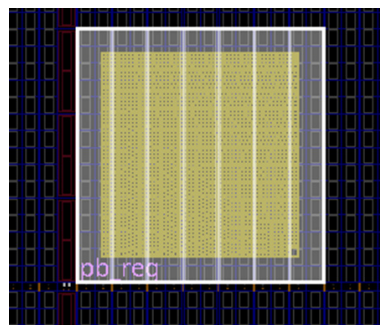


Рис. 3. Иллюстрация к операции выделения Р-блока в ПЛИС с архитектурой FPGA

относящихся к стадиям конвейера, имеет локализованные связи с этими группами. В этом случае установка регионов для расположения стадий конвейеров приведет к компактному размещению связанных с регистрами узлов комбинационной логики, оставляя при этом возможность для САПР выполнять локальные оптимизации. В этом случае ручное управление отдельными компонентами конвейера, включая отдельные триггеры и узлы комбинационной логики, чрезмерно трудоёмко.

Для разработанного в РТУ МИРЭА синтезатора схемы в формате RTL рекомендована модификация в виде генерации имен регистров с внедрением в них фрагментов, однозначно идентифицирующих стадию конвейера, которой принадлежит этот регистр. Данные сведения имеются во внутреннем представлении синтезатора, однако ранее не использовались. Анализ мировых аналогов показал, что идентификация стадии конвейера в них также невозможна, в экспортируемом RTL-представлении обычно используется сквозная нумерация отдельных триггеров схемы. При этом внедрение информации о стадии конвейера позволяет выделять соответствующую этой стадии группу регистров, используя регулярное выражение вида:

```
*/pipeline_unit/*reg_Ki\_*
```

ПРИМЕРЫ ПРАКТИЧЕСКОЙ АПРОБАЦИИ МЕТОДИКИ

Практическая апробация методики была выполнена путем реализации конвейеров нескольких видов. Например, алгоритм CORDIC⁶ [11] заключается в последовательном применении операции поворота вектора. Схожие действия в виде комбинации сложения и сдвига используются при последовательном умножении с накоплением (Multiply and Accumulate).

⁶ Метод CORDIC от англ. COordinate Rotation DIgital Computer – цифровой вычислитель поворота системы координат; метод «цифра за цифрой». [CORDIC is an acronym for COordinate Rotation DIgital Computer; a “digit by digit” method.]

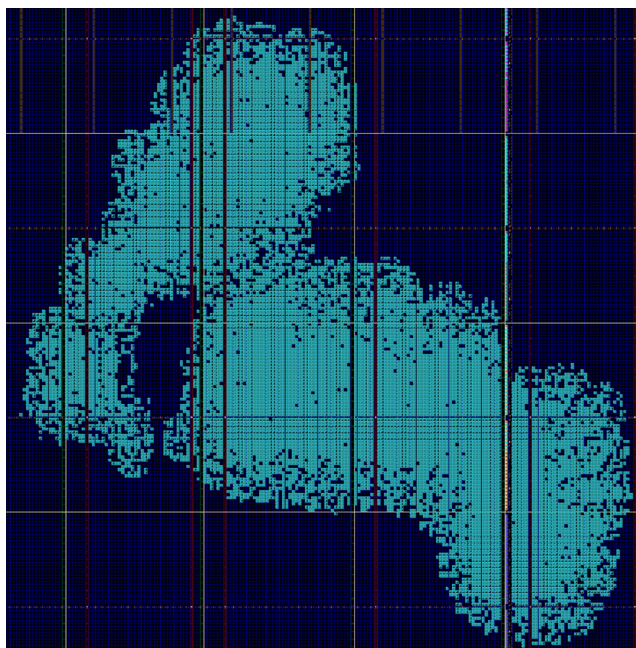


Рис. 4. Размещение стадий конвейера в автоматическом режиме в САПР Vivado

Данные операции можно совместить в одном конфигурируемом конвейере. Фрагмент схемы, сгенерированный в САПР Vivado⁷, приведен на рис. 4. Можно видеть, что автоматическое размещение компонентов конвейера с последующей оптимизацией привело к получению локально оптимального решения.

Поиск оптимального решения с точки зрения САПР не учитывает разбиение на стадии, собственные конвейерной архитектуре. Исходя из этого предположения, характеристики при размещении можно улучшить с применением описанного ранее подхода. В результате экспериментов был сделан вывод о повышении основных схемотехнических показателей схемы в случае, если границы блоков каждой стадии будут частично перекрывать границы соседних блоков. Результат, полученный с применением соответствующей стратегии, представлен на рис. 5.

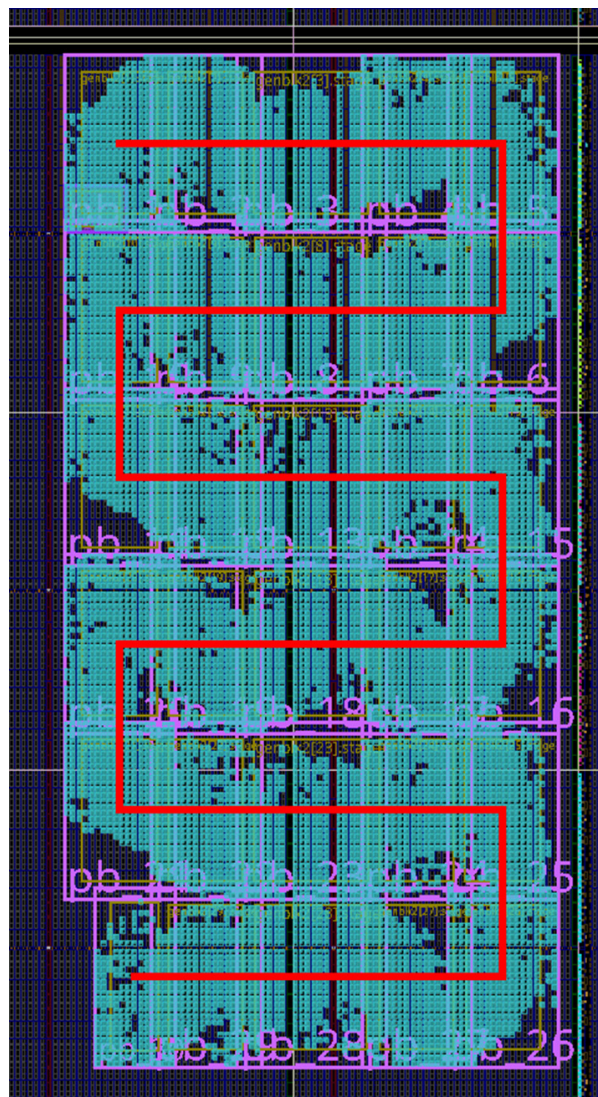


Рис. 5. Размещение стадий конвейера при использовании топологических проектных ограничений

Размещение проводилось в ПЛИС AMD FPGA Virtex™ UltraScale™+xcvu440_CIV-flga2892-3-e⁸ и, по сравнению со стандартным размещением (рис. 6), позволило добиться тактовой частоты в 1 ГГц (рис. 7).

Design Timing Summary

Setup	Hold	Pulse Width
Worst Negative Slack (WNS): -0.179 ns	Worst Hold Slack (WHS): 0.016 ns	Worst Pulse Width Slack (WPWS): -0.176 ns
Total Negative Slack (TNS): -45.105 ns	Total Hold Slack (THS): 0.000 ns	Total Pulse Width Negative Slack (TPWS): -0.176 ns
Number of Failing Endpoints: 970	Number of Failing Endpoints: 0	Number of Failing Endpoints: 1
Total Number of Endpoints: 96075	Total Number of Endpoints: 96075	Total Number of Endpoints: 94415

Рис. 6. Фрагмент отчета САПР с временными характеристиками проекта для схемотехнического решения, полученного в автоматическом режиме

⁷ <https://www.amd.com/en/products/software/adaptive-socs-and-fpgas/vivado.html>. Дата обращения 10.10.2024. / Accessed October 10, 2024.

⁸ <https://www.xilinx.com/products/boards-and-kits/1-66ql3z.html>. Дата обращения 10.10.2024. / Accessed October 10, 2024.

Design Timing Summary

Setup		Hold		Pulse Width	
Worst Negative Slack (WNS):	0.005 ns	Worst Hold Slack (WHS):	0.018 ns	Worst Pulse Width Slack (WPWS):	-0.176 ns
Total Negative Slack (TNS):	0.000 ns	Total Hold Slack (THS):	0.000 ns	Total Pulse Width Negative Slack (TPWS):	-0.176 ns
Number of Failing Endpoints:	0	Number of Failing Endpoints:	0	Number of Failing Endpoints:	1
Total Number of Endpoints:	142287	Total Number of Endpoints:	142287	Total Number of Endpoints:	140609

Рис. 7. Фрагмент отчета САПР с временными характеристиками проекта для схемотехнического решения при использовании топологических проектных ограничений

Еще более ярко заметен результат применения подхода при увеличении числа стадий алгоритма CORDIC. На рис. 8 и 9 представлены результаты размещения вычислителя для 64 стадий при тактовой частоте 600 МГц.

Такой же подход был применен для размещения рассматриваемой схемы на чипах AMD FPGA Artix-7 xc7a100tcs324-1⁹ (16 стадий, тактовая частота 400 МГц) и AMD FPGA Kintex™ UltraScale™ xcku115_CIV-flvf1924-3-e¹⁰ (32 стадии, тактовая частота 850 МГц). Работоспособность подхода доказывается, исходя из сравнения результатов временного анализа для рассматриваемых схем (рис. 10–13).

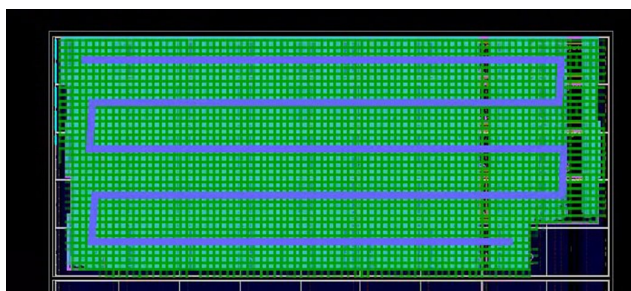


Рис. 8. Размещение стадий конвейера (64 стадии, тактовая частота 600 МГц)

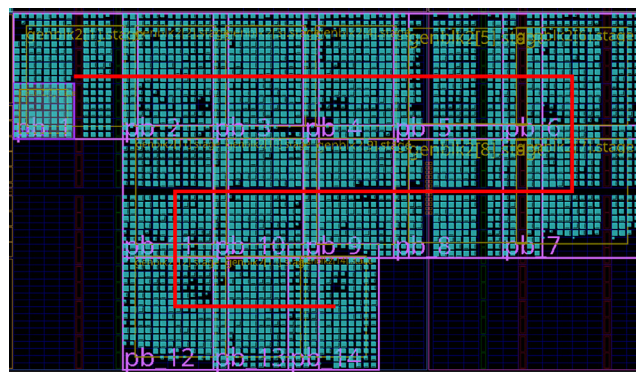


Рис. 10. Размещение стадий конвейера при использовании топологических проектных ограничений (ПЛИС xc7a100tcs324-1)

Design Timing Summary

Setup		Hold	
Worst Negative Slack (WNS):	0.024 ns	Worst Hold Slack (WHS):	0.010 ns
Total Negative Slack (TNS):	0.000 ns	Total Hold Slack (THS):	0.000 ns
Number of Failing Endpoints:	0	Number of Failing Endpoints:	0
Total Number of Endpoints:	16911	Total Number of Endpoints:	16911

All user specified timing constraints are met.

Рис. 11. Фрагмент отчета САПР с временными характеристиками проекта для полученного варианта размещения (ПЛИС xc7a100tcs324-1)

Design Timing Summary

Setup		Hold		Pulse Width	
Worst Negative Slack (WNS):	0.005 ns	Worst Hold Slack (WHS):	0.016 ns	Worst Pulse Width Slack (WPWS):	0.100 ns
Total Negative Slack (TNS):	0.000 ns	Total Hold Slack (THS):	0.000 ns	Total Pulse Width Negative Slack (TPWS):	0.000 ns
Number of Failing Endpoints:	0	Number of Failing Endpoints:	0	Number of Failing Endpoints:	0
Total Number of Endpoints:	780243	Total Number of Endpoints:	780243	Total Number of Endpoints:	775890

All user specified timing constraints are met.

Рис. 9. Фрагмент отчета САПР с временными характеристиками проекта конвейерного вычислителя CORDIC (64 стадии, тактовая частота 600 МГц)

⁹ <https://www.amd.com/en/products/adaptive-socs-and-fpgas/fpga/artix-7.html>. Дата обращения 10.10.2024. / Accessed October 10, 2024.

¹⁰ <https://www.amd.com/en/products/adaptive-socs-and-fpgas/fpga/kintex-ultrascale-plus.html>. Дата обращения 10.10.2024. / Accessed October 10, 2024.

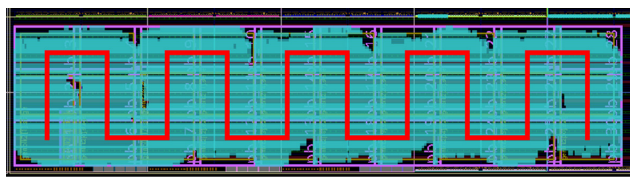


Рис. 12. Размещение стадий конвейера при использовании топологических проектных ограничений (ПЛИС xsku115_CIV-flvf1924-3-e)

Design Timing Summary

Setup	Hold
Worst Negative Slack (WNS): 0.040 ns	Worst Hold Slack (WHS): 0.030 ns
Total Negative Slack (TNS): 0.000 ns	Total Hold Slack (THS): 0.000 ns
Number of Failing Endpoints: 0	Number of Failing Endpoints: 0
Total Number of Endpoints: 142195	Total Number of Endpoints: 142195

Timing constraints are not met.

Рис. 13. Фрагмент отчета САПР с временными характеристиками проекта для полученного варианта размещения (ПЛИС xsku115_CIV-flvf1924-3-e)

Полученные результаты демонстрируют систематическое улучшение характеристик проектов вычислителей с конвейерной структурой при установке проектных ограничений для регистров отдельного модуля, описывающего конвейер. Продолжающийся интерес к конвейерным вычислительным узлам [12] позволяет рассматривать выбранное направление исследований как перспективное. Кроме этого, конвейерные устройства могут выступать в качестве подсистем вычислительных комплексов, увеличивая их эффективность как в широко распространенных задачах [13], так и в качестве ускорителей узкого подкласса вычислений [14, 15].

ЗАКЛЮЧЕНИЕ

Рассмотренные материалы отражают результаты, достигнутые лабораторией специализированных

вычислительных систем РТУ МИРЭА при разработке методики проектирования специализированных ускорителей вычислений с конвейерной архитектурой. Ориентация на простую аппаратную архитектуру с локализованными связями между узлами позволила разработать набор алгоритмов и проектных мероприятий, которые обеспечили систематический эффект улучшения характеристик топологического представления вычислительного устройства из его исходного описания на языке высокого уровня. Полученные результаты могут быть адаптированы к архитектурным шаблонам других типов с целью расширения номенклатуры специализированной электронной компонентной базы.

БЛАГОДАРНОСТИ

Работа выполнена в рамках государственного задания Министерства науки и высшего образования Российской Федерации (№ FSFZ-2022-0004 «Архитектуры специализированных вычислительных комплексов, методики, алгоритмы и инструменты проектирования цифровых вычислительных устройств»).

ACKNOWLEDGMENTS

The research is carried out within the framework of the State Assignment of the Ministry of Science and Higher Education of the Russian Federation (No. FSFZ-2022-0004 “Architectures of special-purpose computing units and procedures, algorithms, and tools for the design of digital computing units”).

Вклад авторов

Все авторы в равной степени внесли свой вклад в исследовательскую работу.

Authors' contribution

All authors equally contributed to the research work.

СПИСОК ЛИТЕРАТУРЫ

1. Saidov B.B., Telezhkin V.F., Gudaev N.N., et al. Development of Equipment for Experimental Study of Digital Algorithms in Nonstationary Signal Processing Problems. *Ural Radio Engineering Journal*. 2022;6(2):186–204. <https://doi.org/10.15826/urej.2022.6.2.004>
2. Jasek R. SHA-1 and MD5 Cryptographic Hash Functions: Security Overview. *Communications (Komunikacie)*. 2015;17(1):73–80.
3. Carrión D.S., Prohaska V., Diez O. Exploration of TPUs for AI Applications. In: Daimi K., Al Sadoon A. (Eds.). *Proceedings of the Second International Conference on Advances in Computing Research (ACR'24)*. ACR 2024. Lecture Notes in Networks and Systems. Springer; 2024. V. 956. P. 559. https://doi.org/10.1007/978-3-031-56950-0_47
4. Тарасов И.Е., Советов П.Н., Люлява Д.В., Мирзоян Д.И. Методика проектирования специализированных вычислительных систем на основе совместной оптимизации аппаратного и программного обеспечения. *Russian Technological Journal*. 2024;12(3):37–45 <https://doi.org/10.32362/2500-316X-2024-12-3-37-45>
5. Алехин В.А. Проектирование электронных систем с использованием SystemC и SystemC–AMS. *Russian Technological Journal*. 2020;8(4):79–95. <https://doi.org/10.32362/2500-316X-2020-8-4-79-95>

6. Pham-Quoc C., Dinh-Duc A.-V. Automatic generation of area constraints for FPGA implementation. In: *2011 IEEE 3rd International Conference on Communication Software and Networks (ICCSN)*. 2011. P. 469–472. <https://doi.org/10.1109/ICCSN.2011.6014937>
7. Li K., Lei L., Guang Q., Shi J.-Y., Hao Y. Improving the performance of an SOC design for network processing based on FPGA with PlanAhead. In: *2011 International Conference on Electronics, Communications and Control (ICECC)*. 2011. P. 297–300. <https://doi.org/10.1109/ICECC.2011.6066640>
8. Sarker A.L. Md, Lee M.H. Synthesis of VHDL code for FPGA design flow using Xilinx PlanAhead tool. In: *2012 International Conference on Education and e-Learning Innovations (ICEELI)*. 2012. <https://doi.org/10.1109/ICEELI.2012.6360614>
9. Song X., Lu R., Guo Z. High-Performance Reconfigurable Pipeline Implementation for FPGA-Based SmartNIC. *Micromachines*. 2024;15(4):449. <https://doi.org/10.3390/mi15040449>
10. Anderson T., Wheeler T.J. An FPGA-based hardware accelerator supporting sensitive sequence homology filtering with profile hidden Markov models. *BMC Bioinformatics*. 2024;25:247. <https://doi.org/10.1186/s12859-024-05879-3>
11. Тарасов И.Е., Советов П.Н. Устройство для вычисления трансцендентных функций и умножения двоичных чисел: пат. 222880 U1 РФ. Заявка № 2023131099; заявл. 28.11.2023; опубл. 22.01.2024. Бюл. № 3.
12. Oishi R., Kadomoto J., Irie H., Sakai S. FPGA-based Garbling Accelerator with Parallel Pipeline Processing. *IEICE Trans. Inform. Syst.* 2023;E106.D(12):1988–1996. <https://doi.org/10.1587/transinf.2023PAP0002>
13. Nurvitadhi E., Sheffield D., Sim J., et al. Accelerating Binarized Neural Networks: Comparison of FPGA, CPU, GPU, and ASIC. In: *2016 International Conference on Field-Programmable Technology (FPT)*. 2016. P. 77–84. <https://doi.org/10.1109/FPT.2016.7929192>
14. Hennessy J.L., Patterson D.A. A new golden age for computer architecture: Domain-specific hardware/software co-design, enhanced security, open instruction sets, and agile chip development. In: *Proceedings of the 2018 ACM/IEEE 45th Annual International Symposium on Computer Architecture (ISCA)*. 2018. P. 27–29. <https://doi.org/10.1109/ISCA.2018.00011>
15. Hennessy J.L., Patterson D.A. *Computer Architecture: A Quantitative Approach*: 6th ed. The Morgan Kaufmann Series in Computer Architecture and Design. 2017. 936 p.

REFERENCES

1. Saidov B.B., Telezhkin V.F., Gudaev N.N., et al. Development of Equipment for Experimental Study of Digital Algorithms in Nonstationary Signal Processing Problems. *Ural Radio Engineering Journal*. 2022;6(2):186–204. <https://doi.org/10.15826/urej.2022.6.2.004>
2. Jasek R. SHA-1 and MD5 Cryptographic Hash Functions: Security Overview. *Communications (Komunikacie)*. 2015;17(1):73–80.
3. Carrión D.S., Prohaska V., Diez O. Exploration of TPUs for AI Applications. In: Daimi K., Al Sadoon A. (Eds.). *Proceedings of the Second International Conference on Advances in Computing Research (ACR'24)*. ACR 2024. Lecture Notes in Networks and Systems. Springer; 2024. V. 956. P. 559. https://doi.org/10.1007/978-3-031-56950-0_47
4. Tarasov I.E., Sovietov P.N., Lulyava D.V., Mirzoyan D.I. Method for designing specialized computing systems based on hardware and software co-optimization. *Russian Technological Journal*. 2024;12(3):37–45. <https://doi.org/10.32362/2500-316X-2024-12-3-37-45>
5. Alekhin V.A. Designing Electronic Systems Using SystemC and SystemC–AMS. *Russian Technological Journal*. 2020;8(4):79–95 (in Russ.). <https://doi.org/10.32362/2500-316X-2020-8-4-79-95>
6. Pham-Quoc C., Dinh-Duc A.-V. Automatic generation of area constraints for FPGA implementation. In: *2011 IEEE 3rd International Conference on Communication Software and Networks (ICCSN)*. 2011. P. 469–472. <https://doi.org/10.1109/ICCSN.2011.6014937>
7. Li K., Lei L., Guang Q., Shi J.-Y., Hao Y. Improving the performance of an SOC design for network processing based on FPGA with PlanAhead. In: *2011 International Conference on Electronics, Communications and Control (ICECC)*. 2011. P. 297–300. <https://doi.org/10.1109/ICECC.2011.6066640>
8. Sarker A.L.Md., Lee M.H. Synthesis of VHDL code for FPGA design flow using Xilinx PlanAhead tool. In: *2012 International Conference on Education and e-Learning Innovations (ICEELI)*. 2012. <https://doi.org/10.1109/ICEELI.2012.6360614>
9. Song X., Lu R., Guo Z. High-Performance Reconfigurable Pipeline Implementation for FPGA-Based SmartNIC. *Micromachines*. 2024;15(4):449. <https://doi.org/10.3390/mi15040449>
10. Anderson T., Wheeler T.J. An FPGA-based hardware accelerator supporting sensitive sequence homology filtering with profile hidden Markov models. *BMC Bioinformatics*. 2024;25:247. <https://doi.org/10.1186/s12859-024-05879-3>
11. Tarasov I.E., Sovietov P.N. *Device for Calculating Transcendental Functions and Multiplying Binary Numbers*: Pat. 222880 U1 RF. Publ. 22.01.2024 (in Russ.).
12. Oishi R., Kadomoto J., Irie H., Sakai S. FPGA-based Garbling Accelerator with Parallel Pipeline Processing. *IEICE Trans. Inform. Syst.* 2023;E106.D(12):1988–1996. <https://doi.org/10.1587/transinf.2023PAP0002>
13. Nurvitadhi E., Sheffield D., Sim J., et al. Accelerating Binarized Neural Networks: Comparison of FPGA, CPU, GPU, and ASIC. In: *2016 International Conference on Field-Programmable Technology (FPT)*. 2016. P. 77–84. <https://doi.org/10.1109/FPT.2016.7929192>

14. Hennessy J.L., Patterson D.A. A new golden age for computer architecture: Domain-specific hardware/software co-design, enhanced security, open instruction sets, and agile chip development. In: *Proceedings of the 2018 ACM/IEEE 45th Annual International Symposium on Computer Architecture (ISCA)*. 2018. P. 27–29. <https://doi.org/10.1109/ISCA.2018.00011>
15. Hennessy J.L., Patterson D.A. *Computer Architecture: A Quantitative Approach*. 6th ed. The Morgan Kaufmann Series in Computer Architecture and Design. 2017. 936 p.

Об авторах

Тарасов Илья Евгеньевич, д.т.н., доцент, заведующий лабораторией специализированных вычислительных систем, ФГБОУ ВО «МИРЭА – Российский технологический университет» (119454, Россия, Москва, пр-т Вернадского, д. 78). E-mail: tarasov_i@mirea.ru. Scopus Author ID 57213354150, SPIN-код РИНЦ 4628-7514, <http://orcid.org/0000-0001-6456-4794>

Советов Петр Николаевич, к.т.н., старший научный сотрудник, лаборатория специализированных вычислительных систем, ФГБОУ ВО «МИРЭА – Российский технологический университет» (119454, Россия, Москва, пр-т Вернадского, д. 78). E-mail: peter.sovietov@gmail.com. Scopus Author ID 57221375427, SPIN-код РИНЦ 9999-1460, <http://orcid.org/0000-0002-1039-2429>

Люлява Даниил Вячеславович, младший научный сотрудник, лаборатория специализированных вычислительных систем, ФГБОУ ВО «МИРЭА – Российский технологический университет» (119454, Россия, Москва, пр-т Вернадского, д. 78). E-mail: lyulyava@mirea.ru. Scopus Author ID 58811698000, SPIN-код РИНЦ 1882-0989, <http://orcid.org/0009-0009-9623-7777>

Дуксин Никита Александрович, инженер, лаборатория специализированных вычислительных систем, ФГБОУ ВО «МИРЭА – Российский технологический университет» (119454, Россия, Москва, пр-т Вернадского, д. 78). E-mail: duksin@mirea.ru. SPIN-код РИНЦ 1082-8956, Scopus Author ID 58811361100, <https://orcid.org/0009-0009-0014-7065>

About the Authors

Ilya E. Tarasov, Dr. Sci. (Eng.), Associated Professor, Head of the Laboratory of Specialized Computing Systems, MIREA – Russian Technological University (78, Vernadskogo pr., Moscow, 119454 Russia). E-mail: tarasov_i@mirea.ru. Scopus Author ID 57213354150, RSCI SPIN-code 4628-7514, <http://orcid.org/0000-0001-6456-4794>

Peter N. Sovietov, Cand. Sci. (Eng.), Senior Researcher, Laboratory of Specialized Computing Systems, MIREA – Russian Technological University (78, Vernadskogo pr., Moscow, 119454 Russia). E-mail: peter.sovietov@gmail.com. Scopus Author ID 57221375427, RSCI SPIN-code 9999-1460, <http://orcid.org/0000-0002-1039-2429>

Daniil V. Lulyava, Junior Researcher, Laboratory of Specialized Computing Systems, MIREA – Russian Technological University (78, Vernadskogo pr., Moscow, 119454 Russia). E-mail: lyulyava@mirea.ru. Scopus Author ID 58811698000, RSCI SPIN-code 1882-0989, <http://orcid.org/0009-0009-9623-7777>

Nikita A. Duxin, Engineer, Laboratory of Specialized Computing Systems, MIREA – Russian Technological University (78, Vernadskogo pr., Moscow, 119454 Russia). E-mail: duksin@mirea.ru. RSCI SPIN-code 1082-8956, Scopus Author ID 58811361100, <https://orcid.org/0009-0009-0014-7065>

Роботизированные комплексы и системы. Технологии дистанционного зондирования
неразрушающего контроля

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

УДК 681.515

<https://doi.org/10.32362/2500-316X-2025-13-3-54-62>

EDN SHAEZM



НАУЧНАЯ СТАТЬЯ

Анализ и синтез интеллектуальных систем автоматического управления с нечетким регулятором I рода

Ю.А. Быковцев[@],
В.М. Лохин

МИРЭА – Российский технологический университет, Москва, 119454 Россия

[@] Автор для переписки, e-mail: bykovcev@mirea.ru

• Поступила: 20.11.2024 • Доработана: 13.02.2025 • Принята к опубликованию: 21.03.2025

Резюме

Цели. Активное развитие интеллектуальных систем автоматического управления, связанное с повышением требований к качеству и точности систем управления современных технических систем, требует разработки новых подходов к их анализу и синтезу. Одним из перспективных классов интеллектуальных управляющих устройств выступают регуляторы, построенные на базе технологии нечеткого логического вывода. Целью настоящей работы является разработка методики комплексного синтеза параметров нечеткого регулятора I рода на основе кругового критерия Якубовича.

Методы. В основу предлагаемой методики положено рассмотрение нечеткого регулятора с позиции соответствующего нелинейного преобразования, что позволяет использовать методы теории нелинейных систем автоматического управления. В качестве показателей качества в работе используются аналогии понятий «степень устойчивости» и «степень колебательности». Синтез параметров нелинейного преобразования сводится к определению достаточных областей абсолютной устойчивости системы со смещенной и расширенной амплитудно-фазовыми частотными характеристиками, полученных с помощью кругового критерия устойчивости Якубовича.

Результаты. В соответствии с теорией нечетких множеств и алгоритмом нечеткого логического вывода Такаги – Сугено показана возможность взаимно-однозначного соответствия нелинейного преобразования и параметров базы знаний нечеткого регулятора при соответствующей организации последней. В работе предложена процедура синтеза параметров нечеткого регулятора I рода, нацеленная на обеспечение комплексных требований к качеству системы управления по «степени устойчивости», «степени колебательности» и точности в установившемся режиме. Предложенная методика также гарантирует абсолютную устойчивость не только положения равновесия, но и процессов, а ее эффективность подтверждена результатами модельных экспериментов.

Выводы. В работе предложена удобная инженерная методика настройки параметров интеллектуального регулятора, построенная по технологии нечеткого логического вывода на основе методов теории автоматического управления. Показано удобство применения таких косвенных показателей качества, как «степень устойчивости», «степень колебательности» и точность в установившемся режиме.

Ключевые слова: интеллектуальная система управления, нечеткий логический вывод, нечеткий регулятор, модель Такаги – Сугено, абсолютная устойчивость процессов

Для цитирования: Быковцев Ю.А., Лохин В.М. Анализ и синтез интеллектуальных систем автоматического управления с нечетким регулятором I рода. *Russian Technological Journal*. 2025;13(3):54–62. <https://doi.org/10.32362/2500-316X-2025-13-3-54-62>, <https://www.elibrary.ru/SHAEZM>

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

Авторы заявляют об отсутствии конфликта интересов.

RESEARCH ARTICLE

Analysis and synthesis of intelligent automatic control systems with type-1 fuzzy regulator

Yuri A. Bykovtsev[@],
Valery M. Lokhin

MIREA – Russian Technological University, Moscow, 119454 Russia

[@] Corresponding author, e-mail: bykovtsev@mirea.ru

• Submitted: 20.11.2024 • Revised: 13.02.2025 • Accepted: 21.03.2025

Abstract

Objectives. The active development of intelligent automatic control systems, which is associated with increasing requirements to the quality and accuracy of control systems of modern technical systems, requires the development of new approaches to their analysis and synthesis. A promising class of intelligent control devices is based on regulators that use fuzzy-logic inference technology. The purpose of this work is to develop a method for the complex synthesis of type-1 fuzzy regulator parameters on the basis of the Yakubovich circle criterion.

Methods. The proposed methodology is based on a consideration of fuzzy regulators in terms of the corresponding nonlinear transformation that support the use of methods derived from the theory of nonlinear automatic control systems. Analogs of the degrees of stability and oscillation are used as quality indicators. The synthesis of the parameters of the nonlinear transformation can be reduced to determining sufficient regions of absolute stability of the system with the shifted and extended Nyquist plot obtained using the Yakubovich circle stability criterion.

Results. In accordance with the theory of fuzzy sets and algorithms of fuzzy logical inference described by Takagi–Sugeno, the possibility of one-to-one correspondence of the nonlinear transformation and the parameters of an appropriately arranged knowledge base of the fuzzy controller is shown. A procedure proposed for synthesizing the parameters of the type-1 fuzzy regulator is aimed at ensuring complex requirements for the quality of the control system according to the degree of stability, the degree of oscillation, and steady-state mode accuracy. The effectiveness of the proposed technique, which guarantees the absolute stability not only of the equilibrium position but also of the processes, is confirmed by the results of model experiments.

Conclusions. The paper proposes a convenient engineering technique for determining the parameters of an intelligent controller constructed using fuzzy logic inference technology based on methods informed by automatic control theory. The convenience of using such indirect quality indicators as the degree of stability, the degree of oscillation, and accuracy in steady-state mode, is demonstrated. These indicators are explicable for developers of applied control systems.

Keywords: intelligent control system, fuzzy logic inference, fuzzy controller, Takagi–Sugeno model, absolute stability of processes

For citation: Bykovtsev Y.A., Lokhin V.M. Analysis and synthesis of intelligent automatic control systems with type-1 fuzzy regulator. *Russian Technological Journal*. 2025;13(3):54–62. <https://doi.org/10.32362/2500-316X-2025-13-3-54-62>, <https://www.elibrary.ru/SHAEZM>

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

The authors declare no conflicts of interest.

ВВЕДЕНИЕ

В последние два–три десятилетия интеллектуальные технологии находят все более широкое применение в различных сферах деятельности. Одной из таких сфер являются системы автоматического управления (САУ) [1], для которых на базе интеллектуальных технологий (экспертных систем, нейроподобных сетей, ассоциативной памяти или нечеткой логики) создается новое поколение регуляторов – интеллектуальных регуляторов, обеспечивающих не только высокие качественные показатели САУ, но и требуемый уровень адаптации к различным факторам неопределенности, воздействующим на систему.

Среди отмеченных интеллектуальных технологий наибольшее распространение (в силу объективных и субъективных причин [2–4]) получила технология нечеткого логического вывода (НЛВ) или нечеткая логика. Она получила достаточно широкое распространение в робототехнике на всех уровнях иерархии интеллектуального управления (стратегического, тактического и исполнительного) автономными и полуавтоматическими роботами различного типа базирования, в системах управления сложным технологическим оборудованием и т.д. В значительной степени это связано с тем, что аппарат НЛВ позволяет строить модели управления даже для сложных объектов на уровне логико-лингвистических рассуждений.

Вместе с тем этот новый и оригинальный формализм не соответствует существующей теории автоматического управления (ТАУ). Возникла серьезная проблема, связанная с созданием новых подходов к решению задач оценки устойчивости и качества нового класса САУ, которая достаточно активно решается в последние два десятилетия. Серьезное обобщение работ проведено в монографии Пегата [5]. Весьма перспективной оказалась концепция, предложенная в начале XXI века в РТУ МИРЭА И.М. Макаровым с соавторами [1]. Обобщая многолетний опыт исследований в области нечетких САУ и, в частности, результаты исследований, приведенных в работах [1–6], можно констатировать следующее:

1. НЛВ позволяет синтезировать логико-лингвистические модели управления для сложных объектов.

2. Несмотря на кажущуюся сложность формализма НЛВ установлено и обосновано, что регуляторы, построенные по данной технологии – нечеткие регуляторы (НР) – являются по существу нелинейными, т.е. реализуют нелинейное преобразование, параметры которого могут незначительно меняться при изменении технологии НЛВ (Мамдани, Сугено и др.).
3. Характер нелинейного преобразования в НР однозначно определяет параметры входных логико-лингвистических переменных.

Представление НР как нелинейного элемента САУ открывает широкую перспективу для привлечения к исследованию интеллектуальных САУ традиционных методов исследования нелинейных систем, принятых в ТАУ и модифицированных с учетом специфики и характера нелинейных преобразований.

ОСОБЕННОСТИ АНАЛИЗА И СИНТЕЗА САУ С НР

Рассмотрим нечеткую логическую систему с одной входной (E) и одной выходной (U) лингвистическими переменными с областями рассуждения на $X_E \subseteq \mathbb{R}$ и $Y_U \subseteq \mathbb{R}$, для которых заданы соответствующие терм-множества T_E и T_U , причем каждое значение лингвистической переменной из базового терм-множества задается нормальным нечетким множеством $A_i^E = \{(\mu_A(e), e) | e \in X_E\}$ и $A_i^U = \{(\mu_A(u), u) | u \in X_U\}$. В настоящее время среди моделей НЛВ активно применяются модели Мамдани, Ларсена, Такаги – Сугено и Цукамото, имеющие свои достоинства и недостатки [7], а нечеткое преобразование в результате (независимо от типа этих моделей) может быть представлено как определенное нелинейное отображение $f: X_E \rightarrow Y_U$.

Несмотря на это, большинство разработчиков нечетких систем управления к настоящему времени сформулировали некоторые общие принципы построения НР, а именно:

1. Количество нечетких множеств в базовых терм-множествах: 5–7.
2. Терм-множество должно содержать, по крайней мере, по одному нечеткому множеству, задаваемому функциями принадлежности (ФП) классов L и γ для ограничения величины управления, что связано с физическими особенностями САУ.

3. Симметричность расположения ФП относительно центральной ФП для обеспечения симметричности процесса регулирования при отклонениях состояния системы от положения равновесия.

В зависимости от способа обработки входных переменных в работе [1] предложено разделение НР на I и II роды. В данной статье рассматриваются САУ с наиболее популярными НР I рода с моделью НЛВ по Такаги – Сугено как наиболее перспективной.

Во-первых, это связано с облегченной процедурой дефазификации, представляющей собой расчет средневзвешенного значения, что требует существенно меньше аппаратных ресурсов и процессорного времени управляющего контроллера, чем другие методы. Во-вторых, при соответствующей организации базы знаний реализуемое нечеткой моделью отображение будет являться кусочно-линейной функцией, что существенно упрощает как анализ, так и синтез нечетких САУ. Рассмотрим последнее обстоятельство подробнее.

На рис. 1 представлен фрагмент нечеткой системы, у которой на некотором интервале рассуждения определены два нечетких множества A_{i-1}^E и A_i^E , задаваемые ФП класса t тройками $\{a_{i-1}, b_{i-1}, c_{i-1}\}$ и $\{a_i, b_i, c_i\}$, соответственно.

Получим выражение для отображения f в диапазоне входных воздействий $[b_{i-1}, b_i]$. Пусть база правил содержит два продукционных правила, вида:

1. Если E есть A_{i-1}^E , то $U = u_{i-1}^*$,
2. Если E есть A_i^E , то $U = u_i^*$,

где $u_{i-1}^* \equiv \text{const}$, $u_i^* \equiv \text{const}$.

В этом случае на этапе фаззификации значения степени принадлежности лингвистической переменной E нечетким множествам A_{i-1}^E и A_i^E определяется как:

$$\mu_{i-1}(e) = \frac{c_{i-1} - e}{c_{i-1} - b_{i-1}}, \quad \mu_i(e) = \frac{e - a_i}{b_i - a_i}. \quad (1)$$

В соответствии с процедурой НЛВ Такаги – Сугено и принятыми ограничениями на значения параметров ФП выходная переменная определяется как:

$$\begin{aligned} u_i(e) &= \frac{\mu_{i-1} u_{i-1}^* + \mu_i u_i^*}{\mu_{i-1} + \mu_i} = \\ &= \left(\frac{b_i - e}{b_i - b_{i-1}} u_{i-1}^* + \frac{e - b_{i-1}}{b_i - b_{i-1}} u_i^* \right) \div \left(\frac{b_i - e}{b_i - b_{i-1}} + \frac{e - b_{i-1}}{b_i - b_{i-1}} \right) = \\ &= \frac{u_i^* - u_{i-1}^*}{b_i - b_{i-1}} e + \frac{b_i u_{i-1}^* - b_{i-1} u_i^*}{b_i - b_{i-1}}. \end{aligned} \quad (2)$$

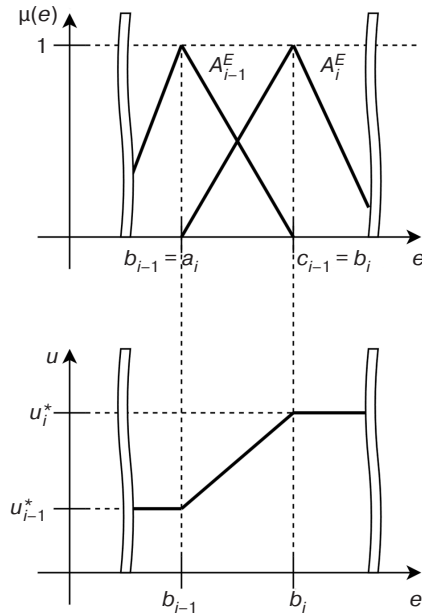


Рис. 1. Связь параметров ФП с видом нелинейного отображения

Таким образом, на интервале $[b_{i-1}, b_i]$ отображение f представляет собой линейную функцию, причем ее область определения задается расположением вершин соседних ФП, а область значения – величиной заключений базы правил.

Пользуясь результатами приведенного анализа, нечеткую САУ, построенную в соответствии с изложенными выше принципами, можно представить как нелинейную систему, у которой статическая характеристика НР имеет кусочно-линейную характеристику (рис. 2). Таким образом, создана возможность всестороннего анализа динамики нечеткой САУ с позиции теории нелинейных систем с учетом особенности нелинейного преобразования.

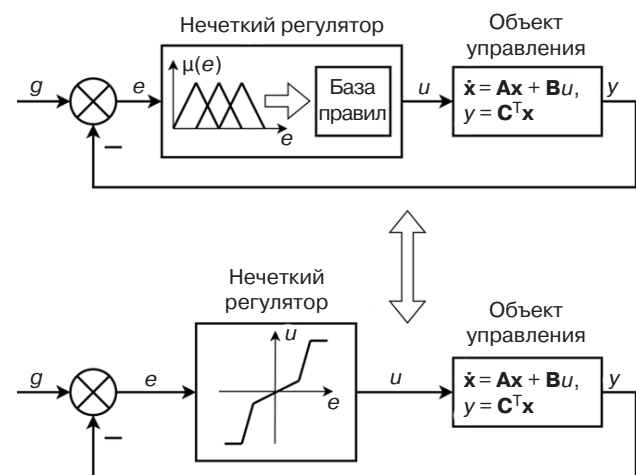


Рис. 2. Преобразование интеллектуальной САУ с НР к нелинейной САУ. g – задающее воздействие, u – управляющее воздействие, y – выход объекта управления, $\mathbf{x}$ – вектор состояния, $\mathbf{A}$ – матрица системы, $\mathbf{B}$ – матрица управления, $\mathbf{C}$ – матрица выхода

К настоящему времени развиваются различные подходы к анализу устойчивости систем с НР. Как отмечается в [5], несмотря на активное появление новых методов, наибольшую популярность приобрели второй метод Ляпунова [8, 9] и критерий абсолютной устойчивости [10, 11]. Для систем с одной входной и одной выходной величиной (Single Input Single Output, SISO) рекомендуется использовать критерий Попова и круговой критерий, в то время как для систем с множеством входов и выходов (Multiple Input Multiple Output, MIMO) наиболее подходящими являются методы, основанные на критерии гиперустойчивости, поскольку эти подходы обеспечивают строгую математическую базу для оценки устойчивости.

На фоне достаточно большого количества работ по устойчивости нечетких систем явно ощущается отсутствие удобных инженерных методик оценки качественных показателей таких систем, как быстродействие и перерегулирование. Такой подход к задаче более конструктивен, поскольку при достижении необходимого уровня качества САУ автоматически решается и вопрос устойчивости. Тем не менее, анализ существующих исследований показывает, что данная проблема недостаточно изучена и проработана.

Помимо аналитических методов исследования нечетких САУ, на данный момент широкую популярность получили подходы, основанные на численных оптимизационных алгоритмах, включая генетические алгоритмы [12, 13], методы, связанные с поведением роя частиц [13, 14], градиентный спуск [15, 16] и прочие. Хотя эти методы отличаются высокой производительностью и способны в автоматическом режиме определять параметры НР, их применение в практических инженерных целях осложняется рядом факторов. Прежде всего, для эффективного использования этих методик необходимо точно определить функцию качества, что представляет собой сложную задачу. Кроме того, алгоритмы не предоставляют инструкции по дополнительной настройке параметров НР после их вычисления, исходя из заданных критериев качества.

Приведенный анализ показывает, что с учетом очевидной перспективы нечетких систем и актуальности внедрения НР в широкий спектр систем управления промышленного и специального назначения, назрела необходимость создания новых подходов к исследованию динамики таких систем, основанных на традиционных методах, принятых в ТАУ. Удобная платформа для этого создается на основе предложенной концепции, при которой нечеткая САУ может рассматриваться как нелинейная.

В качестве исходных показателей качества удобно взять аналоги известных показателей качества линейных САУ – «степени устойчивости» как показателя скорости затухания переходного процесса и «степени

колебательности» как показателя затухания колебаний. За счет рассмотрения НР с позиции ее статической характеристики удастся адаптировать известные методы анализа и синтеза нелинейных САУ и, в частности, круговой критерий устойчивости Якубовича. Как показано в работе [17], за счет применения данного критерия к «смещенной» амплитудно-фазовой частотной характеристике (АФЧХ) линейной части удастся определить секторные ограничения на нелинейное преобразование, обеспечивающие заданную «степень устойчивости» нечеткой САУ, а также гарантирующие абсолютную устойчивость положения равновесия и процессов. С учетом этого, крайне перспективным представляется расширение данного метода с целью учета требований по «степени колебательности» и точности нечеткой САУ.

АЛГОРИТМ НАСТРОЙКИ ПАРАМЕТРОВ НР I РОДА НА ОСНОВЕ КОСВЕННЫХ ПОКАЗАТЕЛЕЙ КАЧЕСТВА

Модель нечеткой системы по аналогии с описанием нелинейных систем, принятым в ТАУ, можно представить в виде:

$$\begin{aligned}\dot{\mathbf{x}} &= \mathbf{A}\mathbf{x} + \mathbf{B}u, \\ u &= f(y), \\ y &= \mathbf{C}^T \mathbf{x},\end{aligned}\tag{3}$$

где $\mathbf{x} \in \mathbb{R}^n$, $u \in \mathbb{R}^1$, $f(y)$ – скалярная функция, являющаяся статической характеристикой «вход-выход» НР и принадлежащая классу $(K_1; K_2)$, т.е. удовлетворяющая неравенству [2]:

$$K_1 \leq \frac{df(y)}{dy} \leq K_2.\tag{4}$$

Ставится задача синтеза соответствующей базы знаний НР, обеспечивающего абсолютную устойчивость процессов в нечеткой САУ и качественные показатели переходного процесса – «степень устойчивости», «степень колебательности» и точность.

Для дальнейшего исследования удобно воспользоваться «смещенной» $\bar{W}(j\omega - \eta)$ и «расширенной» $\hat{W}(j\omega - m\omega)$ АФЧХ линейной части, где $W(j\omega) = \mathbf{C}^T(j\omega\mathbf{I} - \mathbf{A})^{-1}\mathbf{B}$, η – аналог «степени устойчивости», m – аналог «степени колебательности». На рис. 3а и 3б представлено семейство «смещенных» ($\eta_1 > \eta_0 > 0$) и «расширенных» ($m_1 > m_0 > 0$) АФЧХ линейной части 3-го порядка, соответственно. Следует обратить внимание на характерное изменение «смещенной» АФЧХ линейной части при $\eta = \eta_0$ вследствие перехода одного из полюсов $W(s)$ в правую комплексную полуплоскость.

Пусть перед нечеткой системой управления стоит требование обеспечить быстродействие не более, чем

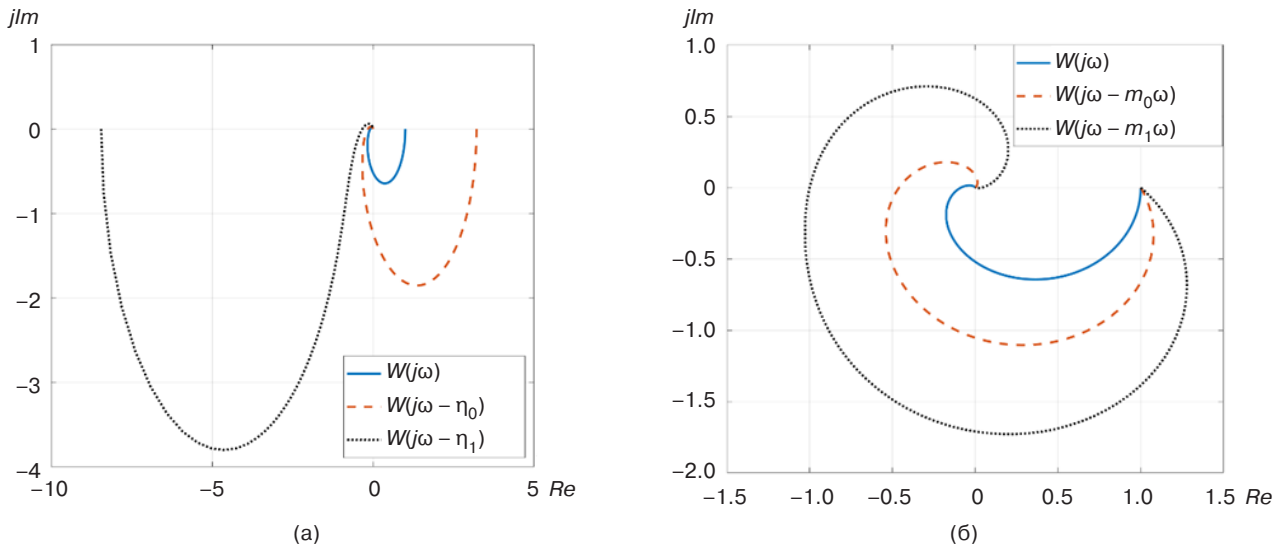


Рис. 3. Семейство «смещенных» (а) и «расширенных» (б) АФЧХ линейной части

время затухания экспоненты $e^{-n_0 t}$, причем n_0 выбрано таким, что степень неустойчивости (r) «смещенной» линейной части $\bar{W}(j\omega - \eta)$ равна единице.

Используя модификацию кругового критерия, предложенную в [17], можно сформулировать следующее достаточное условие: нечеткая система абсолютно устойчива в классе $(K_1^S; K_2^S)$ и имеет степень устойчивости не менее η , если «смещенная» АФЧХ линейной части $\bar{W}(j\omega - \eta)$ охватывает окружность с центром на вещественной оси в точке $-\frac{1}{2}\left(\frac{1}{K_1^S} + \frac{1}{K_2^S}\right)$, проходящей через точки $-\frac{1}{K_1^S}$ и $-\frac{1}{K_2^S}$, также принадлежащие действительной оси, r раз.

Для обеспечения заданной степени колебательности m достаточно, чтобы «расширенная» АФЧХ располагалась вне вышеобозначенной окружности. Причем, если требуемая «степень колебательности» обеспечивается в секторе $(K_1^O; K_2^O)$, то, учитывая результаты, полученные исходя из заданной «степени устойчивости», в секторе $\{\max(K_1^O; K_1^S); \min(K_2^O; K_2^S)\}$ будут обеспечиваться оба требования по качеству. Таким образом, рассматривая нечеткую САУ как нелинейную и применяя круговой критерий Якубовича, удастся определить параметры K_1 и K_2 нелинейного преобразования и соответствующие параметры настройки НР.

Наконец, остановимся на вопросе обеспечения требуемой точности нечеткой САУ. Поскольку ошибка регулирования системы в установившемся режиме определяется участком статической характеристики вблизи положения равновесия, то выбор коэффициента усиления K_1^a определяется из требования желаемой точности (ассигасу) во всем диапазоне возмущающих воздействий f . В работе [18]

показано, что для статических линейных частей ошибка в установившемся режиме в диапазоне возмущающих воздействий $f \leq \frac{b_1(1 + K_1^a K_{лч})}{K_{лч}}$ определяется как:

$$e_{уст} = \frac{f K_{лч}}{1 + K_{лч} K_1^a}, \quad (6)$$

где b_1 – правая граница участка статической характеристики с коэффициентом усиления K_1^a , $K_{лч} = \lim_{\omega \rightarrow 0} W(j\omega)$.

Из (6) нетрудно видеть, что, если известна максимальная величина возмущающего воздействия $f_M = \sup(f)$ и задано допустимое значение ошибки регулирования e_d , то требуемое значение коэффициента усиления K_1^a и граница участка b_1 определяются как:

$$K_1^a = \frac{f_M K_{лч} - e_d}{K_{лч} e_d}, \quad (7)$$

$$b_1 \geq \frac{f_M K_{лч}}{1 + K_{лч} K_1^a}. \quad (8)$$

Таким образом, в секторе $(\max\{\max\{K_1^O, K_1^S\}, K_1^a\}; \min\{K_2^O, K_2^S\})$, гарантирующем абсолютную устойчивость нечеткой САУ, выполняется условие обеспечения требуемых показателей качества переходных процессов и ошибки в установившемся режиме.

Рассмотрим пример синтеза параметров НР I рода для задачи стабилизации положения равновесия нечеткой САУ с линейной частью 3-го порядка с частотной характеристикой $W(j\omega)$. В соответствии с методикой, изложенной выше, на рис. 4 приведены необходимые построения. Из них следует, что:

- требуемое значение η_0 обеспечивается, если характеристика НР располагается в секторе (0.6; 2.5);

- требуемое значение m_0 обеспечивается в секторе $(0.05; 6.1)$.

Если дополнительно учесть требование по ошибке в установившемся режиме и (7), то окончательно искомый сектор будет соответствовать рис. 4б. На рис. 4в представлен переходной процесс $x(t)$ в нечеткой САУ, где НР I рода имеет синтезированную нелинейную характеристику.

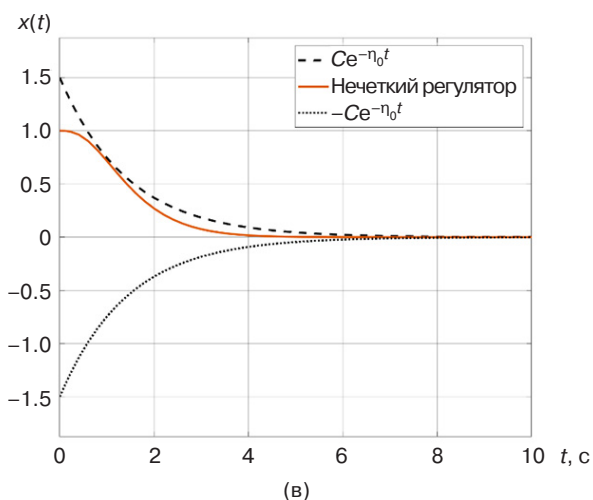
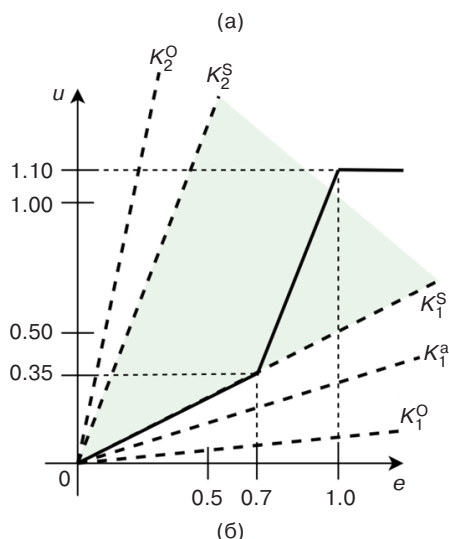
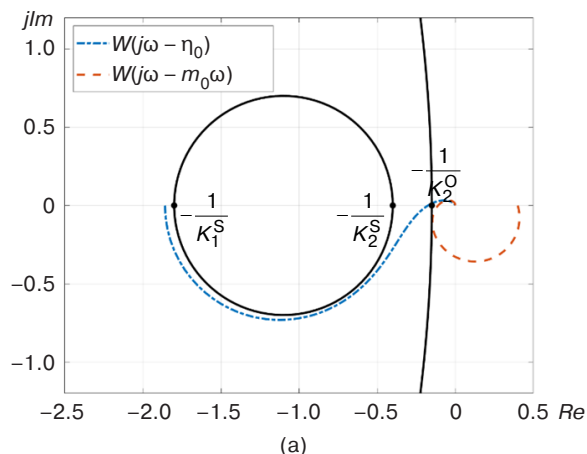


Рис. 4. Круговой критерий Якубовича (а), секторы абсолютной устойчивости с нелинейной характеристикой (б) и переходной процесс в САУ (в)

На основе полученного нелинейного преобразования НР и исходя из приведенных выше рекомендаций по организации структуры базы знаний, нетрудно построить соответствующее ее наполнение:

- поскольку синтезированное нелинейное преобразование имеет 4 точки изменения коэффициента усиления, терм-множество T_E для лингвистической переменной ошибки E будет содержать 5 нечетких множеств $T_E = \{A_0, A_1, A_2, A_3, A_4\}$;
- ФП, задающие нечеткие множества A_0 и A_4 , принадлежат классам L и γ , соответственно (в силу ограничения выходного сигнала);
- остальные ФП принадлежат классу t .

Как было показано ранее, область определения i -го кусочно-линейного участка статической характеристики регулятора зависит от взаимного расположения соседних ФП, а область значений – от величин заключений соответствующих правил из базы правил. Пробегая по всей области определения нелинейной характеристики (рис. 4б) и учитывая ее симметричность относительно начала координат, нетрудно определить параметры ФП (рис. 5) входной переменной, а также значения ФП выходной переменной, заложенные в базу правил:

- ЕСЛИ E есть A_0 , ТО $u = 1.1$;
- ЕСЛИ E есть A_1 , ТО $u = 0.35$;
- ЕСЛИ E есть A_2 , ТО $u = 0$;
- ЕСЛИ E есть A_3 , ТО $u = -0.35$;
- ЕСЛИ E есть A_4 , ТО $u = -1.1$.

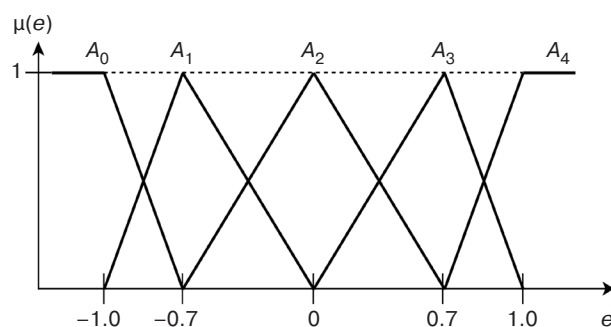


Рис. 5. Расположение функций ФП

ЗАКЛЮЧЕНИЕ

В работе развивается концепция, предложенная Макаровым с соавторами, по которой нечеткое преобразование по Заде, реализуемое в контуре САУ НР I рода, есть, по существу, нелинейное преобразование. В случае использования модели Сугено оно является кусочно-линейным. Для такой нечеткой системы с привлечением методов теории нелинейных САУ решена задача исследования динамики в удобной для инженера-разработчика форме и предложена методика, которая обеспечивает определение не только гарантированной области устойчивости,

но и требуемые показатели качества процесса управления.

Очевидно, что полученные результаты справедливы и при использовании в системах управления моделей НЛВ Мамдани, Ларсена и Цукамото. Действительно, поскольку нелинейные преобразования, соответствующие этим моделям, имеют гладкий характер, то они могут быть аппроксимированы кусочно-линейными участками и, таким образом, решение задачи сводится к предложенной выше методике. При этом линейный участок аппроксимации

в начале координат (с малым наклоном) выбирается исходя из требуемой ошибки в установившемся режиме, а наклон крутого участка – исходя из требований по качеству в соответствии с критерием Якубовича. Такой подход позволяет решать задачи как анализа, так и синтеза нечетких САУ с регулятором I рода.

Вклад авторов. Все авторы в равной степени внесли свой вклад в исследовательскую работу.

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

СПИСОК ЛИТЕРАТУРЫ

1. Макаров И.М., Лохин В.М. *Интеллектуальные системы автоматического управления*. М.: Физматлит; 2001. 576 с. ISBN 978-5-9221-0162-2
2. Пospelov Д.А. (ред.). *Нечеткие множества в моделях управления и искусственного интеллекта*. М.: Наука; 1986. 312 с.
3. Макаров И.М., Лохин В.М., Манько С.В., Романов М.П. *Искусственный интеллект и интеллектуальные системы управления*. М.: Наука; 2006. 333 с.
4. Makarov I.M., Lokhin V.M. *Artificial Intelligence and Complex Objects Control*. Lewiston: Edwin Mellen Press; 2000. 404 p.
5. Piegat A. *Fuzzy Modeling and Control*. Berlin: Physica Heidelberg; 2001. 728 p.
6. Макаров И.М., Лохин В.М., Манько С.В., Романов М.П., Ситников М.С. Устойчивость интеллектуальных систем автоматического управления. *Информационные технологии*. 2013;2:1–32.
7. Рутковская Д., Пилиньский М., Рутковский Л. *Нейронные сети, генетические алгоритмы и нечеткие системы*: пер. с пол. М.: Горячая линия–Телеком; 2006. 452 с.
8. Hashemi S.M., Botez R. Lyapunov-based Robust Adaptive Configuration of the UAS-S4 Flight Dynamics Fuzzy Controller. *The Aeronautical Journal*. 2022;126(1301):1187–1209. <https://doi.org/10.1017/aer.2022.2>
9. Gandhi R., Adhyaru D. Takagi-Sugeno fuzzy regulator design for nonlinear and unstable systems using negative absolute eigenvalue approach. *IEEE/CAA Journal of Automatica Sinica*. 2020;7(2):482–493. <https://doi.org/10.1109/JAS.2019.1911444>
10. Lan L., Tiem N., Co Nhu V. Absolute Stability for a Class of Takagi-Sugeno Fuzzy Control Systems. In: *3rd International Conference on Robotics, Control and Automation Engineering (RCAE)*. 2020. P. 47–51. <https://doi.org/10.1109/RCAE51546.2020.9294352>
11. Sakly A., Zahra B., Benrejeb M. Stability Study of Mamdani's Fuzzy Controllers Applied to Linear Plants. *Studies in Informatics and Control*. 2008;17(4):441–452.
12. Siddikov I., Porubay O., Rakhimov T. Synthesis of the neuro-fuzzy regulator with genetic algorithm. *Int. J. Electric. Comput. Eng. (IJECE)*. 2024;14(1):184–191. <http://doi.org/10.11591/ijece.v14i1.pp184-191>
13. Hamza M., Yap I., Choudhury I. Genetic Algorithm and Particle Swarm Optimization Based Cascade Interval Type 2 Fuzzy PD Controller for Rotary Inverted Pendulum System. *Math. Probl. Eng.* 2015;2015(6). <https://doi.org/10.1155/2015/695965>
14. Mahmoodabadi M., Babak N. Robust fuzzy linear quadratic regulator control optimized by multi-objective high exploration particle swarm optimization for a 4 degree-of-freedom quadrotor. *Aerosp. Sci. Technol.* 2019;97:105598. <https://doi.org/10.1016/j.ast.2019.105598>
15. Sakalli A., Beke A., Kumbasar T. Gradient Descent and Extended Kalman Filter based self-tuning Interval Type-2 Fuzzy PID controllers. In: *2016 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*. 2016. P. 1592–1598. <https://doi.org/10.1109/FUZZ-IEEE.2016.7737880>
16. Islam S.U., Zeb K., Kim S. Design of Robust Fuzzy Logic Controller Based on Gradient Descent Algorithm with Parallel-Resonance Type Fault Current Limiter for Grid-Tied PV System. *Sustainability*. 2022;14(19):12251. <https://doi.org/10.3390/su141912251>
17. Быковцев Ю.А. Синтез нечеткого регулятора на основе оценки степени устойчивости системы управления. *Мехатроника, автоматизация, управление*. 2022;23(6):295–301. <https://doi.org/10.17587/mau.23.295-301>
18. Bykovtsev Y.A., Lokhin V.M. Estimation of the accuracy of a control system with a fuzzy PID controller based on the approximation of the static characteristic of the controller. *Mekhatronika, Avtomatizatsiya, Upravlenie*. 2021;22(12):619–624. <https://doi.org/10.17587/mau.22.619-624>

REFERENCES

1. Makarov I.M., Lokhin V.M. *Intellektual'nye sistemy avtomaticheskogo upravleniya (Intelligent Automatic Control Systems)*. Moscow: Fizmatlit; 2001. 576 p. (in Russ.). ISBN 978-5-9221-0162-2
2. Pospelov D.A. (Ed.). *Nechetkie mnozhestva v modelyakh upravleniya i iskusstvennogo intellekta (Fuzzy Sets in Control Models and Artificial Intelligence)*. Moscow: Nauka; 1986. 312 p. (in Russ.).

3. Makarov I.M., Lokhin V.M., Manko S.V., Romanov M.P. *Iskusstvennyi intellekt i intellektual'nye sistemy upravleniya (Artificial Intelligence and Intelligent Control Systems)*. Moscow: Nauka; 2006. 333 p. (in Russ.).
4. Makarov I.M., Lokhin V.M. *Artificial Intelligence and Complex Objects Control*. Lewiston: Edwin Mellen Press; 2000. 404 p.
5. Piegat A. *Fuzzy Modeling and Control*. Berlin: Physica Heidelberg; 2001. 728 p.
6. Makarov I.M., Lokhin V.M., Manko S.V., Romanov M.P., Sitnikov M.S. Stability of intellectual automatic control systems. *Informatsionnye tekhnologii = Information Technologies*. 2013;2:1–32 (in Russ.).
7. Rutkowska D., Pilinski M., Rutkowski L. *Neironnye seti, geneticheskie algoritmy i nechetkie sistemy (Neural Networks, Genetic Algorithms and Fuzzy Systems)*: transl. from Pol. Moscow: Goryachaya liniya–Telekom; 2006. 452 p. (in Russ.). [Rutkowska D., Piliński M., Rutkowski L. *Sieci Neuronowe, Algorytmy Genetyczne i Systemy Rozmyte*. Warszawa; Łódź: Wydawnictwo Naukowe PWN. 2004.]
8. Hashemi S.M., Botez R. Lyapunov-based Robust Adaptive Configuration of the UAS-S4 Flight Dynamics Fuzzy Controller. *The Aeronautical Journal*. 2022;126(1301):1187–1209. <https://doi.org/10.1017/aer.2022.2>
9. Gandhi R., Adhyaru D. Takagi-Sugeno fuzzy regulator design for nonlinear and unstable systems using negative absolute eigenvalue approach. *IEEE/CAA Journal of Automatica Sinica*. 2020;7(2):482–493. <https://doi.org/10.1109/JAS.2019.1911444>
10. Lan L., Tiem N., Co Nhu V. Absolute Stability for a Class of Takagi-Sugeno Fuzzy Control Systems. In: *3rd International Conference on Robotics, Control and Automation Engineering (RCAE)*. 2020. P. 47–51. <https://doi.org/10.1109/RCAE51546.2020.9294352>
11. Sakly A., Zahra B., Benrejeb M. Stability Study of Mamdani's Fuzzy Controllers Applied to Linear Plants. *Studies in Informatics and Control*. 2008;17(4):441–452.
12. Siddikov I., Porubay O., Rakhimov T. Synthesis of the neuro-fuzzy regulator with genetic algorithm. *Int. J. Electric. Comput. Eng. (IJECE)*. 2024;14(1):184–191. <http://doi.org/10.11591/ijece.v14i1.pp184-191>
13. Hamza M., Yap I., Choudhury I. Genetic Algorithm and Particle Swarm Optimization Based Cascade Interval Type 2 Fuzzy PD Controller for Rotary Inverted Pendulum System. *Math. Probl. Eng.* 2015;2015(6). <https://doi.org/10.1155/2015/695965>
14. Mahmoodabadi M., Babak N. Robust fuzzy linear quadratic regulator control optimized by multi-objective high exploration particle swarm optimization for a 4 degree-of-freedom quadrotor. *Aerosp. Sci. Technol.* 2019;97:105598. <https://doi.org/10.1016/j.ast.2019.105598>
15. Sakalli A., Beke A., Kumbasar T. Gradient Descent and Extended Kalman Filter based self-tuning Interval Type-2 Fuzzy PID controllers. In: *2016 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*. 2016. P. 1592–1598. <https://doi.org/10.1109/FUZZ-IEEE.2016.7737880>
16. Islam S.U., Zeb K., Kim S. Design of Robust Fuzzy Logic Controller Based on Gradient Descent Algorithm with Parallel-Resonance Type Fault Current Limiter for Grid-Tied PV System. *Sustainability*. 2022;14(19):12251. <https://doi.org/10.3390/su141912251>
17. Bykovtsev Y.A. Synthesis of a Fuzzy Controller According to the Degree of Stability of the Control System. *Mekhatronika, Avtomatizatsiya, Upravlenie*. 2022;23(6):295–301 (in Russ.). <https://doi.org/10.17587/mau.23.295-301>
18. Bykovtsev Y.A., Lokhin V.M. Estimation of the accuracy of a control system with a fuzzy PID controller based on the approximation of the static characteristic of the controller. *Mekhatronika, Avtomatizatsiya, Upravlenie*. 2021;22(12):619–624. <https://doi.org/10.17587/mau.22.619-624>

Об авторах

Быковцев Юрий Алексеевич, к.т.н., доцент, кафедра проблем управления, Институт искусственного интеллекта, ФГБОУ ВО «МИРЭА – Российский технологический университет» (119454, Россия, Москва, пр-т Вернадского, д 78). E-mail: bykovtsev@mirea.ru. Scopus Author ID 57302607300, ResearcherID KRQ-5339-2024, SPIN-код РИНЦ 9961-4437, <https://orcid.org/0009-0003-6671-5674>

Лохин Валерий Михайлович, д.т.н., профессор, кафедра проблем управления, Институт искусственного интеллекта, ФГБОУ ВО «МИРЭА – Российский технологический университет» (119454, Россия, Москва, пр-т Вернадского, д. 78). Лауреат государственной премии РФ в области науки и техники. Лауреат премии Правительства РФ в области образования. Член научного Совета РАН по робототехнике и мехатронике. Заслуженный деятель науки РФ. E-mail: kpu-mirea@yandex.ru. Scopus Author ID 6602931640, <https://orcid.org/0000-0001-6708-9124>

About the Authors

Yuri A. Bykovtsev, Cand. Sci. (Eng.), Assistant Professor, Department of Management Problems, Institute of Artificial Intelligence, MIREA – Russian Technological University (78, Vernadskogo pr., Moscow, 119454 Russia). E-mail: bykovtsev@mirea.ru. Scopus Author ID 57302607300, ResearcherID KRQ-5339-2024, RSCI SPIN-code 9961-4437, <https://orcid.org/0009-0003-6671-5674>

Valery M. Lokhin, Dr. Sci. (Eng.), Professor, Department of Management Problems, Institute of Artificial Intelligence, MIREA – Russian Technological University (78, Vernadskogo pr., Moscow, 119454 Russia). Laureate of the State Prize of the Russian Federation in Science and Technology. Laureate of the State Prize of the Russian Federation in Education. Member of the Scientific Council on Robotics and Mechatronics of the Russian Academy of Sciences. Honored Worker of Science of the Russian Federation. E-mail: kpu-mirea@yandex.ru. Scopus Author ID 6602931640, <https://orcid.org/0000-0001-6708-9124>

Роботизированные комплексы и системы. Технологии дистанционного зондирования
неразрушающего контроля

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

УДК 623.09

<https://doi.org/10.32362/2500-316X-2025-13-3-63-72>

EDN SONRQG



НАУЧНАЯ СТАТЬЯ

Идентификационный комплекс обнаружения и подавления беспилотных воздушных судов на основе видеоперехвата для использования в условиях боевых действий

А.А. Коник¹,
Р.Е. Афонин¹,
И.М. Климов²,
Д.С. Зинченко^{2, @}

¹ Белгородский юридический институт МВД России имени И.Д. Путилина, Белгород, 308024 Россия

² ФКУ НПО «СТИС» МВД России, Москва, 111024 Россия

@ Автор для переписки, e-mail: dmitriy.zinchenko.1998@mail.ru

• Поступила: 03.02.2025 • Доработана: 17.02.2025 • Принята к опубликованию: 21.03.2025

Резюме

Цели. В условиях проведения боевых действий государства столкнулись с угрозами атак беспилотных воздушных судов (БВС) на различные объекты, в частности, объекты силовых структур. В настоящее время нет единого, а самое главное, эффективного подхода по обнаружению и подавлению различных видов БВС, а именно, FPV-дронов. Целью работы является создание модернизированных комплексов обнаружения и подавления БВС противника и обоснование полномасштабного внедрения этих комплексов в служебную деятельность силовых структур.

Методы. В работе использовались системно-структурный, сравнительно-правовой, измерительный методы исследования, анализ, наблюдение и натурное моделирование, а также обобщение и систематизация опыта применения БВС в условиях боевых действий.

Результаты. Описаны структура, основные тактико-технические характеристики и принцип работы идентификационного комплекса обнаружения и подавления БВС с возможностью перехвата аналогового радиосигнала, несущего видеоинформацию, позволяющего эффективно обнаруживать и противодействовать БВС противника на значительном расстоянии. Разработан и описан алгоритм действий сотрудников силовых структур при работе с данным комплексом, изложены предложения по созданию на базе комплекса аппаратно-программных систем с возможностью подмены видеопотока.

Выводы. Полученные результаты исследования указывают на необходимость внедрения в служебную деятельность силовых структур идентификационного комплекса обнаружения и подавления БВС с возможностью перехвата аналогового радиосигнала, несущего видеоинформацию, применение которого на широком участке местности, подверженном атакам БВС, позволит существенно улучшить эффективность оповещения всех категорий служащих и гражданского населения, а также будет способствовать установлению контроля над воздушным пространством и повышению эффективности борьбы с БВС.

Ключевые слова: идентификационный комплекс; аналоговый радиосигнал, видеоинформация, беспилотное воздушное судно, антенна, дескремблирование, видеопоток

Для цитирования: Коник А.А., Афонин Р.Е., Климов И.М., Зинченко Д.С. Идентификационный комплекс обнаружения и подавления беспилотных воздушных судов на основе видеоперехвата для использования в условиях боевых действий. *Russian Technological Journal*. 2025;13(3):63–72. <https://doi.org/10.32362/2500-316X-2025-13-3-63-72>, <https://www.elibrary.ru/SONRQG>

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

Авторы заявляют об отсутствии конфликта интересов.

RESEARCH ARTICLE

Identification system for detecting and suppressing unmanned aerial vehicles based on video interception for use in combat conditions

Alexey A. Konik ¹,
Roman E. Afonin ¹,
Ivan M. Klimov ²,
Dmitry S. Zinchenko ^{2, @}

¹ I.D. Putilin Belgorod Law Institute of the Ministry of Internal Affairs of Russia, Belgorod, 308024 Russia

² FKU NGO “STiS”, Ministry of Internal Affairs of Russia, Moscow, 111024 Russia

@ Corresponding author, e-mail: dmitriy.zinchenko.1998@mail.ru

• Submitted: 03.02.2025 • Revised: 17.02.2025 • Accepted: 21.03.2025

Abstract

Objectives. In the context of military operations, states face the threats of attacks by unmanned aerial vehicles (UAVs) on vulnerable assets, in particular, those pertaining to law enforcement agencies. Currently, there is no uniform or—more importantly—effective approach to detecting and suppressing certain types of UAVs, in particular, first person view (FPV) drones. The aim of the work is to develop modernized systems for the detection and suppression of enemy unmanned aerial vehicles and to justify their full-scale implementation in the service activities of law enforcement agencies.

Methods. The work used system-structural, comparative-legal, and measurement research methods along with analysis, observation, and field modeling. In addition, the research refers to generalized and systematized experience of using UAVs in combat conditions.

Results. The structure, basic tactical and technical characteristics are described according to the principle of operation of the identification complex for detecting and suppressing UAVs, having the ability to intercept analog radio signals carrying video information, which makes it possible to effectively detect and counteract enemy UAVs at a considerable distance. An algorithm of actions to be taken by law enforcement officers when using this complex is also developed and described. Proposals for creating a hardware and software system based on the complex with the possibility of spoofing a video stream are outlined.

Conclusions. The results of the study indicate the need to supply law enforcement agencies with an identification

system for detecting and suppressing UAVs having the ability to intercept an analog radio signal carrying video information. The use of such a system across a wide area subject to UAV attacks will significantly improve the effectiveness of alerting all categories of employees and civilians, as well as contributing to the establishment of airspace control and improving the effectiveness of the fight against UAVs.

Keywords: identification system, analog radio signal, video information, unmanned aerial vehicle, antenna, descrambling, video stream

For citation: Konik A.A., Afonin R.E., Klimov I.M., Zinchenko D.S. Identification system for detecting and suppressing unmanned aerial vehicles based on video interception for use in combat conditions. *Russian Technological Journal*. 2025;13(3):63–72. <https://doi.org/10.32362/2500-316X-2025-13-3-63-72>, <https://www.elibrary.ru/SONRQG>

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

The authors declare no conflicts of interest.

ВВЕДЕНИЕ

Специальная военная операция (СВО) продемонстрировала возможности современных беспилотных воздушных судов (БВС). Атаки БВС, осуществляемые с использованием БВС коптерного типа с системой «сброс» и БВС типа FPV-дроны¹ [1], представляют опасность не только для гражданских, промышленных зданий и сооружений, но и для объектов критической инфраструктуры, режимных объектов силовых структур и их личного состава [2].

В условиях военных действий территории многих приграничных регионов регулярно подвергается неизбирательным, а в отдельных случаях хаотичным атакам БВС, снаряженных различными боеприпасами и взрывными устройствами (ВУ) (ПГ-7В² (ВЛ³, ВР⁴, ВМ⁵), РКГ-3⁶) [3], в частности, самодельными ВУ (рис. 1). Так, согласно информации БЕЛ.RU, за 7 месяцев 2024 года Белгородская область подверглась атакам БВС со стороны противника 667 раз, в результате которых пострадали несколько десятков человек, повреждены и уничтожены несколько сотен единиц транспорта⁷. По официальным данным губернатора Белгородской области за одну неделю августа 2024 года зафиксировано более 240 атак



Рис. 1. Самодельное ВУ, примененное при атаке БВС в приграничном регионе

БВС противника, большинство из которых осуществлено при помощи FPV-дронов⁸.

Кроме того, в последнее время разработаны и активно используются системы «сброса» под FPV-дроны. Они способны осуществлять многоразовые атаки с применением различных модернизированных боеприпасов [4]: например, на территории приграничных регионов с помощью FPV-дронов противник регулярно осуществляет дистанционное минирование местности, разбрасывая боеприпасы типа «лепесток» и «колокольчик».

Указанные факторы, а также анализ статистических данных указывают на наиболее массовую опасность для гражданского населения, личного состава и объектов силовых структур приграничных регионов со стороны FPV-дронов, которые атакуют на большой скорости (более 100 км/ч), несут значительную нагрузку в виде боевого заряда и могут проникать на территорию Российской Федерации на расстояние 15–20 км, а в отдельных случаях и до 30 км [5].

⁸ Официальный Telegram-канал губернатора Белгородской области В.В. Гладкова. <https://t.me/vvgladkov>. Дата обращения 20.01.2025. / The official Telegram channel of the Governor of the Belgorod oblast V.V. Gladkov. <https://t.me/vvgladkov>. Accessed January 20, 2025.

¹ First person view – вид от первого лица.

² Выстрел с противотанковой гранатой. [A shot with an anti-tank grenade.]

³ Выстрел с противотанковой гранатой «Луч». [A shot with an anti-tank grenade “Luch.”]

⁴ Выстрел с противотанковой гранатой «Резюме». [A shot with an anti-tank grenade “Resume.”]

⁵ Выстрел с противотанковой гранатой модернизированной. [A modernized anti-tank grenade shot.]

⁶ Ручная кумулятивная граната. [A handheld shaped-charge grenade.]

⁷ Стало известно, сколько человек погибло на белгородских дорогах из-за атак дронов. <https://bel.ru/news/2024-08-12/stalo-izvestno-skolko-chelovek-pogiblo-na-belgorodskih-dorogah-iz-za-atak-dronov-5164260>. Дата обращения 12.12.2024. / It became known how many people died on Belgorod roads due to drone attacks. <https://bel.ru/news/2024-08-12/stalo-izvestno-skolko-chelovek-pogiblo-na-belgorodskih-dorogah-iz-za-atak-dronov-5164260>. Accessed December 12, 2024.

Помимо этого, в ходе боевых действий выяснилось, что противник для управления FPV-дронами использует нестандартные, постоянно смещенные частоты [6], что значительно повышает их устойчивость к обнаружению средствами радиоэлектронной разведки (РЭР) и радиоэлектронной борьбы (РЭБ).

В связи с прогрессивно осложняющейся обстановкой в приграничных регионах и активизацией применения противником FPV-дронов необходимо введение дополнительных мер, позволяющих повысить эффективность противодействия противнику в данном контексте, а также своевременно оповестить гражданское население об угрозах атак БВС. Актуальным направлением представляется разработка модернизированных комплексов обнаружения и подавления БВС, а именно, создание и внедрение в служебную деятельность сотрудников силовых структур приграничных регионов идентификационного комплекса обнаружения и подавления БВС с возможностью перехвата аналогового радиосигнала, несущего видеoinформацию [7].

ХАРАКТЕРИСТИКА ИДЕНТИФИКАЦИОННОГО КОМПЛЕКСА ОБНАРУЖЕНИЯ И ПОДАВЛЕНИЯ БВС С ВОЗМОЖНОСТЬЮ ПЕРЕХВАТА АНАЛОГОВОГО РАДИОСИГНАЛА, НЕСУЩЕГО ВИДЕОИНФОРМАЦИЮ

На основании изложенного, а также установления и подтверждения в ходе СВО того факта, что ключевым элементом РЭБ и эффективным контуром поражения БВС противника является РЭР (раннее обнаружение БВС) [8], специалистами Белгородской области в инициативном порядке разработан идентификационный комплекс обнаружения и подавления БВС (далее – Комплекс), позволяющий с высокой вероятностью обнаруживать и идентифицировать нахождение в воздушном пространстве БВС противника. Функционирование данного Комплекса строится на перехвате аналогового радиосигнала, несущего видеoinформацию, с БВС противника.

Следует отметить, что для обеспечения данной функции в настоящее время разработаны схожие изделия «ЭФИР», «Зонтик» и др. (рис. 2). Основной принцип действия таких комплексов основан на сканировании радиоэфира и регистрации несущей частоты с определением и фиксацией видеопотока.

Однако анализ и оценка тактико-технических характеристик таких изделий показали, что они не позволяют одновременно обнаруживать сигналы БВС в широком диапазоне частот (в изделиях заложено минимальное количество частотных диапазонов, например, 2.4 ГГц и 5.8 ГГц) и на дальних расстояниях (заявленные дальности обнаружения составляют от 500 м до 5 км) [9].



Рис. 2. Детектор БВС «ЭФИР»

Разработанный идентификационный Комплекс включает в себя следующие компоненты:

- направленные приемные антенны на определенные диапазоны частот (можно использовать круговые антенны, но дистанция обнаружения в таком случае значительно меньше);
- приемник аналогового радиосигнала, несущего видеoinформацию;
- дескремблер (дешифратор) аналогового видеопотока;
- кабель передачи видеосигнала с разъемами RCA (длина кабеля зависит от конфигурации установки оборудования, при этом важно учитывать, что с увеличением длины кабеля снижается качество передачи данных);
- монитор с разъемом VGA или HDMI;
- конвертер для подключения монитора AV RCA-VGA или AV RCA-HDMI (зависит от используемого монитора);
- кабель VGA-VGA или HDMI-HDMI;
- источник питания 2 USB, 5V/2.1A.

Принцип действия Комплекса заключается в том, что с помощью круговой антенны (если неизвестно направление вероятного вылета БВС) или направленной антенны осуществляется непрерывный мониторинг окружающего пространства в радиусе от 1.5 до 3 км (при использовании круговой антенны) и до 30 км (при использовании направленной антенны). Указанные расстояния зависят от ряда факторов: мощности передатчика FPV-дрона, усиления принимающей антенны, рельефа местности и т.д. При появлении БВС, использующего аналоговый канал передачи видеосигнала, Комплекс автоматически осуществляет перехват и трансляцию (дублирование) на монитор видео с иллюстрацией местности, над которой летит БВС, и полетными данными, которые наблюдает на экране монитора или в очках для FPV-дрона управляющий им оператор.

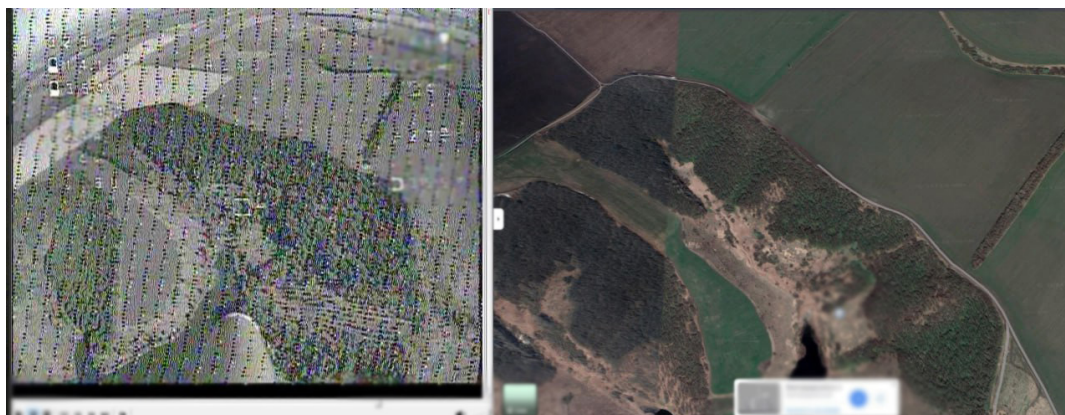


Рис. 3. Выявление направлений полетов БВС противника путем сопоставления данных Комплекса на картах местности

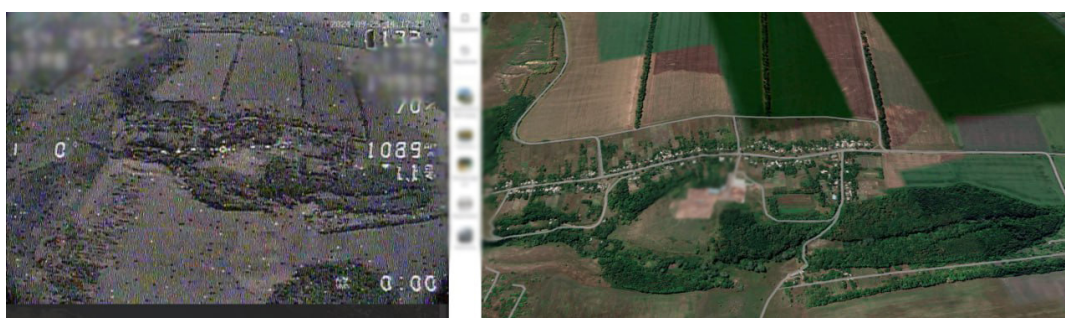


Рис. 4. Фиксация аналогового радиосигнала БВС противника, несущего видеoinформацию, и привязка маршрута полета на карте местности

Разработанный Комплекс способен работать в широком температурном диапазоне, при сильных вибрациях и в различных экстремальных условиях. Важным аспектом функционирования Комплекса является его пассивность – он не излучает собственные сигналы, которые могли бы быть обнаружены сканирующими радиоэфир устройствами.

Отличительной особенностью Комплекса является возможность не только обнаружения БВС путем приема (вывода на монитор) аналогового видеосигнала, но и дескрипирования зашифрованных «Х-сигналов», что позволяет заблаговременно выявлять факт противоправного полета БВС с возможностью определения скорости, высоты полета, а также спрогнозировать траекторию его движения и определить возможные цели. Учитывая дистанцию обнаружения БВС противника с помощью данного Комплекса (до 30 км), имеется время для принятия мер по обеспечению безопасности и организации противодействия.

Одновременно следует отметить, что помимо указанных характеристик Комплекс обладает уникальными возможностями и потенциалом в условиях проведения СВО. Однако в силу конфиденциальности информации, схематика и более подробные тактико-технические характеристики Комплекса не могут быть отражены.

ПРАКТИЧЕСКОЕ ПРИМЕНЕНИЕ РАЗРАБОТАННОГО ИДЕНТИФИКАЦИОННОГО КОМПЛЕКСА ОБНАРУЖЕНИЯ И ПОДАВЛЕНИЯ БВС

Разработанный Комплекс был установлен в одном из приграничных регионов и успешно подтвердил свои тактико-технические характеристики. За несколько месяцев применения Комплекса было обнаружено свыше 500 БВС. Например, только за одни сутки (20 августа 2024 г.) Комплекс способствовал обнаружению на территории 6 поселков приграничного региона 20 FPV-дронов, из которых 9 были подавлены⁹.

С помощью Комплекса при одновременном анализе топографических карт, в т.ч. электронных офлайн-карт, были выявлены направления полетов БВС противника (рис. 3 и 4). Кроме того, при определенных условиях имеется возможность определения точек запуска БВС. Вместе с тем, в целях использования Комплекса на передвижных объектах в настоящее время ведется активная работа по его интеграции в мобильную версию.

⁹ Статистические данные (материалы) предоставлены инициативной группой – разработчиками Комплекса. [Statistical data (materials) are provided by the initiative group, the developers of the complex.]

Также Комплекс был использован в деятельности сотрудников силового блока одного из приграничных регионов, что позволило в короткие сроки существенно повысить эффективность обнаружения и подавления БВС противника на подконтрольном участке (уничтожено 18 FPV-дронов). Вместе с тем, в процессе практического использования Комплекса установлено, что неотъемлемой частью его применения является ведение соответствующей документации (журнал, ведомость и т.д.), в которой необходимо в обязательном порядке отражать полученную информацию (место обнаружения БВС с привязкой к населенному пункту или точными координатами, характер поведения БВС, предпринятые меры и полученный результат и т.д.). С этой целью авторский коллектив ФКУ НПО «СТиС» МВД России¹⁰ и БелЮИ МВД России имени И.Д. Путилина¹¹ разработал оптимальный шаблон журнала учета данных по обнаружению противоправного функционирования БВС.

В связи с этим, для своевременного обнаружения и эффективного подавления БВС предлагается внедрение данного Комплекса в служебную деятельность сотрудников силовых структур и создание сети опорных пунктов, оборудованных Комплексом, средствами РЭБ и иными средствами подавления БВС на базе созданных блокпостов, контрольно-пропускных пунктов, дислокаций, а также на временных позициях базирования силовых структур в приграничных регионах с дальнейшим оповещением заинтересованных сил силового блока по закрытым каналам связи и гражданского населения – по открытым (например, группы в мессенджерах и т.д.).

Одновременно необходимо определить алгоритм действий сотрудников силовых структур [10] при работе с Комплексом:

- выявление угрозы;
- оповещение личного состава, оперативного дежурного, группы противодействия и гражданского населения по подготовленным каналам связи;
- принятие мер по подавлению БВС противника всеми возможными способами (средства РЭБ, физическая ликвидация, в т.ч. с помощью огнестрельного оружия [11] и т.п.);
- принятие мер по устранению последствий подавления БВС или его детонации, а также документированию происходящего.

Оповещение личного состава необходимо выполнять одновременно с оповещением гражданского населения. Сообщение должно носить лаконичный, но информативный характер: направление

движения БВС; предполагаемый тип; меры, которые необходимо предпринять.

Подавление БВС может осуществляться как непосредственно рядом с постом наблюдения, так и с применением тактики выездных мобильных или засадных групп (в распоряжении таких групп обязательно должны быть дополнительные средства РЭР и различные средства подавления БВС (комплексы РЭБ, гладкоствольное оружие и др.), которые будут осуществлять перехват дронов и их подавление на удалении от гражданских и критически важных объектов.

При этом необходимо учитывать, что применение средств РЭБ в непосредственной близости от Комплекса может привести к появлению помех и неисправности оборудования. Поэтому при использовании различных комплексов РЭБ и иных средств подавления БВС в непосредственной близости от Комплекса его необходимо отключать. Важно учитывать, что применение средств РЭБ не является гарантией подавления БВС противника [12], поэтому необходимо быть готовым к применению огнестрельного оружия для поражения дрона. На низких высотах (до 40 м) эффективное подавление БВС возможно обеспечить с использованием гладкоствольного оружия, снаряженного патронами с дробью № 3 и № 5 или созданием «плотности» огня из 3–5 единиц автоматического оружия [13].

После ликвидации БВС противника при необходимости следует оказать первую помощь пострадавшим (при наличии), изолировать место происшествия от посторонних лиц и убедиться, что на уцелевших частях дрона не остались дополнительно заминированные детали, после чего, соблюдая меры безопасности, собрать уцелевшие части дрона для их исследования. Если БВС вызывает сомнения или его подавление произошло без детонации ВУ, категорически запрещается приближаться к нему до исследования объекта специалистами, т.к. ВУ может быть приведено в действие дистанционно или сдетонировать по истечении определенного времени (самоуничтожение через определенный промежуток времени).

Вместе с тем в целях формирования у личного состава знаний и умений по алгоритму действий при обнаружении БВС необходимо регулярно проводить инструктажи.

СТРАТЕГИЧЕСКИЕ ЗАДАЧИ, РЕШАЕМЫЕ С ПОМОЩЬЮ РАЗРАБОТАННОГО ИДЕНТИФИКАЦИОННОГО КОМПЛЕКСА ОБНАРУЖЕНИЯ И ПОДАВЛЕНИЯ БВС

Одной из стратегических целей применения Комплекса является сбор информации с последующим ее обязательным анализом и использованием в служебной деятельности сотрудниками

¹⁰ <https://стис.мвд.рф>. Дата обращения 20.03.2025. / Accessed March 20, 2025.

¹¹ <https://белюи.мвд.рф>. Дата обращения 20.03.2025. / Accessed March 20, 2025.

Минобороны, ОВД, СК, ФСБ, ПУ ФСБ, Росгвардии России и др.

При систематическом использовании Комплекса дополнительно решаются следующие задачи:

- выявление направления и дальнейшего следования БВС. Это позволяет отслеживать маршруты полетов БВС противника, в отдельных случаях – определять точку запуска, наиболее опасные направления, которые можно фиксировать в журналах наблюдения и на картах местности, в т.ч. с помощью сервисов AlpineQuest¹², OfflineMaps¹³ и т.п. Полученные данные можно использовать для усиления средств РЭБ на определенных направлениях и для организации засадных мероприятий;
- идентификация типов и видов БВС [14]. Характер «поведения» дрона в воздухе (скорость перемещения, высота полета, устойчивость и т.д.) с высокой точностью позволяют идентифицировать тип и вид БВС, что дает возможность прогнозировать дальнейшие действия оператора, а также подготовиться к вероятной угрозе;
- выявление групп противника (с наименованием БВС или информацией о позывных подразделений), которые осуществляют террористическую деятельность вблизи (на территории) приграничного региона (при перехвате аналогового радиосигнала, несущего видеоинформацию, данные сведения часто отображаются на экране). В перспективе эти данные после совершения актов нападения могут быть переданы для анализа в следственные органы. Кроме того, данный подход позволит точнее определять количество групп БВС противника и их численность на том или ином участке государственной границы;
- выявление преступников (предателей). В настоящее время имеются установленные факты запуска БВС диверсионно-разведывательными группами противника и преступников (предателей) с территории приграничного региона. Выявить данные факты без специального оборудования достаточно сложно, поэтому свою деятельность подобные группы могут осуществлять длительное время, нанося точечные повреждения объектам жизнеобеспечения и гражданским лицам. При качественной организации функционирования Комплекса место взлета БВС может быть зафиксировано по видеосигналу (при попадании в зону действия антенны), что позволит провести мероприятия по ликвидации и обезвреживанию подобных групп;
- фиксация подавления (ликвидации) БВС или результата применения ВУ. В данном случае речь идет

как о документальном подтверждении ликвидации БВС противника, так и о дальнейшем изъятии его фрагментов, в т.ч. при его подавлении без ликвидации для их дальнейшего исследования. Также известны случаи, когда дроны в дальнейшем находили гражданские лица и получали травмы при детонации ВУ, которыми они оборудованы. При атаке БВС противника и получении ранений использование Комплекса дает возможность оперативного реагирования и направления к указанному месту необходимых служб для оказания помощи;

- определение технических возможностей БВС противника [15]. Применение Комплекса с возможностью перехвата аналогового радиосигнала, несущего видеоинформацию, позволяет фиксировать дальность полета с определением предельных расстояний, отслеживать новые технические приемы, применяемые противником (например, с помощью специальных программ с трофейных БВС регулярно снимаются значения частот, которые используются при полетах), в т.ч. и увеличение мощности аккумуляторных батарей. С помощью Комплекса также можно определить помехоустойчивость БВС противника к применяемому РЭБ и вносить коррективы в его работу.

ЗАКЛЮЧЕНИЕ

Применение идентификационного комплекса обнаружения и подавления БВС с возможностью перехвата аналогового радиосигнала, несущего видеоинформацию, и оборудование им широкого участка местности в приграничных регионах, подверженных атакам БВС, позволит существенно улучшить эффективность оповещения всех категорий служащих и гражданского населения, а также будет способствовать установлению контроля над воздушным пространством и повышению эффективности борьбы с БВС. В свою очередь, сбор и анализ полученных с помощью Комплекса и зафиксированных в журнале учета данных позволит прогнозировать развитие обстановки на том или ином участке местности.

Установленные в ходе научного анализа и практического применения разработанного Комплекса свойства позволяют выделить инновационные характеристики, определяющие его потенциал в условиях проведения СВО:

- широкий частотный диапазон сканирования (более 4 частот; подробная информация носит конфиденциальный характер);
- дистанция обнаружения БВС составляет до 30 км;
- возможность дескремблирования зашифрованных «Х-сигналов»;
- стойкость в экстремальных условиях и широкий температурный диапазон работы: от –25 °С до +40 °С.

¹² <https://alpinequest.net/>. Дата обращения 20.03.2025. / Accessed March 20, 2025.

¹³ <https://offline-maps.softonic.ru/>. Дата обращения 20.03.2025. / Accessed March 20, 2025.

Вместе с тем результаты исследования определяют возможность создания на базе Комплекса аппаратно-программных систем с возможностью подмены видеопотока, подразумевающей постановку помех в виде трансляции оператору БВС стороннего видео, а также различных модификаций Комплекса: стационарного, автомобильного и др.

Вклад авторов

А.А. Коник – создание концепции и методологии исследования, подготовка первоначального варианта рукописи, обзор литературы, редактирование текста рукописи.

Р.Е. Афонин – разработка методологии, подготовка первоначального варианта рукописи, сбор и систематизация информации о практике применения комплекса, обзор литературы.

И.М. Климов – анализ существующих решений, предоставление рекомендаций по формулировке проблемы и разработке комплекса, проведение экспертной оценки полученных результатов.

Д.С. Зинченко – разработка методологии, подготовка первоначального варианта рукописи,

интерпретация и обобщение результатов, общее руководство, подготовка статьи к публикации.

Все авторы внесли общий равнозначный вклад, прочитали и согласовали публикуемый вариант рукописи.

Authors' contributions

A.A. Konik – creation of the research concept and methodology, preparation of the initial version of the manuscript, literature review, editing of the manuscript text.

R.E. Afonin – development of methodology, preparation of the initial version of the manuscript, collection and systematization of information on the practice of using the complex, literature review.

I.M. Klimov – analysis of existing solutions, providing recommendations on problem formulation and development of the complex, conducting an expert assessment of the results obtained.

D.S. Zinchenko – methodology development, preparation of the initial version of the manuscript, interpretation and generalization of the results, general guidance, preparation of the article for publication.

All the authors made a common equivalent contribution, read and agreed on the published version of the manuscript.

СПИСОК ЛИТЕРАТУРЫ

1. Рузняев Е.С. Зарубежные беспилотные летательные аппараты военного назначения. *Матрица научного познания*. 2022;6(2):42–47.
2. Павлов Р.А., Савельев К.П. Применение беспилотных летательных аппаратов в современных военных конфликтах. *Молодой ученый*. 2022;51(446):48–50.
3. Аверченко С.В., Белоусов В.В. Беспилотные летательные аппараты в военных конфликтах второй половины XX – начала XXI веков: основные вехи истории. *Современная научная мысль*. 2023;1:231–242.
4. Зинченко Д.С., Афонин Р.Е. Анализ средств и способов пассивной защиты от атак беспилотных воздушных судов в условиях проведения специальной военной операции. *Проблемы правоохранительной деятельности*. 2024;4(58):56–64. <https://elibrary.ru/ibjkhv>
5. Матарас А.А., Гуляев И.Ю. Анализ применения FPV дронов в ходе боевых действий 2014–2023 гг. В сб.: *Актуальные вопросы повышения эффективности огневой подготовки в силовых структурах: теория и практика (III Макаровские чтения): Всероссийский сборник научно-практических материалов*. Пермь. 2023. Т. 3. С. 135–141. <https://elibrary.ru/xcnjaj>
6. Николаев Н.В., Ильин В.В., Некрасов М.И. Актуальные вопросы противодействия современным автономным беспилотным летательным аппаратам и FPV-дронам. *Вопросы безопасности*. 2024;1:40–56.
7. Сулейманов М.В., Коняев В.М., Гуц С.И. Способы обнаружения малых БВС коммерческого типа. В сб.: *Средства и системы противодействия беспилотным воздушным судам: Сборник докладов научно-практической специальной конференции*. Москва. 2023. С. 22–26. <https://www.elibrary.ru/fcxjrg>
8. Красненко Н.П., Богушевич А.Я., Кураков С.А., Раков А.С., Рыбаков И.А. Обнаружение беспилотных летательных аппаратов: существующие решения и возможности. В сб.: *Всероссийские открытые Арmandовские чтения: Современные проблемы дистанционного зондирования, радиолокации, распространения и дифракции волн*. Муром. 2024. С. 429–438. <https://doi.org/10.24412/2304-0297-2024-1-429-440>
9. Фалилеев В.Ю., Шатовкин Р.Р. Анализ существующих автоматизированных комплексов защиты от дронов. *Воздушно-космические силы. Теория и практика*. 2020;14:130–140. <https://www.elibrary.ru/ayrwcg>
10. Коник А.А., Климов И.М. К вопросу о средствах и способах активной защиты от атак беспилотных воздушных судов в условиях проведения специальной военной операции. *Проблемы правоохранительной деятельности*. 2024;4(58):65–69. <https://www.elibrary.ru/kmijke>
11. Макаренко С.И., Тимошенко А.В. Анализ средств и способов противодействия беспилотным летательным аппаратам. Часть 2. Огневое поражение и физический перехват. *Системы управления, связи и безопасности*. 2020;1:147–197.
12. Тикшаев В.Н., Барвиненко В.В. Проблема борьбы с беспилотными летательными аппаратами и возможные пути ее решения. *Военная мысль*. 2021;1:125–131.
13. Моисеев Н.В. Перспективные варианты поражения БВС. В сб.: *Современная Российская наука: актуальные вопросы, достижения и инновации: сборник статей VI Всероссийской научно-практической конференции*. Пенза: Наука и Просвещение; 2023. С. 106–109.
14. Бердымуратов Д.Б. Анализ систем идентификации и формирования базы данных беспилотных летательных аппаратов. *Universum: технические науки*. 2024;10(127):4–8.
15. Храпков Д.С., Ромашов В.А. Беспилотные летательные аппараты, как современные средства ведения боевых действий. *Научная мысль*. 2021;15(1-1(39)):69–71. <https://www.elibrary.ru/oootziq>

REFERENCES

1. Ruznyaev E.S. Foreign unmanned aerial vehicles for military purposes. *Matritsa Nauchnogo Poznaniya*. 2022;6(2):42–47 (in Russ.).
2. Pavlov R.A., Saveliev R.A. The use of unmanned aerial vehicles in modern military conflicts. *Molodoi uchenyi = Young Scientist*. 2022;51(446):48–50 (in Russ.).
3. Averchenko S.V., Belousov V.V. Unmanned aerial vehicles in military conflicts of the second half of the 20th – beginning of the 21st centuries: the main milestones of history. *Sovremennaya nauchnaya mysl' = Modern Scientific Thought*. 2023;1:231–242 (in Russ.).
4. Zinchenko D.S., Afonin R.E. Analysis of means and methods of passive protection against attacks by unmanned aircraft in the context of a special military operation. *Problemy pravookhranitel'noi deyatel'nosti = Problems of Law Enforcement*. 2024;4(58):56–64 (in Russ.). <https://elibrary.ru/ibjkh>
5. Mataras A.A., Gulyaev I. Yu. Analysis of the use of FPV drones during combat operations 2014–2023. In: *Current Issues of Increasing Effective Fire Training in Law Enforcement Agencies: Theory and Practice (The Third Makarov Readings): All-Russian Collection of Scientific and Practical Materials*. Perm; 2023. V. 3. P. 135–141 (in Russ.). <https://elibrary.ru/xcnjax>
6. Nikolaev N.V., Ilyin V.V., Nekrasov M.I. Actual issues of countering modern autonomous unmanned aerial vehicles and FPV drones. *Voprosy bezopasnosti = Security Issues*. 2024;1:40–56 (in Russ.).
7. Suleimanov M.V., Konyaev V.M., Guts S.I. Methods of detecting small commercial UAVs. In: *Means and Systems for Countering Unmanned Aerial Vehicles: Collection of Reports of the Scientific and Practical Special Conference*. Moscow. 2023. P. 22–26 (in Russ.). <https://www.elibrary.ru/fcxjpg>
8. Krasnenko N.P., Bogushevich A.Ya., Kurakov S.A., Rakov A.S., Rybakov I.A. Detection of unmanned aerial vehicles: existing solutions and possibilities. In: *All-Russian Open Armand Readings: Modern Problems of Remote Sensing, Radar, Wave Propagation and diffraction*. 2024. P. 429–438 (in Russ.). <https://doi.org/10.24412/2304-0297-2024-1-429-440>
9. Falileev V.Yu., Shatovkin R.R. Analysis of existing automated protection against drones. *Vozdushno-kosmicheskie sily. Teoriya i praktika = Aerospace Forces. Theory and Practice*. 2020;14:130–140 (in Russ.). <https://www.elibrary.ru/aypwgc>
10. Konik A.A., Klimov I.M. On the issue of means and methods of active protection against attacks by unmanned aircraft in the context of a special military operation. *Problemy pravookhranitel'noi deyatel'nosti = Problems of Law-Enforcement Activity*. 2024;4(58):65–69 (in Russ.). <https://www.elibrary.ru/kmijke>
11. Makarenko S.I., Timoshenko A.V. Counter Unmanned Aerial Vehicles. Part 2. Rocket and Artillery Fire, Physical Interception. *Sistemy upravleniya, svyazi i bezopasnosti = Systems of Control, Communication and Security*. 2020;1:147–197 (in Russ.).
12. Tikshaev V.N., Barvinenko V.V. The problem of combating unmanned aerial vehicles and possible solutions. *Voennaya mysl' = Military Thought*. 2021;1:125–131 (in Russ.).
13. Moiseev N.V. Prospective variants of BVS lesions. In: *Modern Russian Science: Current Issues, Achievements and Innovations: collection of articles of the VI All-Russian Scientific and Practical Conference*. Penza: Nauka i Prosveshchenie; 2023. P. 106–109 (in Russ.).
14. Berdymuratov D.B. Analysis of UAV identification systems and database development. *UNIVERSUM: Technical Sciences*. 2024;10(127):4–8 (in Russ.).
15. Khrapkov D.S., Romashov V.A. Unmanned aerial vehicles as modern means of conducting combat operations. *Nauchnaya mysl' = Scientific thought*. 2021;15(1-1(39)):69–71 (in Russ.). <https://www.elibrary.ru/ootziq>

Об авторах

Коник Алексей Алексеевич, к.пед.н., доцент, заместитель начальника кафедры тактико-специальной подготовки, ФГКОУ ВО «Белгородский юридический институт Министерства внутренних дел Российской Федерации имени И.Д. Путилина» (308024, Россия, Белгород, ул. Горького, д. 71). E-mail: 89205666067@mail.ru. SPIN-код РИНЦ 9074-3701, <https://orcid.org/0000-0003-4563-3509>

Афонин Роман Евгеньевич, старший преподаватель, кафедра физической подготовки, ФГКОУ ВО «Белгородский юридический институт Министерства внутренних дел Российской Федерации имени И.Д. Путилина» (308024, Россия, Белгород, ул. Горького, д. 71). E-mail: afonin.roman@mail.ru. SPIN-код РИНЦ 4015-0302, <https://orcid.org/0009-0000-6383-3913>

Климов Иван Михайлович, начальник отдела развития средств противодействия функционированию беспилотных аппаратов, Центр развития беспилотных аппаратов и средств противодействия их функционированию, Научно-исследовательский институт специальной техники, ФКУ «Научно-производственное объединение «Специальная техника и связь», Министерство внутренних дел Российской Федерации» (ЦРБАиСПиФ НИИСТ ФКУ НПО «СТИС» МВД России) (111024, Россия, Москва, ул. Пруд Ключики, д. 2). E-mail: klimov.ivan700@gmail.com. <https://orcid.org/0009-0006-9824-8224>

Зинченко Дмитрий Сергеевич, старший научный сотрудник, отдел развития средств противодействия функционированию беспилотных аппаратов, Центр развития беспилотных аппаратов и средств противодействия их функционированию, Научно-исследовательский институт специальной техники, ФКУ «Научно-производственное объединение «Специальная техника и связь», Министерство внутренних дел Российской Федерации» (ЦРБАиСПиФ НИИСТ ФКУ НПО «СТИС» МВД России) (111024, Россия, Москва, ул. Пруд Ключики, д. 2). E-mail: dmitriy.zinchenko.1998@mail.ru. SPIN-код РИНЦ 9816-4245, <https://orcid.org/0009-0001-0948-8239>

About the Authors

Alexey A. Konik, Cand. Sci. (Ped.), Associate Professor, Deputy Head of the Department of Tactical and Special Training, I.D. Putilin Belgorod Law Institute of the Ministry of Internal Affairs of Russia (71, Gor'kogo ul., Belgorod, 308024 Russia). E-mail: 89205666067@mail.ru. RSCI SPIN-code 9074-3701, <https://orcid.org/0000-0003-4563-3509>

Roman E. Afonin, Senior Lecturer, Department of Physical Training, I.D. Putilin Belgorod Law Institute of the Ministry of Internal Affairs of Russia (71, Gor'kogo ul., Belgorod, 308024 Russia). E-mail: afonin.roman@mail.ru. RSCI SPIN-code 4015-0302, <https://orcid.org/0009-0000-6383-3913>

Ivan M. Klimov, Head of the Department for the Development of Means to Counter the Functioning of Unmanned Vehicles, Scientific and Production Association "Special Equipment and Communications," Ministry of Internal Affairs of the Russian Federation (2, Prud Klyuchiki ul., Moscow, 111024 Russia). E-mail: klimov.ivan700@gmail.com. <https://orcid.org/0009-0006-9824-8224>

Dmitry S. Zinchenko, Senior Researcher, Department for the Development of Means to Counter the Functioning of Unmanned Vehicles, Scientific and Production Association "Special Equipment and Communications," Ministry of Internal Affairs of the Russian Federation (2, Prud Klyuchiki ul., Moscow, 111024 Russia). E-mail: dmitriy.zinchenko.1998@mail.ru. RSCI SPIN-code 9816-4245, <https://orcid.org/0009-0001-0948-8239>

УДК 621.396.96

<https://doi.org/10.32362/2500-316X-2025-13-3-73-83>

EDN XDZDDH



НАУЧНАЯ СТАТЬЯ

Об эквивалентности характеристик и «зеркальности» построения традиционных и ММО радиолокационных станций при параллельном обзоре пространства на основе антенных решеток

Б.М. Вовшин^{1, 2, @},
А.А. Пушков¹,
Е.М. Халтурина¹

¹ НПО «Алмаз», Москва, 127411 Россия

² МИРЭА – Российский технологический университет, Москва, 119454 Россия

@ Автор для переписки, e-mail: boris@eleron.net

• Поступила: 30.07.2024 • Доработана: 13.02.2025 • Принята к опубликованию: 31.03.2025

Резюме

Цели. В последние годы в теории и практике радиолокации все больше внимания уделяется вопросам создания ММО (англ., «много входов – много выходов») радиолокационных станций (РЛС), обладающих рядом достоинств перед традиционными РЛС с фазированными антенными решетками. К этим достоинствам следует отнести возможности гибкого обзора пространства, адаптации к меняющейся сигнально-помеховой обстановке и т.д. Технология ММО в радиолокации требует излучения зондирующих сигналов в виде когерентной системы ортогональных сигналов, каждый из которых возбуждает собственный излучатель передающей антенной решетки (АР). Вследствие этого одновременно «освещается» заданная зона поиска цели. Пространственно-временная обработка (ПВО) «собирает» сигналы со всех направлений в облученной зоне на выходе приемника. В связи с этим актуальной является задача поиска оптимальной структуры ПВО в ММО РЛС по сравнению с традиционным подходом. Цель работы – синтез структуры ПВО при одноканальном приеме в ММО РЛС и сравнение полученного построения и характеристик с аналогичными в традиционных РЛС параллельного обзора на основе многолучевой приемной АР.

Методы. Используются методы и принципы теории многолучевых антенн с синтезированной апертурой и методы синтеза оптимальных по критерию Неймана – Пирсона обнаружителей на основе отношения правдоподобия.

Результаты. Для ММО РЛС с АР на передачу и одиночной слабонаправленной антенной на прием синтезирована разделяющаяся ПВО, формирующая оптимальную предпороговую статистику (ППС) обнаружителя на фоне белого гауссова шума. Проведено сравнение полученной ППС с аналогичной ППС в традиционной РЛС параллельного обзора пространства, имеющей «зеркальное» построение.

Выводы. Доказано, что в режиме параллельного поиска цели в одинаковых пространственных секторах показатели качества обнаружения у сравниваемых РЛС при «зеркальном» построении эквивалентны.

Ключевые слова: МИМО РЛС, параллельный обзор пространства, пространственно-временная обработка, алгоритм БПФ, многолучевая антенная решетка, предпороговая статистика

Для цитирования: Вовшин Б.М., Пушков А.А., Халтурина Е.М. Об эквивалентности характеристик и «зеркальности» построения традиционных и МИМО радиолокационных станций при параллельном обзоре пространства на основе антенных решеток. *Russian Technological Journal*. 2025;13(3):73–83. <https://doi.org/10.32362/2500-316X-2025-13-3-73-83>, <https://www.elibrary.ru/XDZDDH>

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

Авторы заявляют об отсутствии конфликта интересов.

RESEARCH ARTICLE

On the equivalence of characteristics and specularity in the construction of traditional and MIMO radars with a parallel view of space based on antenna arrays

Boris M. Vovshin ^{1, 2, @},
Alexander A. Pushkov ¹,
Elizaveta M. Khalturina ¹

¹ NPO ALMAZ, Moscow, 127411 Russia

² MIREA – Russian Technological University, Moscow, 119454 Russia

@ Corresponding author, e-mail: boris@eleron.net

• Submitted: 30.07.2024 • Revised: 13.02.2025 • Accepted: 31.03.2025

Abstract

Objectives. In recent years, more and more attention has been paid in radar theory and practice to the development of multiple-input and multiple-output (MIMO) radar, which offers a number of advantages over traditional radar based on phased antenna arrays (PAAs). These include the possibility to flexibly view space and adapt to a changing signal-interference environment, etc. MIMO technology used in radar requires the emission of a probe signal in the form of a coherent system of orthogonal signals, each of which triggers its own emitter in the transmitting antenna array (AA). As a result, the specified target search area is simultaneously illuminated. Specific spatiotemporal processing (SSP) is used to collect signals from all directions in the irradiated zone at the receiver output. In this regard, the task of finding an SSP structure in MIMO radar that is optimal compared to the traditional approach becomes urgent. The study set out to synthesize the structure of SSP with single-channel reception in MIMO radar and compare the obtained structure and characteristics with those similar in traditional parallel-view radars based on multipath receiving radar.

Methods. The study is based on methods and principles of the theory of multibeam synthesized aperture antennas and methods for the synthesis of optimal Neiman–Pearson detectors based on the likelihood ratio.

Results. For a MIMO radar with AA for transmission and reception provided by a single weakly directional antenna, a split SSP was synthesized to form optimal pre-threshold statistics (PTS) of the detector against a background

of white Gaussian noise. The obtained PTS is compared with a similar PTS in a traditional parallel space survey radar with a mirror structure.

Conclusions. It is shown that the detection quality indicators of the compared radars in the mirror construction are equivalent in the mode of parallel target search in the same spatial sectors.

Keywords: MIMO radar, parallel space survey, space-time processing, FFT algorithm, multipath antenna array, pre-threshold statistics

For citation: Vovshin B.M., Pushkov A.A., Khalturina E.M. On the equivalence of characteristics and specularity in the construction of traditional and MIMO radars with a parallel view of space based on antenna arrays. *Russian Technological Journal*. 2025;13(3):73–83. <https://doi.org/10.32362/2500-316X-2025-13-3-73-83>, <https://www.elibrary.ru/XDZDDH>

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

The authors declare no conflicts of interest.

ВВЕДЕНИЕ

В последние годы в теории и практике радиолокации значительное место занимают исследования так называемых МИМО (от англ. Multiple-Input, Multiple-Output – «много входов – много выходов») радиолокационных станций (РЛС). Интерес к ним возник в связи с открывающимися возможностями снятия ограничений, свойственных традиционным РЛС с фазированными антенными решетками (ФАР) при наблюдении целей. Ожидается, что технология МИМО может оказаться столь же революционной, как в свое время электронное сканирование, которое заменило механическое в антенной технике, придав новые свойства радиолокационным средствам [1, 2].

Первоначально в идею создания МИМО РЛС, по-видимому, было заложено известное свойство РЛС в режиме обзора: отношение сигнал/шум (ОСШ) на входе приемника и соответственно показатели качества обнаружения (ПКО) практически не зависят от ширины луча на передачу $\Delta\theta_{0.5\text{т}}$ при заданных секторе $\Delta\theta_{0.5\text{п}}$ и времени $t_{0.5\text{п}}$ обзора. Данное утверждение обусловлено тем, что снижение коэффициента направленного действия антенны на передачу, а, следовательно, и уменьшение уровня сигнала на цели в РЛС можно скомпенсировать увеличением времени наблюдения. Поясним этот тезис.

За время нахождения цели в луче шириной $\Delta\theta_{0.5}$ размер накопленной пачки отраженных сигналов с периодом T_0 составит $Q = (t_{0.5\text{п}}\Delta\theta_{0.5})/(T_0\Delta\theta_{0.5\text{п}})$. Отсюда следует, что при повышении темпа обзора пропорционально сокращению $t_{0.5\text{п}}$ будет снижаться значение Q , которое при $T_0 = \text{const}$ можно увеличить только за счет расширения луча $\Delta\theta_{0.5}$. Если ширину луча согласовать с сектором обзора $\Delta\theta_{0.5} = \Delta\theta_{0.5\text{п}}$, то эхо-сигналы от всех целей с априори неизвестными угловыми координатами внутри зоны обзора можно «собрать» при использовании приемной многолучевой антенной решетки (МАР). Такой вид обзора получил название «параллельный», и при его применении число когерентно накапливаемых импульсов Q

ограничено только временным интервалом корреляции самой цели $t_{\text{ц.кор}}$ [3, 4]. Традиционные РЛС с ФАР параллельного обзора на практике имеют следующие известные недостатки [5, 6]:

1. Широкая диаграмма направленности (ДН) на передачу, равная $\Delta\theta_{0.5\text{п}}$, обычно реализуется слабо-направленной антенной (СНА), а, следовательно, при заданной мощности излучения она должна обладать повышенной электрической прочностью.
2. При выполнении условия ортогональности число формируемых МАР лучей не может превышать число излучателей, а для перекрытия зоны лимитируется шаг d между ними. Оба этих фактора определяют разрешающую способность РЛС по угловым координатам.
3. Также как и любая ФАР, МАР обладает дисперсионными свойствами, накладывающими ограничения на широкополосность используемых зондирующих сигналов (ЗС) [4].

В ряде работ [7–9] отмечалось, что указанные недостатки и ограничения можно скомпенсировать или вообще снять при использовании в РЛС технологии МИМО. Ее суть заключается в следующем. Передающая M -элементная антенная решетка (АР) излучает M -компонентную систему взаимноортогональных когерентных ЗС. Ширина парциальных ДН этой АР $\Delta\theta_{\text{эл}}$ должна быть согласована с $\Delta\theta_{0.5\text{п}}$. В свою очередь ортогональность ЗС позволяет считать, что после отражения от цели суперпозиция принятых эхо-сигналов может быть разделена на M независимых каналов с некоррелированными

$$\overline{n_i n_j} = \begin{cases} \sigma_{\text{ш}0}^2 & \text{при } i = j, \\ 0 & \text{при } i \neq j, \end{cases} \quad i, j \in M, \quad \text{где } \sigma_{\text{ш}0}^2 -$$

дисперсия шума, для простоты принятая одинаковой во всех каналах, черта сверху – знак усреднения.

Для МИМО РЛС параллельный обзор пространства вследствие фактического отсутствия интерференции ЗС на цели является единственно возможным. Как и в традиционной РЛС, в режиме поиска здесь также предполагается многолучевое

диаграммообразование на прием. Реализацию параллельного обзора в МИМО РЛС будем рассматривать в совокупности с пространственно-временной обработкой (ПВО), формирующей предпороговую статистику (ППС), оптимальную по критерию Неймана – Пирсона [10].

Такой комплексный подход позволяет провести корректное сравнение МИМО РЛС с РЛС традиционного построения на уровне ППС, которая обеспечивает эквивалентные ПКО в режиме поиска целей. После решения этой задачи можно определить условия и принципы построения МИМО РЛС, имеющие преимущества по сравнению с традиционными РЛС, несмотря на возможные сложности практической реализации.

Цель статьи – синтезировать структуру ПВО в МИМО РЛС и сравнить ее построение с аналогичной обработкой в традиционной РЛС параллельного обзора при условии достижения одинаковых ПКО.

ПАРАЛЛЕЛЬНЫЙ ОБЗОР И ОБНАРУЖЕНИЕ ЦЕЛЕЙ В ТРАДИЦИОННЫХ РЛС С МАР

Рассмотрим принципы построения традиционных РЛС с МАР при параллельном обзоре заданного углового сектора $\Delta\theta_{обз}$. Наиболее общим случаем является применение СНА на передачу и МАР на прием, как показано на рис. 1. Без потери общности

будем полагать, что ширина ДН СНА $\Delta\theta_{СНА} = \Delta\theta_{обз}$, а число независимых лучей линейной МАР равно числу излучателей N , расставленных с эквидистантным

шагом $d_r \leq \frac{\lambda_0}{1 + |\sin(\Delta\theta_{обз}/2)|}$, где λ_0 – рабочая

длина волны. Кроме того, на приемную МАР наложим условие узкополосности возбуждающего сигнала $u(t)$ с шириной полосы Δf_c [2, 8]:

$$T_a = (N-1)d_r \sin\left(\frac{\Delta\theta_{обз}}{2}\right) \ll \frac{1}{\Delta f_c}, \quad (1)$$

где T_a – время заполнения апертуры МАР импульсом с эквивалентной длительностью $\tau_{и.э.} = 1/\Delta f_c$.

Определим комплексную огибающую ЗС $\dot{U}_{ref}(t)$ после отражения от точечной цели с трехкомпонентным вектором информационных параметров $\mathbf{k} = \{R_{ц}, V_{ц}, \theta_{ц}\}$, где $R_{ц}$, $V_{ц}$, $\theta_{ц}$ – дальность, радиальная скорость и угловая координата цели. На выходе МАР она имеет вид:

$$\begin{aligned} \dot{U}_{ref}(t, \mathbf{k}) &= F_{tr}(\theta_{ц}) \dot{U}_{tr}(\gamma_{ц}t - \tau_{ц}) e^{j2\pi f_0 \gamma_{ц}t} = \\ &= U_{tr}(t' - \tau_{ц}) e^{j2\pi f_0 t'}, \end{aligned} \quad (2)$$

где $F_{tr}(\theta_{ц})$ – уровень ДН СНА в направлении $\theta_{ц}$, $\tau_{ц} = 2R_{ц}/c$ – время задержки сигнала, c – скорость света, $\gamma_{ц} = 1 \pm 2V_{ц}/c$ – доплеровский коэффициент

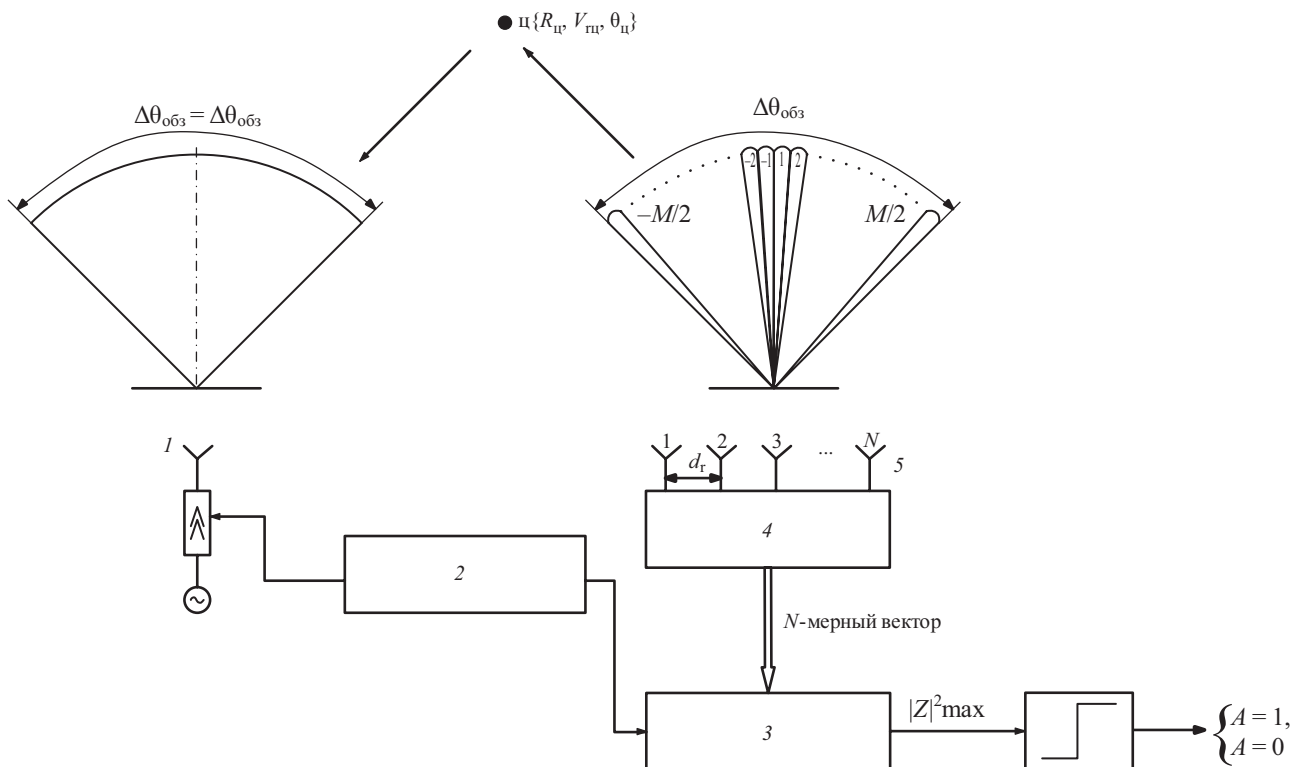


Рис. 1. Традиционная РЛС с параллельным обзором пространства:

1 – СНА, 2 – формирователь ЗС, 3 – временная обработка (формирование ППС), 4 – пространственная обработка – диаграммообразующая схема (ДОС), 5 – МАР, $\mathbf{ц}\{R_{ц}, V_{ц}, \theta_{ц}\}$ – цель, A – порог, по которому принимается решение о наличии или отсутствии сигнала

изменения масштаба времени, для которого $F_d = 2V_r/\lambda_0$ – частота Доплера, $t' = \gamma_{\text{ц}} t$.

В дальнейшем можно полагать $F_{\text{тр}}(\theta_{\text{ц}}) \approx \text{const}$, а также то, что в силу условия (1) в одноцелевой ситуации комплексная огибающая принимаемого сигнала факторизуется на скалярную временную функцию и N -мерный вектор $\mathbf{b}(\theta_{\text{ц}})$ пространственных фаз:

$$U_r(t' - \theta_{\text{ц}}) = U_{\text{тр}}(t' - \tau_{\text{ц}}) \mathbf{b}(\theta_{\text{ц}}),$$

$$\mathbf{b}(\theta_{\text{ц}}) = \exp \left\{ j \frac{2\pi d_r}{\lambda_0} (n-1) \sin \theta_{\text{ц}} \right\}_{n=1}^N. \quad (3)$$

Рассмотрим задачу многоканального обнаружения цели в классической постановке приема смеси $y(t) = A \dot{U}_r(t) + n(t)$, где $A = \{1, 0\}$ в зависимости от наличия или отсутствия полезного сигнала, а $n(t)$ – белый гауссовский шум (БГШ). В отсутствие коррелированных внешних помех при одинаковой интенсивности БГШ в каналах МАР: $\sigma_i^2 = \sigma_j^2 = \sigma_0^2$ ($i, j \in M$), а $(N \times N)$ -мерная корреляционная матрица-функция (КМФ) помех имеет вид:

$$\Phi(t - s') = N_0 \mathbf{I} \delta(t - s), \quad (4)$$

где $\mathbf{I}$ – единичная диагональная матрица, N_0 – спектральная плотность мощности шумов, δ – дельта-функция Дирака.

В этих условиях оптимальная по критерию Неймана – Пирсона ПВО сводится к вычислению ППС в форме квадрата модуля весового интеграла:

$$\xi = |Z|^2 = \frac{1}{N_0} \left| \int \dot{Y}_{\Sigma}(t') \dot{U}_{\text{гес}}^*(t') dt' \right|^2, \quad (5.1)$$

$$\dot{Y}_{\Sigma}(t') = \mathbf{Y}^T(t', \theta_{\text{ц}}) \mathbf{b}^*(\theta_{\text{ц}}), \quad (5.2)$$

где $\mathbf{Y}$ – входной вектор.

В формулах (5.1) и (5.2) знаки $(*)$ и $(^T)$ означают комплексное сопряжение и транспонирование соответственно. Выражение (5.2) определяет комплексную амплитуду на выходе i -го «вторичного» канала МАР, сфазированного в ожидаемое направление $\theta_{\text{ц}i}$ ($i \in 1, N$). Как уже отмечалось, N лучей, формируемых МАР, должны перекрыть весь заданный сектор обзора $\Delta\theta_{\text{обз}}$. Практически формирование ортогональных лучей без потерь удобно реализовать на основе алгоритма быстрого преобразования Фурье (БПФ), если $N = 2^q$, где q – целое число (в аналоговой форме – матрица Батлера [1]). Алгоритм БПФ преобразует отсчеты N -мерного вектора входного сигнала (3) («первичные» каналы МАР) в вектор из N ортогональных лучей («вторичных» каналов (ДН)) с использованием $(N \times N)$ матрицы

преобразования $\mathbf{W}(\theta) = \{w_{ni}\}_{n,i}^{N,N}$, где $w_{ni} = \exp(j2\pi n/N)$; n, i – номера «первичных» и «вторичных» каналов МАР соответственно.

После реализации БПФ формулу (5.1) можно рассматривать как выражение для ППС в любом из N «вторичных» каналов, если в ней полагать:

$$\dot{Y}_{\Sigma}(t') = \mathbf{Y}^T(t', \theta_{\text{ц}}) \mathbf{W}^*(\theta_{\text{ц}i}),$$

где $\theta_{\text{ц}i} = \arcsin[(i-1/2)\lambda_0/(N-1)d_r]$ (см. рис. 1).

Вышеизложенные принципы параллельного обзора являются общепринятыми в традиционных РЛС с МАР. Они обладают следующими особенностями, в дальнейшем важными с точки зрения сравнения с МИМО РЛС. К ним относятся:

1. Цель одномоментно облучается единственным когерентным ЗС $\dot{U}_{\text{тр}}(t)$ на несущей частоте f_0 с заданной средней мощностью.
2. Нарастивание числа элементов в СНА, например, выполненной в виде малоразмерной АР, часто принципиально невозможно, т.к. сопровождается сужением $\Delta\theta_{\text{СНА}}$, что не позволит эффективно «осветить» зону поиска $\Delta\theta_{\text{обз}}$.
3. Разрешающая способность лучей МАР определяется геометрическим размером ее апертуры $L_{\text{МАР}} = (N-1)d_r$. При этом согласно п. 2 число ортогональных лучей ограничено числом первичных каналов N , а увлечение шага d_r лимитируется шириной заданного сектора обзора $\Delta\theta_{\text{обз}}$. Эти факторы не позволяют сузить лучи МАР, а также увеличить их число.
4. Построение традиционной РЛС с параллельным обзором пространства (рис. 1) реализует факторизованное представление ППС (3), что позволяет разделить последовательную ПВО на пространственную (ДН) и частотно-временную (функцию Вудворта) обработки в таком порядке.

Отметим, что обратный порядок нецелесообразен, т.к. потребует до пространственного накопления реализовать одинаковое временное накопление в каждом первичном канале МАР, которое удобно осуществить единственный раз.

Вышеуказанные особенности и ограничения редко встречаются и в основном снимаются в МИМО РЛС благодаря увеличению размерности задачи. Вместо единого ЗС в пространство одновременно излучается N ортогональных, но когерентных сигналов, что позволяет для радиолокационного наблюдения использовать дополнительную степень свободы.

Перейдем к анализу отличий МИМО РЛС от традиционных РЛС при условии достижения эквивалентных ППС и характеристик обнаружения.

ПАРАЛЛЕЛЬНЫЙ ОБЗОР И ОБНАРУЖЕНИЕ ЦЕЛЕЙ В МИМО РЛС

Начнем с основной отличительной особенности МИМО РЛС, которая заключается в «освещении» зоны обзора системой ортогональных ЗС, возбуждающих АР с шагом d_{tr} , показанной в левой части рис. 2. Решетка в целом излучает векторный сигнал с комплексной огибающей $\dot{U}_{tr}(t) = \{\dot{U}_{tr m}\}_{m=1}^M$. Условие ортогональности этих компонент должно выполняться для всех направлений θ внутри сектора $\Delta\theta_{обз}$. Тогда нормированный коэффициент корреляции между p -й и q -й компонентами будет описываться выражением

$$\rho_{pq} = \frac{\int \dot{U}_{ptr} \left[t + (p-1) \frac{d_{tr}}{c} \sin \theta \right] dt \int \dot{U}_{qtr}^* \left[t + (q-1) \frac{d_{tr}}{c} \sin \theta \right] dt}{\left[\int |\dot{U}_{ptr}(t)|^2 dt \int |\dot{U}_{qtr}(t)|^2 dt \right]^{1/2}} \quad (6)$$

Комплексная огибающая суммарного сигнала, достигающего цель, имеет вид

$$\dot{U}_{tr}(t, \theta_{ц}) = \sum_{m=1}^M \dot{U}_{mtr} \left[t + (m-1) \frac{d_{tr}}{c} \right] \exp \left[j \frac{2\pi d_{tr}}{\lambda_0} (m-1) \sin \theta_{ц} \right] \quad (7)$$

Как и в предыдущем случае, будем полагать ее узкополосной, удовлетворяющей условию (1). Заметим, что для МИМО РЛС традиционные ограничения на широкополосность АР практически сняты вследствие работы на ортогональных сигналах, например, разнесенных по несущим частотам на Δf_0 [11, 12]. При приеме они могут быть разделены по полосам, а, следовательно, их интерференцией на цели можно пренебречь. Для корректности сравнения будем полагать, что полоса сигнала, излучаемого АР МИМО РЛС такая же, как и в традиционной РЛС и равна $\Delta f_c = (N-1)\Delta f_0$ при условии $\Delta f_c \ll \Delta f_0$.

Среднюю мощность векторного сигнала (7), достигшего цели, можно представить квадратичной эрмитовой формой:

$$P_{tr} = \sum_{p=1}^M \sum_{q=1}^M \overline{\dot{U}_p \dot{U}_q} \exp \left[j \frac{2\pi d_{tr}}{\lambda_0} (p-q) \sin \theta_{ц} \right] = \alpha^T(\theta_{ц}) I \alpha^*(\theta_{ц}). \quad (8)$$

Будем полагать, что мощность (8) такая же, как и в предыдущем варианте, т.е. $P_{tr} = P_{ср} G_{СНА}$, где $P_{ср}$ – мощность передатчика на входе СНА традиционной РЛС, $G_{СНА}$ – коэффициент направленного действия СНА.

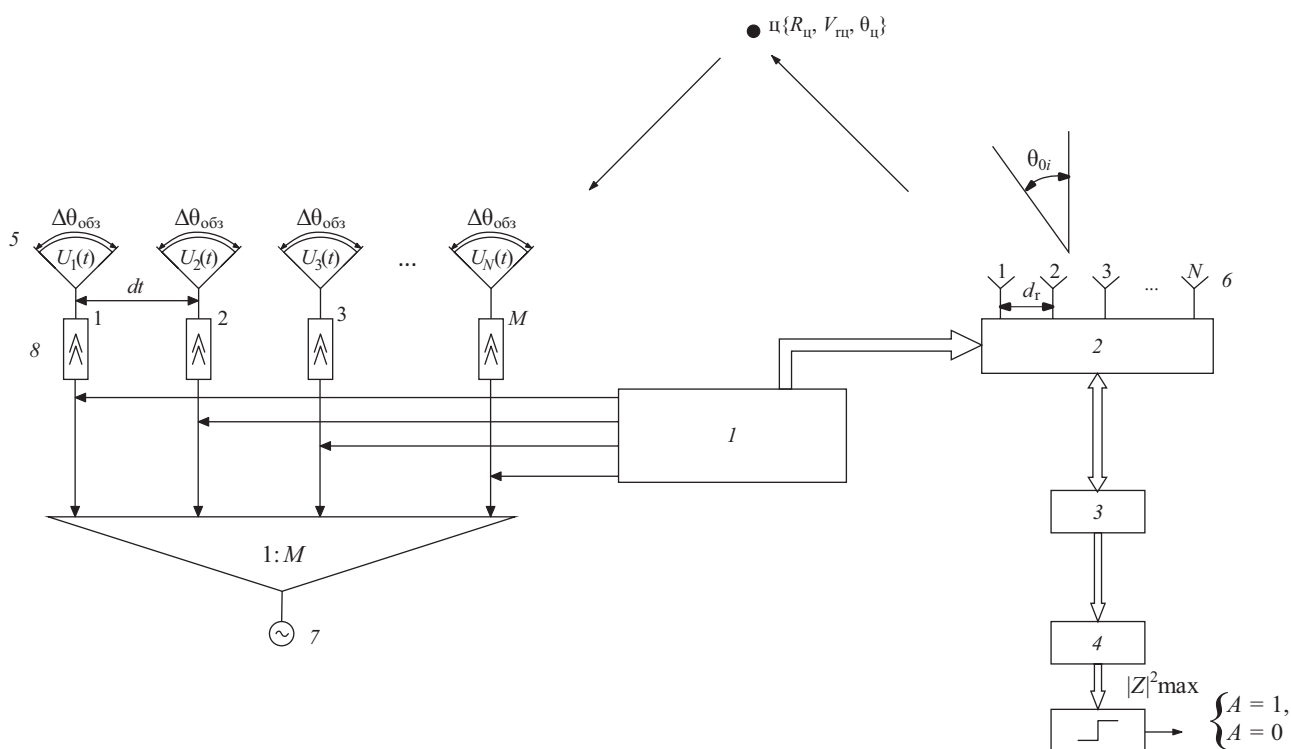


Рис. 2. МИМО РЛС с параллельным обзором пространства:

1 – формирователь ортогональных ЗС, 2 – временная обработка, 3 – пространственная обработка, 4 – формирователь ППС, 5 – передающая АР, 6 – приемная АР, 7 – задающий генератор (ЗГ), 8 – усилитель мощности (УМ)

Для точечной цели с вектором информационных параметров $\kappa = \{t', V_{\text{гц}}, \theta_{\text{ц}}\}$ комплексная амплитуда отраженного сигнала имеет вид:

$$\dot{U}_{\text{ref}}(t', \tau_{\text{ц}}, \theta_{\text{ц}}) = F_{\text{тр}}(\theta_{\text{ц}}) \times \sum_{m=1}^M \dot{U}_{\text{mtr}}(t' - \tau_{\text{ц}}) \exp \left[j \frac{2\pi d_{\text{тр}}}{\lambda_0} (m-1) \sin \theta_{\text{ц}} \right]. \quad (9)$$

Будем считать, что все излучатели передающей АР имеют одинаковые слабонаправленные ДН, согласованные с сектором обзора, т.е. $F_{\text{тр}}(\theta_{\text{ц}}) = \text{const}$ для всех $\theta_{\text{ц}} \in \Delta\theta_{\text{обз}}$. Тогда выражение (9) можно рассматривать как M -мерный вектор $\dot{U}_{\text{ref}}(t) = \left\{ \dot{U}_{\text{mtr}}(t, \tau_{\text{ц}}) e^{j\alpha_m(\theta_{\text{ц}})} \right\}_{m=1}^M$, который совпадает с вектором, возбуждающим одиночную приемную СНА в правой части рис. 2, $\alpha_m(\theta_{\text{ц}})$ – фазовый набег m -го парциального сигнала.

Перейдем к синтезу ПВО на выходе этой СНА, учитывая, что принятый сигнал $\dot{U}_{\text{rec}}(t) = \dot{U}_{\text{ref}}(t)$ факторизуется на M временных и пространственных множителей. Поэтому, как показано на рис. 3, разделим его на M независимых каналов:

$$\dot{U}_{\text{rec}}(t', \tau_{\text{ц}}, \theta_{\text{ц}}) = \mathbf{E}^T \otimes \dot{U}_{\text{ref}}(t', \tau_{\text{ц}}, \theta_{\text{ц}}) = \left\{ \dot{U}_{\text{mtr}}(t' - \tau_{\text{ц}}) e^{j\alpha_m(\theta_{\text{ц}})} \right\}_{m=1}^M, \quad (10)$$

где $\mathbf{E} = \{1\}_{m=1}^M$ – единичный M -мерный вектор, $\otimes$ – знак кронекеровского произведения.

С точки зрения шумовых характеристик эти каналы независимы, их КМФ описывается соотношением (4), справедливым и в данном случае. На фоне БГШ оптимальная временная обработка реализуется набором согласованных фильтров (СФ) для каждого парциального сигнала $U_m(t)$ в приемном канале. Однако, в отличие от (5), в одноцелевой ситуации для ожидаемого направления $\theta_{\text{ц}}$ в выражении для ППС временной множитель сигнала является вектором, а не скаляром, и имеет вид:

$$\begin{aligned} \xi = |Z|^2 &= \mathbf{Y}^T(t', \theta_{\text{ц}}) \mathbf{U}^*(t', \theta_{\text{ц}}) = \\ &= \frac{1}{N_0} \left| \sum_{m=1}^M \dot{U}_m(t', \theta_{\text{ц}}) \dot{U}_{0m}^*(t') e^{-j \frac{2\pi d_{\text{тр}}}{\lambda_0} (\sin \theta_{\text{ц}} - \sin \theta_i)} \right|^2 = \\ &= \frac{1}{N_0} \left| \sum_{m=1}^M Y_{m \text{ mf}}(0) e^{-j \frac{2\pi d_{\text{тр}}}{\lambda_0} (m-1) (\sin \theta_{\text{ц}} - \sin \theta_i)} \right|^2, \end{aligned} \quad (11)$$

где $\dot{U}_{0m}(t')$ – комплексная огибающая ожидаемого сигнала в m -м парциальном канале, $Y_{m \text{ mf}}(0)$ (mf – match filter, англ.) – результат временного накопления сигнала,

что соответствует максимуму амплитуды при $t' = 0$ на выходе СФ в m -м канале.

Важной особенностью схемы на рис. 3 является то, что временное накопление здесь предшествует пространственному, что указывает на обратный порядок по отношению к традиционной ПВО.

Если в (11) положить $t' \neq 0$, то выходной сигнал $\dot{Y}_{m \text{ mf}}(t')$ можно рассматривать как частотно-временное рассогласование по дальности и скорости, а формулу в целом – как многомерную функцию рассогласования в координатах «дальность-скорость-угол», аналогично [13].

Выражению (11) можно дать еще одну интерпретацию. Оно соответствует традиционной формуле множителя АР, в которой вектор $\dot{\mathbf{Y}}_{\text{mf}}(t')$ имеет вид зависящего от времени амплитудного распределения M -элементной эквидистантной АР с шагом $d_{\text{тр}}$. Если во всех каналах $Y_{\text{mf}}(0) = \text{const}$, то в момент $t' = 0$, соответствующий максимуму сигнала, это эквивалентное распределение будет равномерным. В теории антенн подобная трансформация одиночного приемного излучателя в АР, идентичную передающей (рис. 4), соответствует понятию «синтезированная апертура». Однако в большинстве публикаций, посвященных ММО РЛС, пользуются термином «виртуальная подрешетка/решетка» (ВПП/ВР) [7, 14, 15].

Из-за априорной неопределенности углового положения точечной цели ($\theta_{\text{ц}}$) после временного накопления на выходе ВПП требуется реализовать многолучевую ДОС. Как было показано выше, для формирования ортогональных лучей можно воспользоваться цифровым алгоритмом БПФ, и таким образом получить M «вторичных» приемных лучей:

$$\begin{aligned} |Z|^2 &= \max | \mathbf{Y}_{\text{mf}}^T(0) \mathbf{W}^*(\theta_{0i}) |^2, \\ \mathbf{W}(\theta_{0i}) &= \left\{ \exp \left(j \frac{2\pi m i}{M} \right) \right\}_{m,i=-\frac{M}{2}}^{\frac{M}{2}}. \end{aligned} \quad (12)$$

Из полученной структуры ММО РЛС, осуществляющей параллельный обзор пространства, следует вывод о «зеркальности» ее построения по сравнению с традиционной РЛС, как показано на рис. 4.

На рис. 4 в сравниваемых вариантах не показаны каналы доплеровской обработки на входах пороговых устройств. В обоих случаях системы доплеровской фильтрации идентичны и могут быть реализованы различными способами в виде череспериодной компенсации или набора фильтров, настроенных на ожидаемые радиальные скорости $V_{\text{гц}} = \frac{\lambda \Delta \Phi_0}{2 \Delta T}$,

где $\Delta \Phi_0$ – разность фаз между смежными импульсами пачки, ΔT – период повторения. Это означает, что в узкополосном случае ММО РЛС по информативному параметру «скорость» не имеет специфических

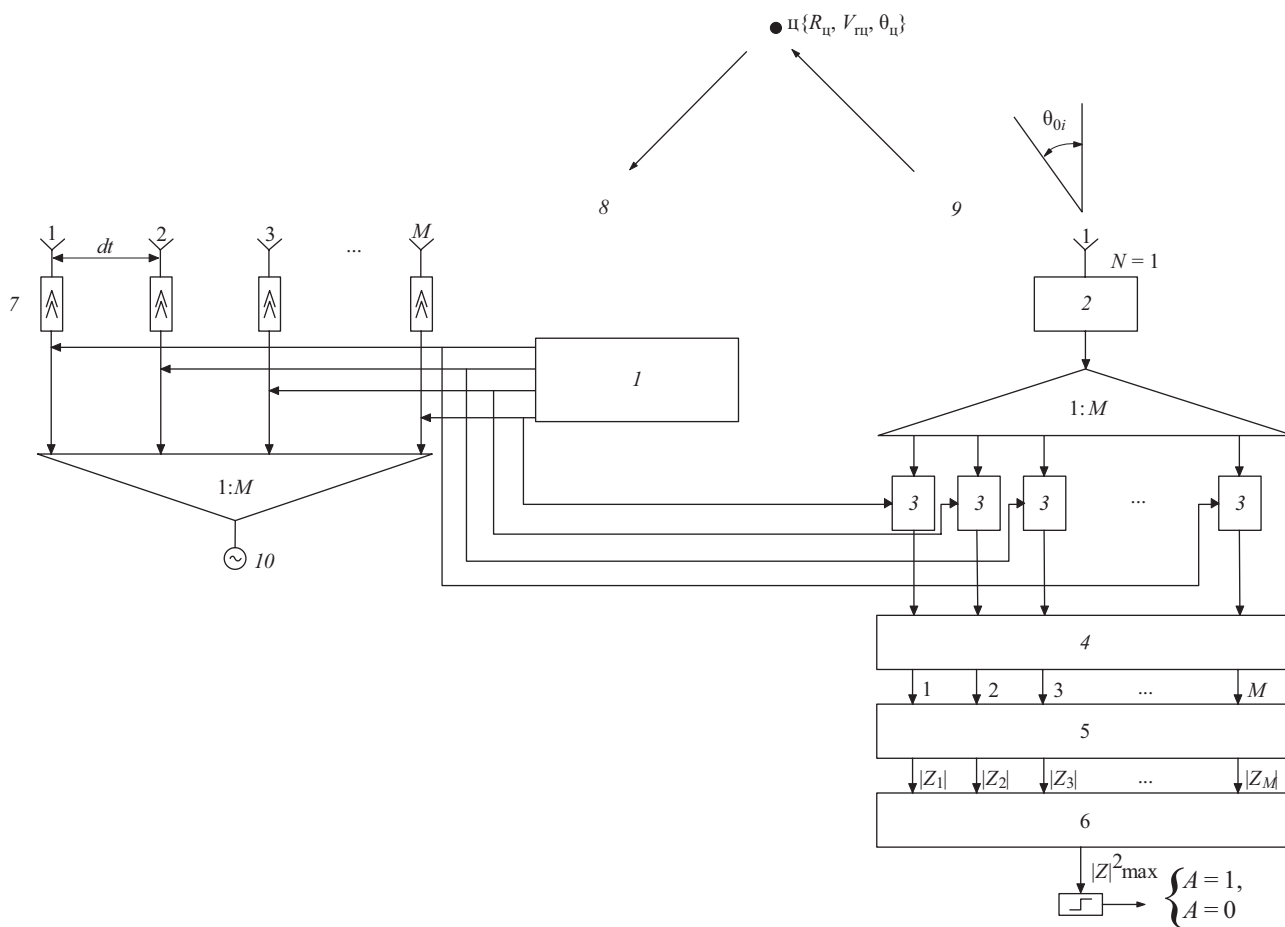


Рис. 3. Последовательность этапов ПВО в МИМО РЛС:
1 – формирователь ортогональных ЗС, 2 – маломощный усилитель (МШУ), 3 – СФ, 4 – ДОС (M-точечный БПФ),
5 – формирователь ППС, 6 – отбор максимума, 7 – УМ, 8 – передающая АР, 9 – приемная АР, 10 – ЗГ

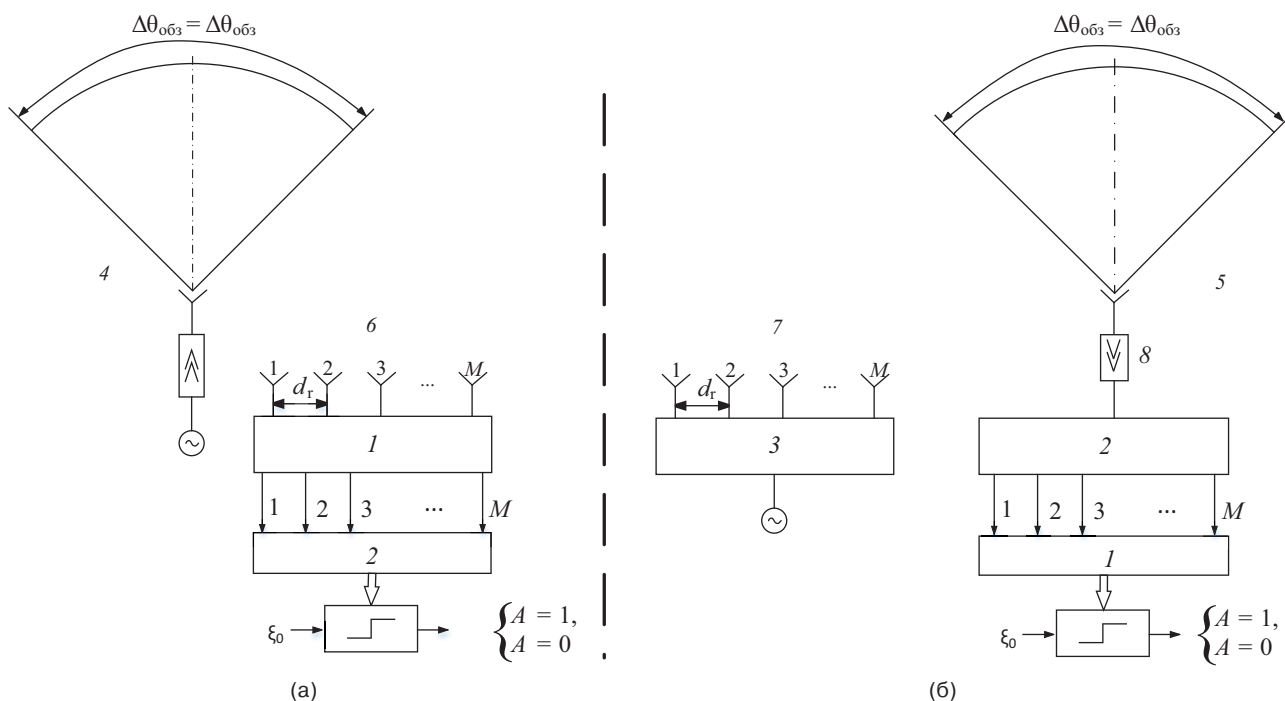


Рис. 4. Зеркальность построения традиционной (а) и МИМО (б) РЛС с параллельным обзором пространства:
1 – ДОС, 2 – временная обработка, 3 – формирователь ортогональных ЗС, 4 – передающая СНА,
5 – приемная СНА, 6 – приемная МАР, 7 – передающая АР, 8 – МШУ

особенностей. В выбранных условиях эквивалентность ППС обеих вариантов обеспечивает идентичность их статистических ПКО.

Для МИМО РЛС с АР на передачу и одноканальным построением (СНА) на прием можно выделить следующие основные особенности:

1. Цель одномоментно облучается вектором когерентных ортогональных ЗС $\vec{U}_{\text{тр}}(t)$, излучаемых элементами АР, например, на разных несущих частотах f_{0m} . Нарастивание числа излучателей АР не приводит к сужению сектора обзора $\Delta\theta_{\text{обз}}$ и при фиксированной мощности излучения снижает требования к электрической прочности тракта.
2. Разрешающая способность по угловым координатам в данной МИМО РЛС определяется только размером передающей АР $L_{\text{тр}} = (M - 1)d_{\text{тр}}$, в которую, благодаря первоначальной частотно-временной обработке, трансформируется приемная СНА, которая «превращается» в ВПР.
3. Также, как и в традиционном случае, при использовании одиночной СНА на прием, здесь ПВО факторизуется на временную и пространственную. Их последовательность не принципиальна, но из практических соображений оказывается обратной по сравнению с вариантом МАР.
4. Достоинства МИМО РЛС, указанные в п. 1, частично нивелируются определенными трудностями практической реализации передающей АР. К этим трудностям следует отнести необходимость формирования набора когерентных сигналов и сохранение их когерентности при разводке сверхвысокочастотных сигналов по каналам передающей АР.

Согласованная временная обработка в каналах ВПР проводится при более низком уровне полезного сигнала по сравнению с традиционным вариантом, где СФ «вторичных» каналов работают после когерентного пространственного накопления. Однако одинаковое значение ОСШ при сохранении линейности приводит к теоретически одинаковому качеству обнаружения.

ЗАКЛЮЧЕНИЕ

Полученные в статье результаты позволяют сделать следующие выводы:

1. Сравнение РЛС традиционного построения и одноканальной на прием МИМО РЛС в режиме параллельного поиска целей показало

эквивалентность их статистических ПКО в следующих условиях:

- одинаковое результирующее ОСШ и ППС в обнаружителе;
 - одинаковые широкополосность когерентных ЗС и линейность при приеме;
 - одинаковые сектора обзоров $\Delta\theta_{\text{обз}}$;
 - возможность разделения ПВО на частотно-временную и пространственную.
2. Теоретическая эквивалентность сравниваемых вариантов достигнута при «зеркальности» структурных схем их построения (см. рис. 4).

Традиционный вариант:

- СНА на передачу и M -элементная МАР на прием;
- пространственное накопление предшествует временному.

МИМО РЛС:

- M -элементная АР на передачу и СНА на прием;
- временное накопление предшествует пространственному.

Общее число каналов ПВО одинаково.

3. Основные достоинства МИМО РЛС по сравнению с РЛС традиционного построения проявляются, если в них одиночную СНА на прием заменить на многоканальную приемную АР. Этот более сложный случай рассматривается в следующей статье.

Вклад авторов

Б.М. Вовшин – общая постановка задачи, разработка принципов построения МИМО РЛС, разработка методики сравнения МИМО РЛС и традиционной РЛС параллельного обзора.

А.А. Пушков – разработка структурных схем построения, вывод аналитических соотношений, анализ результатов сравнения вариантов построения РЛС.

Е.М. Халтурина – разработка структурных схем построения, вывод аналитических соотношений, написание текста, оформление материала.

Authors' contributions

B.M. Vovshin – general statement of the problem, development of principles for constructing MIMO radars, development of a methodology for comparing MIMO radars and traditional parallel-survey radars.

A.A. Pushkov – development of structural design schemes, derivation of analytical ratios, analysis of the results of comparison of radar design options.

E.M. Khalturina – development of structural construction schemes, derivation of analytical ratios, writing text, design of material.

СПИСОК ЛИТЕРАТУРЫ

1. Черняк В.С. О новых и старых идеях в радиолокации: МИМО РЛС. *Успехи современной радиоэлектроники*. 2011;2:5–20.
2. Li J., Stoica P. *MIMO Radar Signal Processing*. New York: Wiley; 2009. 448 p.
3. *Проблемы антенной техники*; под ред. Л.Д. Бахраха и Д.И. Воскресенского. М.: Радио и связь; 1989. 368 с.
4. Mailloux R.J. *Phased Array Antenna Handbook*. 2nd Ed. London: Artech House; 2005. 479 p.
5. Черняк В.С. Обнаружение сигналов в МИМО РЛС. *Успехи современной радиоэлектроники*. 2014;7:35–48.
6. Вовшин Б.М. Безынерционный обзор пространства сверхширокополосными радиолокационными сигналами. *Антенны*. 2006;7(102):92–100.
7. Bergin J., Guerci J.R. *MIMO Radar: Theory and Application*. London: Artech House; 2018. 221 p.
8. Dorey J., Garnier G., Auvray G. RIAS Radar a impulsion et Antenne Synthetique. In: *Proceedings Colloque International sur le Radar*. Paris. April. 1989. P. 112–115 (in French).
9. Vovshin B.M. Parallel Surveillance Ultra-wide-band Radars with Orthogonal Ranging Signals. In: *Proceedings on International Radar Symposium (IRS-2007)*. Cologne, Germany. P. 461–466.
10. *Радиоэлектронные системы. Основы построения и теория*; под ред. Я.Д. Ширмана. М.: Радиотехника; 2007. 512 с.
11. Вовшин Б.М., Иммогеев И.Я. Влияние дисперсионных свойств фазированной антенной решетки на отношение сигнал/шум в РЛС с широкополосными сигналами. *Радиотехника*. 1985;7:74–92.
12. Калинин В.И., Чапурский В.В., Черепенин В.А. Суперразрешение в системах радиолокации и радиолографии на основе МИМО антенных решеток с рециркуляцией сигналов. *Радиотехника и электроника*. 2021;66(6):614–624. <https://doi.org/10.31857/S0033849421060139>
13. Чапурский В.В., Слукин Г.П., Филатов А.А., Коротеев Д.Е. Виртуальное МИМО РЛС и их сравнение на основе обобщенных функций неопределенности. *Успехи современной радиоэлектроники*. 2023;77(5):5–19. <https://doi.org/10.18127/j20700784-202305-01>
14. Brookner E. MIMO Radars and their Conventional Equivalents. In: *Proc. on IEEE International Radar conference*. May 10–15, 2015, Arlington, VA, USA: P. 918–924. <https://doi.org/10.1109/RADAR.2015.7131126>
15. Tebaldini S., Manzoni M., Ferro-Famil L., Banda F., Giudici D. FDM MIMO SAR Tomography. In: *EUSAR 2024 – 15th European Conference on Synthetic Aperture Radar*. Munich, Germany: 2024. P. 709–714.

REFERENCES

1. Chernyak V.S. About new and old ideas in radar: MIMO radars. *Uspekhi sovremennoi radioelektronik = Achievements of Modern Radioelectronics*. 2011;2:5–20 (in Russ.).
2. Li J., Stoica P. *MIMO Radar Signal Processing*. New York: Wiley; 2009. 448 p.
3. Bakhrakh L.D., Voskresenskii D.I. (Eds.) *Problemy antennoi tekhniki (Antenna Technology Problems)*. Moscow: Radio i svyaz'; 1989. 368 p. (in Russ.).
4. Mailloux R.J. *Phased Array Antenna Handbook*. 2nd Ed. London: Artech House; 2005. 479 p.
5. Chernyak V.S. Signal detection with MIMO radars. *Uspekhi sovremennoi radioelektronik = Achievements of Modern Radioelectronics*. 2014;7:35–48 (in Russ.).
6. Vovshin B.M. The inertialless air surveillance by ultra wideband radar signals. *Antenny*. 2006;7(102):92–100 (in Russ.).
7. Bergin J., Guerci J.R. *MIMO Radar: Theory and Application*. London: Artech House; 2018. 221 p.
8. Dorey J., Garnier G., Auvray G. RIAS Radar a impulsion et Antenne Synthetique. In: *Proceedings Colloque International sur le Radar*. Paris. April. 1989. P. 112–115 (in French).
9. Vovshin B.M. Parallel Surveillance Ultra-wide-band Radars with Orthogonal Ranging Signals. In: *Proceedings on International Radar Symposium (IRS-2007)*. Cologne, Germany. P. 461–466.
10. Shirman Ya.D. (Ed.). *Radioelektronnye sistemy. Osnovy postroeniya i teoriya (Radioelectronic systems. Fundamentals of Construction and Theory)*. Moscow: Radiotekhnika; 2007. 512 p. (in Russ.).
11. Vovshin B.M., Immoreev I.Y. Influence of dispersion properties phased array antenna on the signal to noise ratio in radars with broadband signals. *Radiotekhnika*. 1985;7:74–92 (in Russ.).
12. Kalinin V.I., Chapursky V.V., Cherepenin V.A. Super-Resolution of Radar and Radio Holography Systems Based on a MIMO Retrodirective Antenna Array. *J. Commun. Technol. Electron.* 2021;66(6):727–736. <https://doi.org/10.1134/S1064226921060139>
[Original Russian Text: Kalinin V.I., Chapursky V.V., Cherepenin V.A. Super-Resolution of Radar and Radio Holography Systems Based on a MIMO Retrodirective Antenna Array. *Radiotekhnika i elektronika*. 2021;66(6):614–624 (in Russ.). <https://doi.org/10.31857/S0033849421060139>]
13. Chapurskii V.V., Slukin G.P., Filatov A.A., Koroteev D.E. Virtual MIMO radars and their comparison based on generalized uncertainty functions. *Uspekhi sovremennoi radioelektroniki = Achievements of Modern Radioelectronics*. 2023;77(5):5–19 (in Russ.). <https://doi.org/10.18127/j20700784-202305-01>
14. Brookner E. MIMO Radars and their Conventional Equivalents. In: *Proc. on IEEE International Radar conference*. May 10–15, 2015, Arlington, VA, USA, P. 918–924. <https://doi.org/10.1109/RADAR.2015.7131126>
15. Tebaldini S., Manzoni M., Ferro-Famil L., Banda F., Giudici D. FDM MIMO SAR Tomography. In: *EUSAR 2024 – 15th European Conference on Synthetic Aperture Radar*. Munich Germany: 2024. P. 709–714.

Об авторах

Вовшин Борис Михайлович, д.т.н., профессор, кафедра радиоволновых процессов и технологий, Институт радиоэлектроники и информатики, ФГБОУ ВО «МИРЭА – Российский технологический университет» (119454, Россия, Москва, пр-т Вернадского, д. 78); главный научный сотрудник, ПАО «Научно-производственное объединение «Алмаз» имени академика А.А. Расплетина (ПАО «НПО «Алмаз») (127411, Россия, Москва, Дмитровское шоссе, д. 110). E-mail: boris@eleron.net. Scopus Author ID 7801355022, <https://orcid.org/0009-0003-1357-5866>

Пушков Александр Александрович, к.т.н., начальник научно-технического центра, ПАО «Научно-производственное объединение «Алмаз» имени академика А.А. Расплетина (ПАО «НПО «Алмаз») (127411, Россия, Москва, Дмитровское шоссе, д. 110). E-mail: aapushkov@mail.ru. Scopus Author ID 54407412700, <https://orcid.org/0009-0005-3831-4528>

Халтурина Елизавета Максимовна, инженер 3 категории, ПАО «Научно-производственное объединение «Алмаз» имени академика А.А. Расплетина (ПАО «НПО «Алмаз») (127411, Россия, Москва, Дмитровское шоссе, д. 110). E-mail: e.m.halturina@yandex.ru. <https://orcid.org/0009-0003-8527-419X>

About the Authors

Boris M. Vovshin, Dr. Sci. (Eng.), Professor, Department of Radio Wave Processes and Technology, Institute of Radio Electronics and Informatics, MIREA – Russian Technological University (78, Vernadskogo pr., Moscow, 119454 Russia); Chief Researcher, PJSC “ALMAZ R&P Corp.” (110, Dmitrovskoe sh., Moscow, 127411 Russia). E-mail: boris@eleron.net. Scopus Author ID 7801355022, <https://orcid.org/0009-0003-1357-5866>

Alexander A. Pushkov, Cand. Sci. (Eng.), Head of the Scientific and Technical Center, PJSC “ALMAZ R&P Corp.” (110, Dmitrovskoe sh., Moscow, 127411 Russia). E-mail: aapushkov@mail.ru. <https://orcid.org/0009-0005-3831-4528>

Elizaveta M. Khalturina, Engineer, PJSC “ALMAZ R&P Corp.” (110, Dmitrovskoe sh., Moscow, 127411 Russia). E-mail: e.m.halturina@yandex.ru. <https://orcid.org/0009-0003-8527-419X>

УДК 621.372.85

<https://doi.org/10.32362/2500-316X-2025-13-3-84-91>

EDN NHBOSK



КРАТКОЕ СООБЩЕНИЕ

Управление амплитудно-частотной характеристикой узкополосного фильтра для X-диапазона частот на основе фотонного кристалла с подвижным цилиндрическим дефектом

Е.А. Рябов[@],
А.А. Андреев,
С.А. Сергеев,
А.И. Михайлов

Саратовский национальный исследовательский государственный университет имени Н.Г. Чернышевского, Саратов, 410012 Россия

[@] Автор для переписки, e-mail: k1u2r3ka@mail.ru

• Поступила: 12.11.2024 • Доработана: 14.02.2025 • Принята к опубликованию: 21.03.2025

Резюме

Цели. Цель работы – исследовать возможность и эффективность использования в конструкции фотонного кристалла подвижного цилиндрического дефекта с металлическими штырями для управления амплитудно-частотной характеристикой (АЧХ) узкополосного фильтра на прямоугольном волноводе с сечением 23×10 мм в трехсантиметровом диапазоне (X-диапазоне), определить способы наиболее эффективного управления АЧХ.

Методы. Для численного моделирования АЧХ фильтра используется программный пакет *openEMS*, в основе которого лежит система уравнений Максвелла, решаемая методом конечных разностей во временной области. Проведено также экспериментальное исследование АЧХ действующего макета предложенной и созданной конструкции фильтра в трехсантиметровом диапазоне (X-диапазоне).

Результаты. Результаты численного моделирования показывают, что введение в центр конструкции фотонного кристалла подвижного цилиндрического дефекта с двумя металлическими штырями приводит к появлению в полосе запираания на АЧХ фильтра резонансного пика пропускания, положение которого эффективно управляется поворотом цилиндрического дефекта вокруг его оси. При неизменном положении цилиндрического дефекта уменьшение периода фотонного кристалла приводит к увеличению частоты пика пропускания. На частоту резонансного пика пропускания наиболее сильное влияние оказывает изменение размера отверстий в конструкции фотонного кристалла, что может использоваться как эффективный фактор для управления положением пика пропускания и формой всей АЧХ; при этом значение коэффициента пропускания при повороте цилиндрического дефекта вокруг его оси практически не изменяется. Проведены также экспериментальные исследования АЧХ фотонных кристаллов, изготовленных с использованием технологии 3D-печати из пластика PETG (полиэтилентерефталатгликоль), данные которых согласуются с результатами моделирования.

Выводы. Предложенные спроектированные и изготовленные экспериментальные модели узкополосных фильтров в трехсантиметровом диапазоне (X-диапазоне) на основе фотонного кристалла показали достаточные для практики изменения значения коэффициента пропускания и возможности эффективного управления частотой резонансного пика и всей формой АЧХ, что делает их весьма перспективными для практических применений в радиоэлектронной аппаратуре.

Ключевые слова: узкополосный фильтр, резонансный фильтр, СВЧ-диапазон, фотонный кристалл, 3D-печать, openEMS

Для цитирования: Рябов Е.А., Андреев А.А., Сергеев С.А., Михайлов А.И. Управление амплитудно-частотной характеристикой узкополосного фильтра для X-диапазона частот на основе фотонного кристалла с подвижным цилиндрическим дефектом. *Russian Technological Journal*. 2025;13(3):84–91. <https://doi.org/10.32362/2500-316X-2025-13-3-84-91>, <https://www.elibrary.ru/NHBOSK>

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

Авторы заявляют об отсутствии конфликта интересов.

SHORT COMMUNICATION

Control of the frequency response of a narrow-band filter for the X-band frequency based on a photonic crystal with a movable cylindrical defect

Evgeny A. Ryabov[@],
Anton A. Andreev,
Sergey A. Sergeev,
Alexander I. Mikhailov

Saratov National Research State University, Saratov, 410012 Russia

[@] Corresponding author, e-mail: k1u2r3ka@mail.ru

• Submitted: 12.11.2024 • Revised: 14.02.2025 • Accepted: 21.03.2025

Abstract

Objectives. The work set out to investigate the possibility and effectivity of using a movable cylindrical defect with metal pins in the design of a photonic crystal to control the frequency response of a narrow-band filter in a rectangular waveguide having a cross-section of 23×10 mm in the X-band, as well as to determine the most effective methods for controlling frequency response.

Methods. A numerical simulation of the frequency response of the filter was carried out using the *openEMS* software package, which is based on Maxwell's equations solved by the finite-difference time-domain method. The frequency response of the currently proposed and implemented filter construction in the X-band was further investigated in an experimental study.

Results. Numerical simulation shows that a resonant transmission peak in the stopband of the frequency response can be caused to appear by introducing a movable cylindrical defect having two metal pins into the center of a photonic crystal structure. In addition, the position of this peak on the frequency response can be effectively controlled by rotating the cylindrical defect around its axis. If the position of the defect remains unchanged,

an increase in the frequency of the transmission peak occurs as a result of decreasing the period of the photonic crystal. However, the frequency of this resonant transmission peak is most strongly influenced by changes in the size of holes in the photonic structure. These changes can be used to control both the position and shape of the transmission peak, as well as the overall frequency response. At the same time, the difference in transmission remains practically unchanged when the cylinder rotates around its axis. The simulation results were confirmed by the data of an experimental study of the frequency response of photonic crystals made from PETG plastic using 3D printing technology.

Conclusions. The proposed, designed, and manufactured experimental samples of narrow-band filters in the X-band based on a photonic crystal demonstrated reliably variable transmission values and the possibility of controlling the resonant peak frequency and thus the entire frequency response, including operational control. This makes them very promising for practical use in radio-electronic equipment.

Keywords: narrow-band filter, resonant filter, microwave range, photonic crystal, 3D printing, openEMS

For citation: Ryabov E.A., Andreev A.A., Sergeev S.A., Mikhailov A.I. Control of the frequency response of a narrow-band filter for the X-band frequency based on a photonic crystal with a movable cylindrical defect. *Russian Technological Journal*. 2025;13(3):84–91. <https://doi.org/10.32362/2500-316X-2025-13-3-84-91>, <https://www.elibrary.ru/NHBOSK>

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

The authors declare no conflicts of interest.

ВВЕДЕНИЕ

Электромагнитные волны широко используются в различных областях науки и техники: от радиолокации и навигации до информационно-телекоммуникационных технологий и медицинского оборудования. Одним из наиболее распространенных типов линий передачи сверхвысокочастотных (СВЧ) электромагнитных волн является прямоугольный волновод. Широко используются так называемые частично и полностью заполненные какими-либо конструкциями волноводы. Обычно конструкции для заполнения представляют собой диэлектрические пластины различных формы и размеров. Существует также возможность использования в качестве таких конструкций искусственных материалов и структур, в частности, основанных на фотонных кристаллах или метаматериалах [1–13]. Фотонный кристалл в СВЧ-диапазоне – определенная конструкция на базе отрезка волновода, заполнение которого представляет собой периодическую структуру, состоящую из отдельных ячеек, выполненных из материалов с разными показателями преломления, в результате чего в его спектре пропускания формируются фотонные запрещенные и разрешенные зоны (диапазоны частот), аналогичные энергетическим зонам в твердых телах [1, 2, 4–9]. Благодаря этому использование фотонных кристаллов, в частности, на основе прямоугольных волноводов дает возможность создавать частотно-селективные устройства [6, 7], которые способны выделять и практически без потерь пропускать на выход устройства СВЧ-излучение как в определенной полосе частот (полосовой фильтр), так и в области частот ниже (фильтр нижних частот) или выше (фильтр верхних частот) определенного уровня частоты среза в заданном частотном диапазоне.

Добавление в фотонный кристалл единичного дефекта приводит к нарушению периодичности его структуры и появлению на его амплитудно-частотной характеристике (АЧХ) резонансного пика пропускания [10–12]. Управление положением дефекта, его формой и размерами открывает возможности создания резонансных (узкополосных) фильтров, всевозможных управляемых датчиков, поглотителей и прочих полезных устройств [1–4, 7–10].

В данной работе представлены исследования фотонных кристаллов с подвижным вращающимся цилиндрическим дефектом. Также изготовлены с использованием технологии 3D-печати образцы фотонных кристаллов и проведено сравнение результатов экспериментальных исследований с характеристиками их математических моделей, выполненных с использованием специализированного программного обеспечения *openEMS*¹.

ОПИСАНИЕ КОНСТРУКЦИИ

В настоящее время для прототипирования изделий в СВЧ-диапазоне начали широко применять весьма перспективную технологию 3D-печати [13–16]. В данной работе для изготовления фотонных кристаллов используется технология моделирования методом послойного наплавления (англ. – fused deposition modeling, FDM), являющаяся одним из видов технологии 3D-печати. Основой технологии FDM является пластиковая нить, в результате ее плавления и нанесения образуется слой, который располагается на поверхности предыдущего слоя, в дальнейшем формируется структура заданной

¹ <https://www.openems.de>. Дата обращения 20.03.2025. / Accessed March 20, 2025.

Таблица. Наиболее значимые размеры элементов проектируемого фотонного кристалла. Форма отверстий – прямоугольная

Форма отверстий	Количество отверстий (i)	Период следования отверстий (Λ), мм	Размер отверстия вдоль оси волновода (w), мм	Расстояние между отверстиями (L), мм	Размер отверстия вдоль широкой стенки волновода (g), мм
Прямоугольная	4	24–31	5–9	19–22	18

модели. Для изготовления конструкции сначала проектируется 3D-модель, которая обычно разрабатывается в системе автоматизированного проектирования (САПР). Модель может быть сохранена в формате файла STL, который хранит трехмерную модель в виде числовых массивов, что широко используется для 3D-печати. Проектирование фотонного кристалла осуществлялось с использованием САПР *OpenSCAD*². Выбор *OpenSCAD* обусловлен ее мощными возможностями для создания сложных трехмерных моделей с высокой степенью параметризации. Это означает, что изменение одного параметра автоматически пересчитывает всю геометрию, значительно упрощая процесс оптимизации конструкции. Созданный фотонный кристалл представляет собой заполненный участок прямоугольного волновода с полностью заполненным сечением 23×10 мм. Воздушные отверстия располагаются периодически вдоль оси волновода. Сечение отверстий имеет прямоугольную форму. В качестве материала для изготовления выбран пластик полиэтилентерефталатгликоль (англ. – polyethylene terephthalate glycol, PETG), в исследуемом диапазоне частот относительная диэлектрическая проницаемость $\epsilon' \approx 2.5$ [15, 16].

На рис. 1 представлено схематическое изображение предлагаемой конструкции и приведены обозначения наиболее значимых размеров элементов конструкции, а в таблице приведены конкретные значения размеров. Аналогичные конструкции фотонных кристаллов предлагались и подробно анализировались в ряде работ (например, [6, 13]).

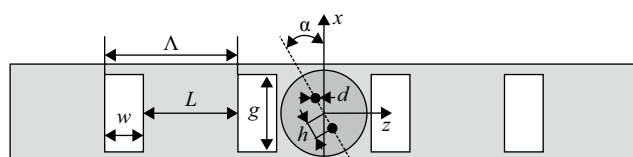


Рис. 1. Схематическое изображение предлагаемой конструкции заполнения фотонного кристалла с подвижным вращающимся дефектом (вид со стороны широкой стенки волновода – в плоскости xz)

В центре спроектированного фотонного кристалла расположен дефект – подвижный, вращающийся вокруг своей оси цилиндр, в котором симметрично на расстоянии $h = 4$ мм по обе стороны от оси располагаются 2 тонких металлических стержня круглого сечения диаметром $d = 2$ мм. Положению первого дефекта (угол поворота $\alpha = 0^\circ$) соответствует расположение пары стержней дефекта поперек оси волновода – вдоль оси x . Положению второго дефекта (угол поворота $\alpha = 90^\circ$) соответствует расположение пары стержней дефекта вдоль оси волновода – вдоль оси z .

РЕЗУЛЬТАТЫ И ИХ ОБСУЖДЕНИЕ

После того, как была выбрана оптимальная конструкция фотонного кристалла, начался этап численного моделирования его характеристик. Для этого использовался пакет *openEMS*. Данный программный пакет основан на методе конечных разностей во временной области, который является одним из популярных численных методов в вычислительной электродинамике [17]. Программа поддерживает импорт и экспорт геометрических моделей из различных форматов файлов (например, PLY, STL), что значительно упрощает процесс моделирования, особенно при использовании САПР для проектирования. Для удобства управления моделированием и обработки результатов *openEMS* интегрируется со скриптовыми языками, такими как *MATLAB*³, *Octave*⁴ и *Python*⁵. Эти языки позволяют автоматизировать процесс настройки параметров модели, запуска расчетов и обработки полученных данных.

На рис. 2 изображены изготовленные с применением технологии 3D-печати конструкции фотонных кристаллов с цилиндрическим дефектом. Рисунок 3 демонстрирует АЧХ фотонных кристаллов, полученные в ходе численного моделирования и экспериментальных исследований. Конкретные

³ <https://www.mathworks.com/products/matlab.html>. Дата обращения 20.03.2025. / Accessed March 20, 2025.

⁴ <https://octave.org/>. Дата обращения 20.03.2025. / Accessed March 20, 2025.

⁵ <https://www.python.org/>. Дата обращения 20.03.2025. / Accessed March 20, 2025.

² <https://openscad.org>. Дата обращения 20.03.2025. / Accessed March 20, 2025.

данные приведены для двух положений дефекта: $\alpha = 0^\circ$ (положение 1) и $\alpha = 90^\circ$ (положение 2). Как видно из результатов численного моделирования и из экспериментальных результатов, при положении 2 дефекта на АЧХ наблюдается наиболее отчетливый резонансный пик пропускания. На частоте резонансного пика изменение значения коэффициента пропускания (ΔT) при изменении положения дефекта с 1 на 2 достигает более 15 дБ.

Зависимости положения частоты пика пропускания от периода фотонного кристалла представлены на рис. 4. При уменьшении периода фотонного кристалла наблюдается смещение пика пропускания в область высоких частот. Изменение размера отверстий w оказывает более сильное влияние на положение частоты пика пропускания по сравнению с изменением расстояния между отверстиями L .



Рис. 2. Общий вид изготовленных фотонных кристаллов с дефектом

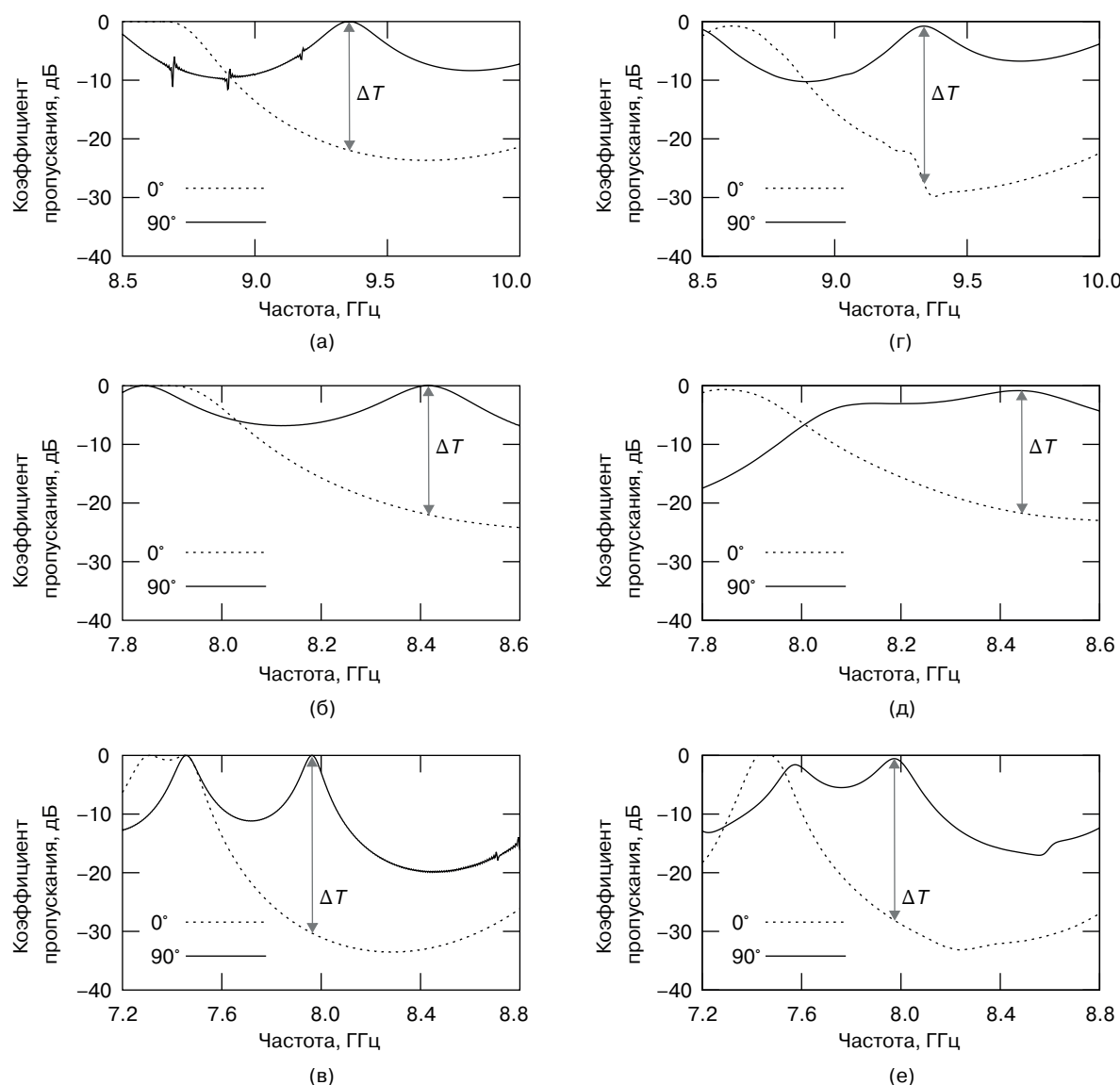


Рис. 3. АЧХ, полученные в результате численного моделирования (а, б, в) и экспериментального исследования (г, д, е) фотонных кристаллов при значениях угла поворота дефекта α , равных 0° и 90° : $L = 19$ мм, $w = 5$ мм (а, г); $L = 22$ мм, $w = 5$ мм (б, д); $L = 22$ мм, $w = 9$ мм (в, е)

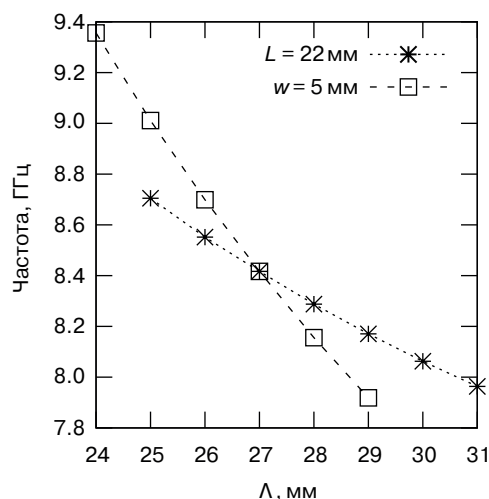


Рис. 4. Зависимости частоты пика пропускания фотонного кристалла от периода Λ при фиксированном размере отверстия w (пунктирная линия) и расстоянием между отверстиями L (штриховая линия)

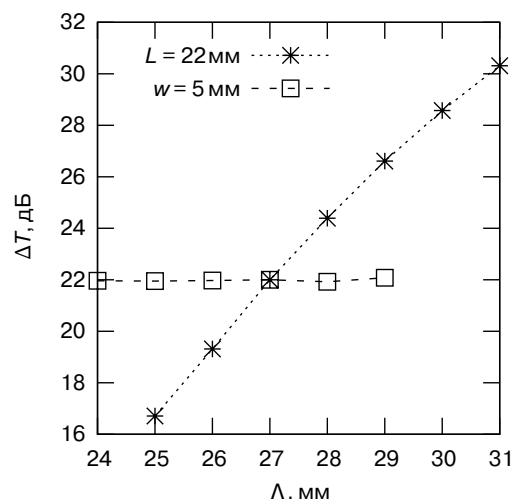


Рис. 5. Зависимость изменения значения коэффициента пропускания (ΔT) от периода фотонного кристалла Λ при фиксированном размере отверстия w (пунктирная линия) и расстоянии между отверстиями L (штриховая линия)

На рис. 5 приведены зависимости изменения значения коэффициента пропускания ΔT от периода фотонного кристалла Λ . Изменение значения коэффициента пропускания ΔT достигает 22 дБ при фиксированном размере отверстия $w = 5$ мм, а влияние расстояния между отверстиями L минимально. При увеличении размера отверстия w от 3 до 9 мм и фиксированном расстоянии между отверстиями $L = 22$ мм изменение значения коэффициента пропускания ΔT фотонного кристалла достигает 14 дБ (с 16 до 30 дБ).

ЗАКЛЮЧЕНИЕ

В работе продемонстрирована возможность эффективного использования в конструкции фотонного кристалла подвижного вращающегося вокруг своей оси цилиндрического дефекта с двумя симметрично расположенными относительно оси одинаковыми металлическими штырями для управления АЧХ узкополосного фильтра в прямоугольном волноводе X-диапазона с сечением 23×10 мм. Результаты

численного моделирования показывают, что при уменьшении периода следования отверстий Λ фотонного кристалла происходит смещение пика пропускания в сторону более высоких частот. Наиболее сильное частотное смещение пика пропускания при увеличении периода следования отверстий происходит при фиксированном расстоянии между отверстиями. При этом для значений угла поворота дефекта 90° (положение 2) наблюдается минимальное изменение значения коэффициента пропускания. Спроектированные и экспериментально разработанные модели узкополосных фильтров, основанные на фотонных кристаллах данной конструкции, демонстрируют изменение значения коэффициента пропускания в диапазоне от 16 до 30 дБ при изменении угла поворота дефекта с 0° на 90° . Эти результаты вполне подходят для использования в реальных технических приложениях.

Вклад авторов. Все авторы в равной степени внесли свой вклад в исследовательскую работу.

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

СПИСОК ЛИТЕРАТУРЫ

1. Wan B.F., Xu Y., Zhou Z.W., Zhang D., Zhang H.F. Theoretical investigation of a sensor based on one-dimensional photonic crystals to measure four physical quantities. *IEEE Sens. J.* 2020;21(3):2846–2853. <https://doi.org/10.1109/JSEN.2020.3027759>
2. Aly A.H., Mohamed D., Mohaseb M.A. Theoretical and simulation study in defective semiconductor layer that incorporated with superconducting-dielectric photonic crystal. *Int. J. Modern Phys. B.* 2019;33(32):1950397. <https://doi.org/10.1142/S0217979219503971>

3. Бучнев И.Ю., Осипов О.В. Исследование электромагнитных свойств поперечной вставки на основе планарного слоя кирального метаматериала в прямоугольном волноводе. *Физика волновых процессов и радиотехнические системы*. 2023;26(1):93–105. <https://doi.org/10.18469/1810-3189.2023.26.1.93-105>
4. Ghasemi F., Aliasghary M., Razi S. Magneto-sensitive photonic crystal optical filter with tunable response in 12–19 GHz; cross over from design to prediction of performance using machine learning. *Phys. Lett. A*. 2021;401:127328. <https://doi.org/10.1016/j.physleta.2021.127328>
5. Zhao L., Li Y., Chen Z.M., Liang Z.H., Wang J., Shen X., Zhang Q. A Band-Pass Filter Based on Half-Mode Substrate Integrated Waveguide and Spoof Surface Plasmon Polaritons. *Sci Rep*. 2019;9(1):13429. <https://doi.org/10.1038/s41598-019-50056-9>
6. Михайлов А.И., Рябов Е.А., Сергеев С.А. Оценка возможностей 3D-печати для изготовления волноводных фотонных кристаллов. *Физика волновых процессов и радиотехнические системы*. 2022;25(3):29–35. <https://doi.org/10.18469/1810-3189.2022.25.3.29-35>
7. Ищенко Е.А., Пастернак Ю.Г., Пендюрин В.А., Фёдоров С.М., Черноиваненко И.А. Активный волноводный полосовой фильтр на основе метаматериала. *Вестник Воронежского государственного технического университета*. 2022;18(3):57–60. <https://doi.org/10.36622/VSTU.2022.18.3.007>
8. Liu L., Ye M., Yu Z., Xue W. Notch microwave photonic filter with narrow bandwidth and ultra-high all-optical tuning efficiency based on a silicon nanobeam cavity. *J. Lightwave Technol.* 2023;41(15):5051–5058. <https://doi.org/10.1109/JLT.2023.3248611>
9. Yu B., Yang J., Song Y., Wang Z., Zhang T., Yan B., Xu R. Terahertz Metamaterial Waveguide with I-Shaped Resonators for Phase and Absorption Modulation. *Photonics*. 2023;10(7):816. <https://doi.org/10.3390/photonics10070816>
10. Скрипаль А.В., Пономарев Д.В., Шаронов В.Е. Резонансные характеристики сверхвысокочастотных фотонных кристаллов с включениями в виде проводящих нанослоев. *Письма в Журнал технической физики (Письма в ЖТФ)*. 2023;49(19):27–30. <https://doi.org/10.61011/PJTF.2023.19.56269.19645>
11. Kumar N., Pandey G.N., Dhayal S., Dhayal S.S. Microwave Propagation Characteristics in Magnetized-Cold-Plasma-Based Binary Photonic Crystal with Defect of MCP Layer. *Macromol. Symp.* 2023;407(1):2100515. <https://doi.org/10.1002/masy.202100515>
12. Усанов Д.А., Скрипаль А.В., Посадский В.Н., Тяжлов В.С., Байкин А.В. Дефектная мода в СВЧ волноводных брэгговских структурах с металлическими штырями. *Журнал технической физики*. 2019;89(10):1606–1610. <https://doi.org/10.21883/JTF.2019.10.48180.6-19>
13. Хайрушев И.В., Рябов Е.А., Сергеев С.А. Теоретические и экспериментальные исследования фотонных кристаллов, изготовленных технологией 3D-печати, в X-диапазоне. *Электроника и микроэлектроника СВЧ*. 2022;1:546–549. URL: <https://mwelectronics.etu.ru/assets/files/2022/546-549.pdf>
14. Pei Z., Xu Y., Wei F., Liu T., Su D. Electromagnetic property of a novel gradient honeycomb composite fabricated by 3D forming. *J. Magn. Magn. Mater.* 2020;493:165742. <https://doi.org/10.1016/j.jmmm.2019.165742>
15. Андреев А.А., Рябов Е.А., Хайрушев И.В., Сергеева Б.В., Сергеев С.А. Влияние температуры на диэлектрическую проницаемость пластиков в СВЧ диапазоне. *Электроника и микроэлектроника СВЧ*. 2023;1:388–392. URL: <https://mwelectronics.etu.ru/assets/files/2023/novoe/388-392.pdf>
16. Zechmeister J., Lacik J. Complex Relative Permittivity Measurement of Selected 3D-Printed Materials up to 10 GHz. In: *2019 Conference on Microwave Techniques (COMITE)*. IEEE; 2019. P. 1–4. <https://doi.org/10.1109/COMITE.2019.8733590>
17. Liebig T., Rennings A., Held S., Erni D. OpenEMS – a free and open source equivalent-circuit (EC) FDTD simulation platform supporting cylindrical coordinates suitable for the analysis of traveling wave MRI applications. *Int. J. Numerical Modelling: Electronic Networks, Devices and Fields*. 2013;26(6):680–696. <https://doi.org/10.1002/jnm.1875>

REFERENCES

1. Wan B.F., Xu Y., Zhou Z.W., Zhang D., Zhang H.F. Theoretical investigation of a sensor based on one-dimensional photonic crystals to measure four physical quantities. *IEEE Sens. J.* 2020;21(3):2846–2853. <https://doi.org/10.1109/JSEN.2020.3027759>
2. Aly A.H., Mohamed D., Mohaseb M.A. Theoretical and simulation study in defective semiconductor layer that incorporated with superconducting-dielectric photonic crystal. *Int. J. Modern Phys. B*. 2019;33(32):1950397. <https://doi.org/10.1142/S0217979219503971>
3. Buchnev I.Yu., Osipov O.V. Investigation of the electromagnetic properties of a transverse insert based on a planar layer of a chiral metamaterial in a rectangular waveguide. *Fizika volnovykh protsessov i radiotekhnicheskie sistemy = Physics of Wave Processes and Radio Systems*. 2023;26(1):93–105 (in Russ.). <https://doi.org/10.18469/1810-3189.2023.26.1.93-105>
4. Ghasemi F., Aliasghary M., Razi S. Magneto-sensitive photonic crystal optical filter with tunable response in 12–19 GHz; cross over from design to prediction of performance using machine learning. *Phys. Lett. A*. 2021;401:127328. <https://doi.org/10.1016/j.physleta.2021.127328>
5. Zhao L., Li Y., Chen Z.M., Liang Z.H., Wang J., Shen X., Zhang Q. A Band-Pass Filter Based on Half-Mode Substrate Integrated Waveguide and Spoof Surface Plasmon Polaritons. *Sci Rep*. 2019;9(1):13429. <https://doi.org/10.1038/s41598-019-50056-9>
6. Mikhailov A.I., Ryabov E.A., Sergeev S.A. Evaluation of the possibilities of 3D-printing for the making of waveguide photonic crystals. *Fizika volnovykh protsessov i radiotekhnicheskie sistemy = Physics of Wave Processes and Radio Systems*. 2022;25(3):29–35 (in Russ.). <https://doi.org/10.18469/1810-3189.2022.25.3.29-35>
7. Ishchenko E.A., Pasternak Yu.G., Pendyurin V.A., Fyedorov S.M., Chernov Ivanenko I.A. Active rectangular waveguide bandpass filter based on the metamaterial. *Vestnik Voronezhskogo gosudarstvennogo tekhnicheskogo universiteta = Bulletin of the Voronezh State Technical University*. 2022;18(3):57–60 (in Russ.). <https://doi.org/10.36622/VSTU.2022.18.3.007>

8. Liu L., Ye M., Yu Z., Xue W. Notch microwave photonic filter with narrow bandwidth and ultra-high all-optical tuning efficiency based on a silicon nanobeam cavity. *J. Lightwave Technol.* 2023;41(15):5051–5058. <https://doi.org/10.1109/JLT.2023.3248611>
9. Yu B., Yang J., Song Y., Wang Z., Zhang T., Yan B., Xu R. Terahertz Metamaterial Waveguide with I-Shaped Resonators for Phase and Absorption Modulation. *Photonics*. 2023;10(7):816. <https://doi.org/10.3390/photonics10070816>
10. Skripal A.V., Ponomarev D.V., Sharonov V.E. Resonance characteristics of microwave photonic crystals with inclusions in the form of conducting nanolayers. *Technical Physics Letters*. 2023;10:23–26.
[Original Russian Text: Skripal A.V., Ponomarev D.V., Sharonov V.E. Resonance characteristics of microwave photonic crystals with inclusions in the form of conducting nanolayers. *Pis'ma v Zhurnal tekhnicheskoi fiziki*. 2023;49(19):27–30 (in Russ.). <https://doi.org/10.61011/PJTF.2023.19.56269.19645>]
11. Kumar N., Pandey G.N., Dhayal S., Dhayal S.S. Microwave Propagation Characteristics in Magnetized-Cold-Plasma-Based Binary Photonic Crystal with Defect of MCP Layer. *Macromol. Symp.* 2023;407(1):2100515. <https://doi.org/10.1002/masy.202100515>
12. Usanov D.A., Skripal' A.V., Posadskii V.N., et al. Defect Mode in Microwave Waveguide Bragg Structures with Metal Pins. *Tech. Phys.* 2019;64(10):1523–1526. <https://doi.org/10.1134/S1063784219100232>
[Original Russian Text: Usanov D.A., Skripal' A.V., Posadskii V.N., Tyazhlov V.S., Baikin A.V. Defect Mode in Microwave Waveguide Bragg Structures with Metal Pins. *Zhurnal tekhnicheskoi fiziki*. 2019;89(10):1606–1610 (in Russ.). <https://doi.org/10.21883/JTF.2019.10.48180.6-19>]
13. Khairushev I.V., Ryabov E.A., Sergeev S.A. Theoretical and experimental studies of photonic crystals manufactured by 3D printing technology in the X-band. *Microwave Electronics and Microelectronics*. 2022;1:546–549 (in Russ.). Available from URL: <https://mwelectronics.etu.ru/assets/files/2022/546-549.pdf>
14. Pei Z., Xu Y., Wei F., Liu T., Su D. Electromagnetic property of a novel gradient honeycomb composite fabricated by 3D forming. *J. Magn. Magn. Mater.* 2020;493:165742. <https://doi.org/10.1016/j.jmmm.2019.165742>
15. Andreev A.A., Ryabov E.A., Khairushev I.V., Sergeeva B.V., Sergeev S.A. The influence of temperature on the dielectric constant of plastics in the microwave range. *Microwave Electronics and Microelectronics*. 2023;1:388–392 (in Russ.). Available from URL: <https://mwelectronics.etu.ru/assets/files/2023/novoe/388-392.pdf>
16. Zechmeister J., Lacik J. Complex Relative Permittivity Measurement of Selected 3D-Printed Materials up to 10 GHz. In: *2019 Conference on Microwave Techniques (COMITE)*. IEEE; 2019. P. 1–4. <https://doi.org/10.1109/COMITE.2019.8733590>
17. Liebig T., Rennings A., Held S., Erni D. OpenEMS – a free and open source equivalent-circuit (EC) FDTD simulation platform supporting cylindrical coordinates suitable for the analysis of traveling wave MRI applications. *Int. J. Numerical Modelling: Electronic Networks, Devices and Fields*. 2013;26(6):680–696. <https://doi.org/10.1002/jnm.1875>

Об авторах

Рябов Евгений Александрович, ассистент, кафедра физики твердого тела, Институт физики, ФГБОУ ВО «Саратовский национальный исследовательский государственный университет имени Н.Г. Чернышевского» (410012, Россия, Саратов, ул. Астраханская, д. 83). E-mail: k1u2r3ka@mail.ru. SPIN-код РИНЦ 9110-4151, <https://orcid.org/0000-0003-4777-7346>

Андреев Антон Андреевич, инженер, учебная лаборатория по полупроводниковой электронике, Институт физики, ФГБОУ ВО «Саратовский национальный исследовательский государственный университет имени Н.Г. Чернышевского» (410012, Россия, Саратов, ул. Астраханская, д. 83). E-mail: andreev25304@mail.ru. SPIN-код РИНЦ 6173-0839, <https://orcid.org/0009-0003-3212-6484>

Сергеев Сергей Алексеевич, к.ф.-м.н., доцент, кафедра физики твердого тела, Институт физики, ФГБОУ ВО «Саратовский национальный исследовательский государственный университет имени Н.Г. Чернышевского» (410012, Россия, Саратов, ул. Астраханская, д. 83). E-mail: ssergeev@bk.ru. SPIN-код РИНЦ 6883-7787, <https://orcid.org/0000-0002-4442-6797>

Михайлов Александр Иванович, д.ф.-м.н., профессор, кафедра физики твердого тела, Институт физики, ФГБОУ ВО «Саратовский национальный исследовательский государственный университет имени Н.Г. Чернышевского» (410012, Россия, Саратов, ул. Астраханская, д. 83). E-mail: mikhailovai13@mail.ru. SPIN-код РИНЦ 2491-0488, <https://orcid.org/0000-0002-4158-9195>

About the Authors

Evgeny A. Ryabov, Assistant, Department of Solid State Physics, Institute of Physics, Saratov National Research State University (83, Astrakhanskaya ul., Saratov, 410012 Russia). E-mail: k1u2r3ka@mail.ru. RSCI SPIN-code 9110-4151, <https://orcid.org/0000-0003-4777-7346>

Anton A. Andreev, Engineer, Semiconductor Technology Educational Laboratory, Institute of Physics, Saratov National Research State University (83, Astrakhanskaya ul., Saratov, 410012 Russia). E-mail: andreev25304@mail.ru. RSCI SPIN-code 6173-0839, <https://orcid.org/0009-0003-3212-6484>

Sergey A. Sergeev, Cand. Sci. (Phys.-Math.), Associate Professor, Department of Solid State Physics, Institute of Physics, Saratov National Research State University (83, Astrakhanskaya ul., Saratov, 410012 Russia). E-mail: ssergeev@bk.ru. RSCI SPIN-code 6883-7787, <https://orcid.org/0000-0002-4442-6797>

Alexander I. Mikhailov, Dr. Sci. (Phys.-Math.), Professor, Department of State Physics, Institute of Physics, Saratov National Research State University (83, Astrakhanskaya ul., Saratov, 410012 Russia). E-mail: mikhailovai13@mail.ru. RSCI SPIN-code 2491-0488, <https://orcid.org/0000-0002-4158-9195>

УДК 621.372.8

<https://doi.org/10.32362/2500-316X-2025-13-3-92-102>

EDN OEBXOF



НАУЧНАЯ СТАТЬЯ

Разработка сверхвысокочастотного фильтра нижних частот на основе проекционной модели микрополосковой линии

А.Д. Ярлыков[®],
О.А. Демин

МИРЭА – Российский технологический университет, Москва, 119454 Россия

[®] Автор для переписки, e-mail: yarlykov@mirea.ru

• Поступила: 18.09.2024 • Доработана: 03.02.2025 • Принята к опубликованию: 24.03.2025

Резюме

Цели. Отрезки микрополосковых линий конечной длины применяются для разработки интегральных схем и устройств сверхвысоких частот (СВЧ) различного назначения, таких как делители мощности, направленные ответвители, аттенюаторы и фильтры, имеющих, в большинстве случаев, сложную топологическую структуру. В частности, фильтры нижних частот (ФНЧ) СВЧ-диапазона представляют собой ступенчатое соединение регулярных отрезков микрополосковых линий с различными геометрическими параметрами. Однако современные подходы к расчету фильтров СВЧ при помощи коммерческих программ требуют больших вычислительных и временных затрат, связанных, в частности, с предложенными подходами к электродинамическому анализу микрополосковых линий. Целью статьи является разработка алгоритма и методики расчета фильтров с использованием проекционного подхода к электродинамическому анализу микрополосковых линий, позволяющих сократить время расчета характеристик фильтров СВЧ при сохранении высокой точности полученных результатов.

Методы. Предложен проекционный подход к проведению электродинамического анализа микрополосковой линии, позволяющий быстро и с высокой точностью проводить расчет ее основных электродинамических параметров – коэффициента замедления и волнового сопротивления в широком диапазоне изменения геометрических параметров линии, ее диэлектрической проницаемости и частоты.

Результаты. На базе аналитических выражений для расчета электродинамических параметров микрополосковой линии получены формулы для описания характера изменений элементов матрицы рассеяния многокаскадных ФНЧ в заданной полосе частот. Разработана компьютерная программа, позволяющая рассчитывать значения элементов матрицы рассеяния ФНЧ в широком диапазоне диэлектрической проницаемости подложки и частоты. Проведено сравнение полученных результатов с характеристиками фильтров, рассчитанных при помощи коммерческих программ.

Выводы. Предложенный подход к расчету электродинамических параметров микрополосковых линий и, как следствие, элементов матрицы рассеяния многокаскадных ФНЧ позволяет значительно сократить время расчетов при достижении достаточно высокой точности полученных результатов, что значительно снижает трудозатраты при проектировании фильтров СВЧ в инженерной практике.

Ключевые слова: микрополосковая линия, проекционный подход, фильтр нижних частот, коэффициент за-медления, волновое сопротивление, матрица рассеяния, коэффициент отражения, коэффициент передачи

Для цитирования: Ярлыков А.Д., Демин О.А. Разработка сверхвысокочастотного фильтра нижних частот на основе проекционной модели микрополосковой линии. *Russian Technological Journal*. 2025;13(3):92–102. <https://doi.org/10.32362/2500-316X-2025-13-3-92-102>, <https://www.elibrary.ru/OEBXOF>

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

Авторы заявляют об отсутствии конфликта интересов.

RESEARCH ARTICLE

Development of a microwave low-pass filter based on a microstrip line projection model

Alexey D. Yarlykov[@],
Oleg A. Demin

MIREA – Russian Technological University, Moscow, 119454 Russia

[@] Corresponding author, e-mail: yarlykov@mirea.ru

• Submitted: 18.09.2024 • Revised: 03.02.2025 • Accepted: 24.03.2025

Abstract

Objectives. Sections of microstrip lines having finite length are widely used to develop integrated circuits and microwave devices for various purposes, such as power dividers, directional couplers, attenuators, and filters. In particular, low-pass filters in the microwave range are comprised of a cascade connection of regular sections of microstrip lines having various geometric parameters. However, modern approaches to calculating microwave filters using commercial software require large computational and time-consuming resources, especially when carrying out electrodynamic analysis of microstrip lines. The work set out to develop an algorithm and a method for calculating filters using a projection approach to the electrodynamic analysis of microstrip lines that reduces the time required to calculate characteristics of microwave filters while maintaining high accuracy of the obtained results.

Methods. The proposed projection approach to the electrodynamic analysis of a microstrip line can be used to rapidly and accurately calculate the main electrodynamic parameters of retardation coefficient and wave impedance across a wide range of changes in the geometrical parameters of the line, as well as its dielectric constant and frequency.

Results. Formulas obtained on the basis of analytical expressions for calculating the electrodynamic parameters of a microstrip line are used to describe the nature of changes in the elements of the scattering matrix of multistage low-pass filters in a given frequency band. A developed computer program was used to calculate the values of the elements of the low-pass filter scattering matrix across a wide range of substrate dielectric constant and frequency parameters. The obtained results were compared with the characteristics of filters calculated using commercial software.

Conclusions. The proposed approach to calculating the electrodynamic parameters of microstrip lines and consequent elements of the scattering matrix of multistage low-pass filters can significantly reduce the calculation time while achieving a sufficiently high accuracy of the obtained results to significantly reduce labor costs when calculating microwave filters in engineering practice.

Keywords: microstrip line, projection approach, low-pass filter, retardation coefficient, wave impedance, scattering matrix, reflection coefficient, transmission coefficient

For citation: Yarlykov A.D., Demin O.A. Development of a microwave low-pass filter based on a microstrip line projection model. *Russian Technological Journal*. 2025;13(3):92–102. <https://doi.org/10.32362/2500-316X-2025-13-3-92-102>, <https://www.elibrary.ru/OEBXOF>

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

The authors declare no conflicts of interest.

ВВЕДЕНИЕ

В настоящее время большинство устройств и модулей сверхвысокочастотного (СВЧ) диапазона конструктивно выполняются на основе микрополосковых линий (МПЛ) передачи. Это связано с их малыми массогабаритными параметрами, а также удобством перехода к их топологии от элементов с сосредоточенными параметрами при низкочастотном прототипировании СВЧ-устройств [1]. Так, например, простейшая топология фильтра нижних частот (ФНЧ) СВЧ-диапазона представляет собой каскадное соединение регулярных отрезков МПЛ конечной длины с различным значением ширины полоскового проводника [2]. Геометрические параметры МПЛ, а также диэлектрическая проницаемость ее подложки и частота определяют величину основных электродинамических параметров линии – коэффициента замедления и волнового сопротивления, играющих ключевую роль при расчете характеристик фильтров СВЧ [3]. Однако современные подходы к их расчету при помощи ряда коммерческих программ предполагают достаточно высокие вычислительные и временные затраты как при расчете основных электродинамических параметров МПЛ [4], так и при проектировании фильтров СВЧ в целом [5], в особенности высоких порядков. С учетом этого актуальной является задача применения проекционной модели МПЛ, подробно рассмотренной в [6]. В рамках данной модели открытая МПЛ моделируется экранированной в широком диапазоне изменения ее геометрических параметров, диэлектрической проницаемости подложки и частоты. Помимо этого, в [7] определены минимальные размеры экрана, при которых обеспечивается моделирование открытой МПЛ с заданной точностью, а в [8] представлены методы повышения эффективности предложенной модели МПЛ. В данной статье представлено применение предложенной в [9] модели МПЛ к расчету матрицы рассеяния СВЧ ФНЧ и продемонстрирована высокая точность полученных результатов при значительном снижении временных затрат.

1. МЕТОДИКА ПРОЕКТИРОВАНИЯ СВЧ-ФИЛЬТРА НИЖНИХ ЧАСТОТ

Основная методика проектирования фильтров СВЧ приведена в [10–12]. Согласно классическому подходу, первым этапом проектирования фильтра СВЧ является расчет его низкочастотного прототипа

на сосредоточенных элементах, который начинается с определения числа звеньев в рассчитываемом фильтре. Оно определяется исходя из требуемого типа аппроксимации фильтра и величины затухания в полосе заграждения. Так, для фильтра Баттерворта число звеньев можно определить по формуле:

$$N = \frac{\lg\left(10^{\lceil L(\omega)/10 \rceil} - 1\right)}{2 \lg\left(\omega/\omega_{\text{ср}}\right)}, \quad (1)$$

а для фильтра Чебышева:

$$N = \frac{\text{arch}\left\{\left(10^{\lceil L(\omega)/10 \rceil} - 1\right) / \left(10^{\lceil G_T/10 \rceil} - 1\right)\right\}^{1/2}}{\text{arch}\left(\omega/\omega_{\text{ср}}\right)}, \quad (2)$$

где $L(\omega)$ – величина затухания на частоте ω в полосе заграждения, $\omega_{\text{ср}}$ – частота среза фильтра, G_T – амплитуда пульсаций (Throb) в полосе пропускания (в децибелах).

После определения числа звеньев в фильтре можно составить его эквивалентную низкочастотную схему, где каждое звено фильтра представляется в виде сосредоточенного элемента (индуктивности или емкости) в соответствии с таблицей 5.2 из [12]. Примером такой эквивалентной схемы для пятизвенного ФНЧ является схема, приведенная на рис. 1.

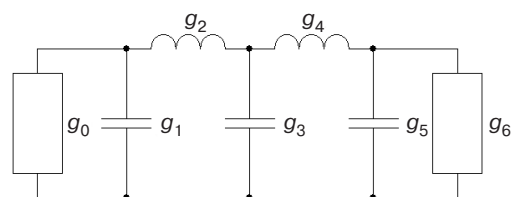


Рис. 1. Эквивалентная схема низкочастотного прототипа пятизвенного ФНЧ. g_i – нормированные параметры эквивалентной схемы

Нормированные параметры эквивалентной схемы (g -параметры) определяются исходя из типа аппроксимации фильтра и общего числа звеньев N . Для фильтра Баттерворта g -параметры определяются по формулам:

$$\begin{aligned} g_0 &= g_{N+1} = 1, \\ g_k &= 2 \sin\left[\frac{(2k-1)\pi}{2N}\right], \quad k = \overline{1, N}, \end{aligned} \quad (3)$$

а для фильтра Чебышева:

$$\begin{aligned} g_0 &= 1, g_1 = \frac{2a_1}{\psi}, \\ g_k &= \frac{4a_{k-1}a_k}{b_{k-1}g_{k-1}}, k = \overline{2, N}, \\ g_{N+1} &= \begin{cases} 1 & \text{при нечетном } N, \\ \text{cth}^2(\beta/4) & \text{при четном } N, \end{cases} \\ \beta &= \ln[\text{cth}(G_T/17.37)], \\ \psi &= \text{sh}\left(\frac{\beta}{2N}\right), \\ a_k &= \sin\left[\frac{(2k-1)\pi}{2N}\right], b_k = \psi^2 + \sin^2\left(\frac{\pi k}{N}\right). \end{aligned} \quad (4)$$

После расчета g -параметров необходимо их денормировать и определить абсолютные величины емкостей C_k и индуктивностей L_k в эквивалентной схеме, а также сопротивление генератора и нагрузки R_k , определяющих, по сути, волновое сопротивление питающей линии. Денормировка осуществляется по правилу:

$$R_k = R_H g_k, C_k = \frac{g_k}{R_H \omega_{cp}}, L_k = \frac{g_k R_H}{\omega_{cp}}, \quad (5)$$

где R_H – сопротивление нагрузки, равное волновому сопротивлению питающей линии.

После расчета параметров эквивалентной схемы необходимо осуществить переход к топологии фильтра на элементах с распределенными параметрами в соответствии с таблицей 5.3 из [10]. Пример топологии пятизвенного ФНЧ на МПЛ приведен на рис. 2. Как видно из рисунка, он образуется линейным полосковым проводником с изменяющейся шириной полоски W вдоль его длины. Участок l_1 МПЛ имеет большое волновое сопротивление относительно волнового сопротивления Z подводящей линии, а участок l_2 имеет меньшее волновое сопротивление. Если $l_1 < \frac{\lambda_{cp}}{8}$ и $l_2 < \frac{\lambda_{cp}}{8}$ $\left(\lambda_{cp} = \frac{6\pi \cdot 10^8}{\omega_{cp}} \right)$, где λ_{cp} – длина волны, соответствующая частоте среза, то участок l_1 имеет индуктивное сопротивление, а l_2 – емкостное сопротивление. Поэтому при выборе волнового сопротивления и расчете длин отрезков следует осуществлять проверку выполнения условий выше.

Для обеспечения одноволнового режима в линии (отсутствия поперечного резонанса) целесообразно ширину полоскового проводника выбирать не больше $\frac{\lambda_{cp}}{4}$.

Для обеспечения скачка сопротивлений при переходе от индуктивного элемента к емкостному

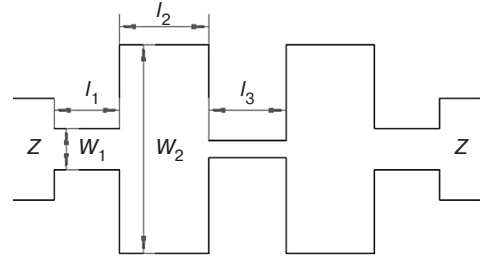


Рис. 2. Топология пятизвенного СВЧ-фильтра нижних частот на МПЛ. W_1, W_2 – ширина полосковых проводников

и наоборот необходимо обеспечить отношение волновых сопротивлений этих элементов не менее чем в 3 раза. Исходя из требуемых волновых сопротивлений, с учетом выполнения описанных условий определяются ширина полосковых проводников, диэлектрическая проницаемость подложки и ее высота.

После выбора волновых сопротивлений и ширины полосковых проводников следует определить длину всех отрезков линии в фильтре. Длина отрезка, реализующего индуктивность, определяется по формуле:

$$l_L = \frac{3 \cdot 10^8}{\omega_{cp} \sqrt{\epsilon}} \arcsin\left(\frac{\omega_{cp} L}{Z_L}\right), \quad (6)$$

а отрезка, реализующего емкость:

$$l_C = \frac{3 \cdot 10^8}{\omega_{cp} \sqrt{\epsilon}} \arcsin(\omega_{cp} C Z_C), \quad (7)$$

где ϵ – диэлектрическая проницаемость подложки, L и C – рассчитанные на этапе низкочастотного прототипирования значения индуктивности и емкости соответственно, Z_L и Z_C – волновые сопротивления индуктивного и емкостного отрезков МПЛ соответственно.

Последним этапом проектирования топологии фильтра является корректировка длин его емкостных и индуктивных отрезков линии с учетом влияния концевых емкостей и индуктивностей. Их величины вычитаются из исходных значений индуктивности и емкости для каждого звена фильтра, и с учетом полученных значений по формулам (6) и (7) вновь рассчитывается скорректированная длина отрезков линии.

На всем этапе проектирования наиболее трудоемким является процесс определения геометрических параметров отрезков полосковых проводников исходя из требуемого соотношения волновых сопротивлений линии, которые, впрочем, можно определить по специальным графикам, представленным, например, в [13]. Однако при расчете матрицы рассеяния фильтра СВЧ следует учесть, что все элементы матрицы являются

функциями, зависящими от частоты. Это обуславливает необходимость расчета волновых сопротивлений и постоянных распространения отрезков МПЛ как дисперсионных характеристик, что и приводит к значительному времени счета при моделировании фильтров при помощи коммерческих программ.

2. ПРОЕКЦИОННЫЙ ПОДХОД К РАСЧЕТУ ЭЛЕКТРОДИНАМИЧЕСКИХ ПАРАМЕТРОВ МПЛ

Для уменьшения вычислительных и временных затрат в [14] рассмотрен проекционный подход, при котором поверхностная плотность тока на полосковом проводнике представляется в виде системы базисных функций в виде полиномов Чебышева, учитывающих особенности поля на краях полоскового проводника. При учете, согласно [14], в разложении продольной составляющей поверхностной плотности тока по «чебышевскому» базису лишь одной базисной функции, можно получить дисперсионное уравнение, из которого определяется коэффициент замедления n_0 МПЛ:

$$\sum_{m=1}^{\infty} \left[\frac{1}{\chi_m^2} (n_0^2 G_m^{\mathcal{E}} + \alpha_m^2 G_m^{\mathcal{M}}) \right] J_0^2(m\alpha) \sin^2(m\beta) = 0, \quad (8)$$

где $G_m^{\mathcal{E}} = \left(\frac{\varepsilon}{\beta_{m1}} \operatorname{ctg}(k_0 \beta_{m1} h) + \frac{1}{\beta_{m2}} \operatorname{ctg}[k_0 \beta_{m2} (b-h)] \right)^{-1}$, $G_m^{\mathcal{M}} = (\beta_{m1} \operatorname{ctg}(k_0 \beta_{m1} h) + \beta_{m2} \operatorname{ctg}[k_0 \beta_{m2} (b-h)])^{-1}$ – функции, полученные при решении электрической (Э) и магнитной (М) задач о собственных волнах, соответственно; $J_0(m\alpha)$ – функция Бесселя; m – целое число, определяющее структуру поля в МПЛ; h – высота подложки; $\beta_{m1} = \sqrt{\varepsilon - \chi_m^2}$; $\beta_{m2} = \sqrt{1 - \chi_m^2}$; $\chi_m^2 = \alpha_m^2 + \Gamma^2$; $\alpha_m = \frac{\pi}{k_0 a} m$; $\alpha = \frac{\pi}{2} \cdot \frac{W}{a}$; $\beta = \alpha \left(1 + \frac{S}{W} \right)$; $k_0 = 2\pi f/c$ – волновое число; f – частота; c – скорость света в вакууме; W – ширина полоски; a , b – размеры экрана; S – расстояние от края полоски до стенки экрана.

Волновое сопротивление Z_0 определяется через мощность, переносимую через поперечное сечение линии, и ток в полосковом проводнике:

$$Z_0 = \frac{240\pi}{k_0 a} n \times \sum_{m=1}^{\infty} \left[- \left(n_0^2 (G_m^{\mathcal{E}})' + \alpha_m^2 (G_m^{\mathcal{M}})' \right) + \frac{\alpha_m^2}{\chi_m^2} (G_m^{\mathcal{M}} - G_m^{\mathcal{E}}) \right] \times \quad (9)$$

$$\times \frac{1}{\chi_m^2} J_0^2(m\alpha) \sin^2(m\beta),$$

где $(G_m^{\mathcal{E},\mathcal{M}})'$ – производная функции $G_m^{\mathcal{E},\mathcal{M}}$ по n_0^2 .

В [9] учтена проблема, связанная с расчетом коэффициента замедления и волнового сопротивления экранированной МПЛ по формулам (8) и (9) – ряды, входящие в данные выражения, сходятся медленно, что приводит к значительному времени расчета параметров, необходимого для обеспечения сходимости рядов. Поэтому в [9] предложены простые формулы для расчета основных электродинамических параметров линии в квазистатическом приближении и определены границы применимости указанных формул. На основе полученных выражений учтена зависимость коэффициента замедления и волнового сопротивления от частоты и предложены простые формулы для расчета дисперсионных характеристик экранированной МПЛ:

$$n(f) \approx \xi_n n_0(f), \quad Z(f) \approx \xi_Z Z_0(f), \quad (10)$$

где $Z_0(f)$ – волновое сопротивление, рассчитанное в «первом приближении» на некоторой частоте f ; ξ_n , ξ_Z – коэффициенты, зависящие от ширины полоскового проводника и определяющиеся по аппроксимационным формулам:

$$\xi_n \approx \begin{cases} 1, & \text{при } W/h < 2, \\ 1 + 3.6 \cdot 10^{-3} (W/h - 2), & \text{при } 2 \leq W/h \leq 10, \end{cases} \quad (11)$$

$$\xi_Z \approx \begin{cases} 1, & \text{при } W/h < 2, \\ 1 - 8.725 \cdot 10^{-3} (W/h - 2), & \text{при } 2 \leq W/h \leq 10. \end{cases}$$

С учетом предложенных формул в таблице приведены значения постоянной распространения и волнового сопротивления МПЛ в широком диапазоне изменения ширины полоскового проводника, диэлектрической проницаемости подложки и частоты. В первых строках каждой ячейки приведены значения постоянной распространения $\Gamma = nk_0$ и волнового сопротивления Z для линии с кварцевой подложкой ($\varepsilon = 3.8$), во вторых строках – для линии с поликоревой подложкой ($\varepsilon = 9.6$), а в третьих строках – для линии с арсенид-галлиевой подложкой ($\varepsilon = 13.3$). При этом следует учесть, что ширина полоскового проводника и частота являются нормированными к высоте подложки h для унификации значений электродинамических параметров линии. Данные таблицы позволят инженеру-разработчику, исходя из заданной частоты среза, определиться с материалом и высотой подложки, а также геометрическими параметрами топологии фильтра СВЧ. Также приведенные для выбранных геометрических и физических параметров фильтра численные значения постоянной распространения и коэффициента замедления необходимы для расчета элементов матрицы рассеяния.

Таблица. Постоянная распространения и волновое сопротивление МПЛ

$f\hbar$, ГГц · мм		0.1	1	3	5	7	10	15
$W/h = 0.1$	Γ	0.19	1.93	5.79	9.67	13.57	19.45	29.38
		0.29	2.90	8.72	14.60	20.55	29.63	45.15
		0.34	3.37	10.16	17.04	24.03	34.72	53.13
	Z	163.70	163.71	163.85	164.16	164.67	165.81	168.81
		109.01	108.99	109.05	109.34	109.92	111.45	116.12
		93.66	93.63	93.68	94.01	94.70	96.56	102.46
$W/h = 0.5$	Γ	0.20	1.97	5.92	9.90	13.90	19.97	30.26
		0.30	2.98	8.99	15.10	21.31	30.85	47.28
		0.35	3.48	10.50	17.67	25.01	36.31	55.92
	Z	101.64	101.65	101.75	102.00	102.40	103.29	105.57
		67.23	67.21	67.23	67.44	67.90	69.06	72.36
		57.69	57.66	57.67	57.91	58.44	59.83	63.82
$W/h = 1$	Γ	0.20	2.01	6.03	10.09	14.18	20.40	30.97
		0.30	3.05	9.22	15.51	21.94	31.83	48.90
		0.36	3.56	10.79	18.20	25.81	37.58	57.96
	Z	75.74	75.74	75.82	76.03	76.36	77.09	78.91
		49.82	49.80	49.80	49.97	50.35	51.31	53.82
		42.70	42.67	42.66	42.86	43.31	44.42	47.31
$W/h = 1.5$	Γ	0.20	2.03	6.12	10.24	14.41	20.74	31.51
		0.31	3.11	9.41	15.85	22.44	32.59	50.09
		0.36	3.63	11.02	18.63	26.44	38.53	59.40
	Z	61.45	61.45	61.52	61.69	61.98	62.61	64.13
		40.24	40.21	40.21	40.37	40.70	41.51	43.51
		34.46	34.42	34.41	34.60	34.99	35.93	38.16
$W/h = 2$	Γ	0.21	2.06	6.19	10.37	14.59	21.02	31.95
		0.32	3.16	9.56	16.12	22.85	33.20	50.98
		0.37	3.69	11.22	18.98	26.96	39.28	60.46
	Z	52.01	52.01	52.06	52.22	52.48	53.05	54.36
		33.93	33.91	33.91	34.07	34.40	35.15	36.89
		29.03	29.00	28.99	29.17	29.54	30.36	32.16
$W/h = 3$	Γ	0.21	2.09	6.31	10.57	14.88	21.45	32.61
		0.32	3.23	9.80	16.55	23.47	34.08	52.22
		0.38	3.79	11.52	19.51	27.73	40.35	61.89
	Z	40.06	40.05	40.10	40.23	40.46	40.93	41.97
		25.99	25.97	25.98	26.15	26.46	27.09	28.41
		22.21	22.18	22.19	22.37	22.69	23.34	24.60
$W/h = 4$	Γ	0.21	2.12	6.39	10.72	15.10	21.76	33.06
		0.33	3.29	9.99	16.87	23.91	34.69	53.01
		0.38	3.85	11.75	19.90	28.26	41.06	62.78
	Z	32.71	32.71	32.75	32.87	33.08	33.49	34.35
		21.14	21.12	21.15	21.32	21.60	22.14	23.17
		18.05	18.03	18.05	18.22	18.51	19.04	19.99
$W/h = 5$	Γ	0.21	2.14	6.46	10.83	15.26	22.00	33.40
		0.33	3.33	10.13	17.11	24.24	35.12	53.55
		0.39	3.91	11.92	20.19	28.65	41.56	63.36
	Z	27.70	27.70	27.74	27.86	28.04	28.41	29.13
		17.85	17.84	17.88	18.05	18.30	18.75	19.56
		15.24	15.21	15.24	15.42	15.67	16.12	16.87
$W/h = 6$	Γ	0.22	2.16	6.51	10.93	15.39	22.18	33.64
		0.34	3.37	10.24	17.30	24.49	35.44	53.94
		0.39	3.95	12.06	20.42	28.95	41.92	63.78
	Z	24.06	24.05	24.09	24.21	24.38	24.71	25.33
		15.47	15.45	15.50	15.67	15.89	16.27	16.92
		13.19	13.17	13.21	13.38	13.61	13.99	14.61

Таблица. Продолжение

$fh, \Gamma\Gamma_{\text{ц}} \cdot \text{мм}$		0.1	1	3	5	7	10	15
$W/h = 7$	Γ	0.22	2.17	6.56	11.00	15.50	22.32	33.84
		0.34	3.40	10.33	17.45	24.69	35.69	54.23
		0.40	3.99	12.17	20.60	29.17	42.19	64.09
	Z	21.27	21.27	21.31	21.42	21.58	21.88	22.42
		13.65	13.64	13.70	13.85	14.05	14.36	14.90
		11.64	11.62	11.67	11.83	12.04	12.37	12.91
$W/h = 8$	Γ	0.22	2.19	6.60	11.06	15.59	22.44	33.99
		0.34	3.42	10.41	17.57	24.84	35.88	54.45
		0.40	4.02	12.27	20.74	29.35	42.40	64.32
	Z	19.08	19.08	19.12	19.23	19.38	19.65	20.13
		12.23	12.23	12.28	12.42	12.59	12.86	13.31
		10.42	10.41	10.46	10.61	10.80	11.09	11.57
$W/h = 10$	Γ	0.22	2.20	6.65	11.16	15.72	22.62	34.21
		0.34	3.46	10.53	17.75	25.07	36.16	54.76
		0.40	4.07	12.41	20.96	29.62	42.71	64.65
	Z	15.84	15.83	15.88	15.98	16.11	16.33	16.72
		10.12	10.12	10.18	10.30	10.43	10.63	10.98
		8.63	8.61	8.68	8.81	8.97	9.21	9.60

Для определения численных значений постоянной распространения и волнового сопротивления в пределах двух соседних частот, их можно аппроксимировать линейной функцией. Это позволяет с достаточной степенью точности рассчитывать элементы матрицы рассеяния.

3. РАСЧЕТ ЭЛЕМЕНТОВ МАТРИЦЫ РАССЕЯНИЯ МНОГОКАСКАДНОГО ФНЧ

Согласно [12], наиболее удобным подходом к расчету матрицы рассеяния многокаскадных фильтров СВЧ является переход к матрицам передачи, рассчитанным в отдельности для каждой нерегулярности в фильтре, их перемножение, а затем обратный переход от итоговой матрицы передачи к итоговой матрице рассеяния. Рассмотрим более подробно расчет матрицы рассеяния и передачи для i -го ступенчатого перехода, изображенного на рис. 3.

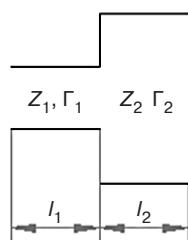


Рис. 3. Топология ступенчатого перехода МПЛ.
 Z_1, Z_2 – волновое сопротивление отрезка МПЛ;
 Γ_1, Γ_2 – постоянные распространения этого отрезка

При подключении источника к левому плечу, а нагрузки – к правому, будем полагать, что

сопротивление источника $Z_{\text{н}}(f) = Z_1(f)$, а сопротивление нагрузки $Z_{\text{н}}(f) = Z_2(f)$. В этом случае коэффициент отражения от первого плеча S_{11} и коэффициент передачи из первого плеча во второе S_{21} будут являться функциями от частоты и определяться как:

$$S_{11}(f) = \frac{Z_2(f) - Z_1(f)}{Z_2(f) + Z_1(f)} e^{-i\Gamma_1(f)2l_1}, \quad (12)$$

$$S_{21}(f) = \sqrt{1 - |S_{11}(f)|^2} e^{-i(\Gamma_1(f)l_1 + \Gamma_2(f)l_2)} = \frac{2\sqrt{Z_1(f)Z_2(f)}}{Z_2(f) + Z_1(f)} e^{-i(\Gamma_1(f)l_1 + \Gamma_2(f)l_2)}. \quad (13)$$

Аналогичные выражения можно получить для коэффициента отражения от второго плеча S_{22} и коэффициента передачи из второго плеча в первое S_{12} , заменив в выражениях (12) и (13) индекс 1 на 2, а индекс 2 – на 1:

$$S_{22}(f) = \frac{Z_1(f) - Z_2(f)}{Z_1(f) + Z_2(f)} e^{-i\Gamma_2(f)2l_2}, \quad (14)$$

$$S_{12}(f) = \sqrt{1 - |S_{22}(f)|^2} e^{-i(\Gamma_2(f)l_2 + \Gamma_1(f)l_1)} = \frac{2\sqrt{Z_2(f)Z_1(f)}}{Z_1(f) + Z_2(f)} e^{-i(\Gamma_2(f)l_2 + \Gamma_1(f)l_1)}. \quad (15)$$

Таким образом, матрицу рассеяния S_i для ступенчатого перехода можно представить в виде:

$$\mathbf{S}_i = \begin{bmatrix} \frac{Z_2(f) - Z_1(f)}{Z_2(f) + Z_1(f)} e^{-i\Gamma_1(f)2l_1} & \frac{2\sqrt{Z_2(f)Z_1(f)}}{Z_1(f) + Z_2(f)} e^{-i(\Gamma_2(f)l_2 + \Gamma_1(f)l_1)} \\ \frac{2\sqrt{Z_1(f)Z_2(f)}}{Z_2(f) + Z_1(f)} e^{-i(\Gamma_1(f)l_1 + \Gamma_2(f)l_2)} & \frac{Z_1(f) - Z_2(f)}{Z_1(f) + Z_2(f)} e^{-i\Gamma_2(f)2l_2} \end{bmatrix}. \quad (16)$$

Переход от матрицы рассеяния $\mathbf{S}_i$ к матрице передачи $\mathbf{T}_i$, согласно [12], осуществляется по правилу:

$$\mathbf{T}_i = \begin{bmatrix} T_{11}(f) & T_{12}(f) \\ T_{21}(f) & T_{22}(f) \end{bmatrix} = \begin{bmatrix} \frac{1}{S_{21}(f)} & -\frac{S_{22}(f)}{S_{21}(f)} \\ \frac{S_{11}(f)}{S_{21}(f)} & S_{12}(f) - \frac{S_{11}(f)S_{22}(f)}{S_{21}(f)} \end{bmatrix}. \quad (17)$$

После расчета матрицы передачи $\mathbf{T}_i$ для каждой i -й нерегулярности в топологии фильтра можно получить итоговую матрицу передачи ФНЧ $\mathbf{T}$ путем перемножения матриц передачи всех его нерегулярностей:

$$\mathbf{T} = \prod_{i=1}^{N+1} \mathbf{T}_i. \quad (18)$$

Переход к матрице рассеяния ФНЧ $\mathbf{S}$ от матрицы передачи $\mathbf{T}$ осуществляется, согласно [12], по правилу:

$$\mathbf{S} = \begin{bmatrix} \frac{T_{21}(f)}{T_{11}(f)} & T_{22}(f) - \frac{T_{21}(f)T_{12}(f)}{T_{11}(f)} \\ \frac{1}{T_{11}(f)} & -\frac{T_{12}(f)}{T_{11}(f)} \end{bmatrix}. \quad (19)$$

4. РЕЗУЛЬТАТЫ ЧИСЛЕННОГО АНАЛИЗА

На основании приведенного алгоритма с использованием языка программирования *GNU Octave*¹ разработана компьютерная программа, позволяющая проводить расчет характеристик ФНЧ в широком диапазоне изменения ширины полоскового проводника ($0.1 \leq W/h \leq 10$), диэлектрической проницаемости подложки ($2 \leq \epsilon \leq 20$) и частоты ($0.1 \text{ ГГц} \leq f \leq 15 \text{ ГГц}$). Структура программы включает в себя основное тело, в котором происходят математические вычисления согласно приведенному в данной работе алгоритму, а также подпрограмму для аппроксимации данных из таблицы по частоте и диэлектрической проницаемости. По результатам работы программы проведен расчет ФНЧ при различной частоте среза (1, 5 и 10 ГГц) и проведено сравнение его характеристик с результатами расчета при помощи коммерческих программ. На рис. 4 представлены амплитудно-частотные характеристики соответствующих фильтров (слева, сплошная линия – расчет фильтра при помощи предложенного

подхода, пунктирная линия – расчет фильтра при помощи коммерческой программы) и значения абсолютной погрешности расчета коэффициента передачи (справа). Время расчета коэффициента передачи при помощи предложенного подхода составляет лишь несколько секунд, в то время как на расчет коэффициента передачи при помощи коммерческой программы было затрачено несколько минут.

Исходя из анализа полученных графиков, можно сделать вывод, что абсолютное значение погрешности коэффициента передачи во всем исследуемом диапазоне не превышает 0.08. При этом для ФНЧ с частотой среза 1 ГГц максимальное значение абсолютной погрешности коэффициента передачи наблюдается в полосе заграждения на частоте 7 ГГц и составляет 0.072. Для ФНЧ с частотой среза 5 ГГц максимальное значение погрешности коэффициента передачи на частоте 12 ГГц составляет 0.04, а для ФНЧ с частотой среза 10 ГГц на частоте 10 ГГц составляет 0.008.

¹ <https://octave.org/>. Дата обращения 20.03.2025. / Accessed March 20, 2025.

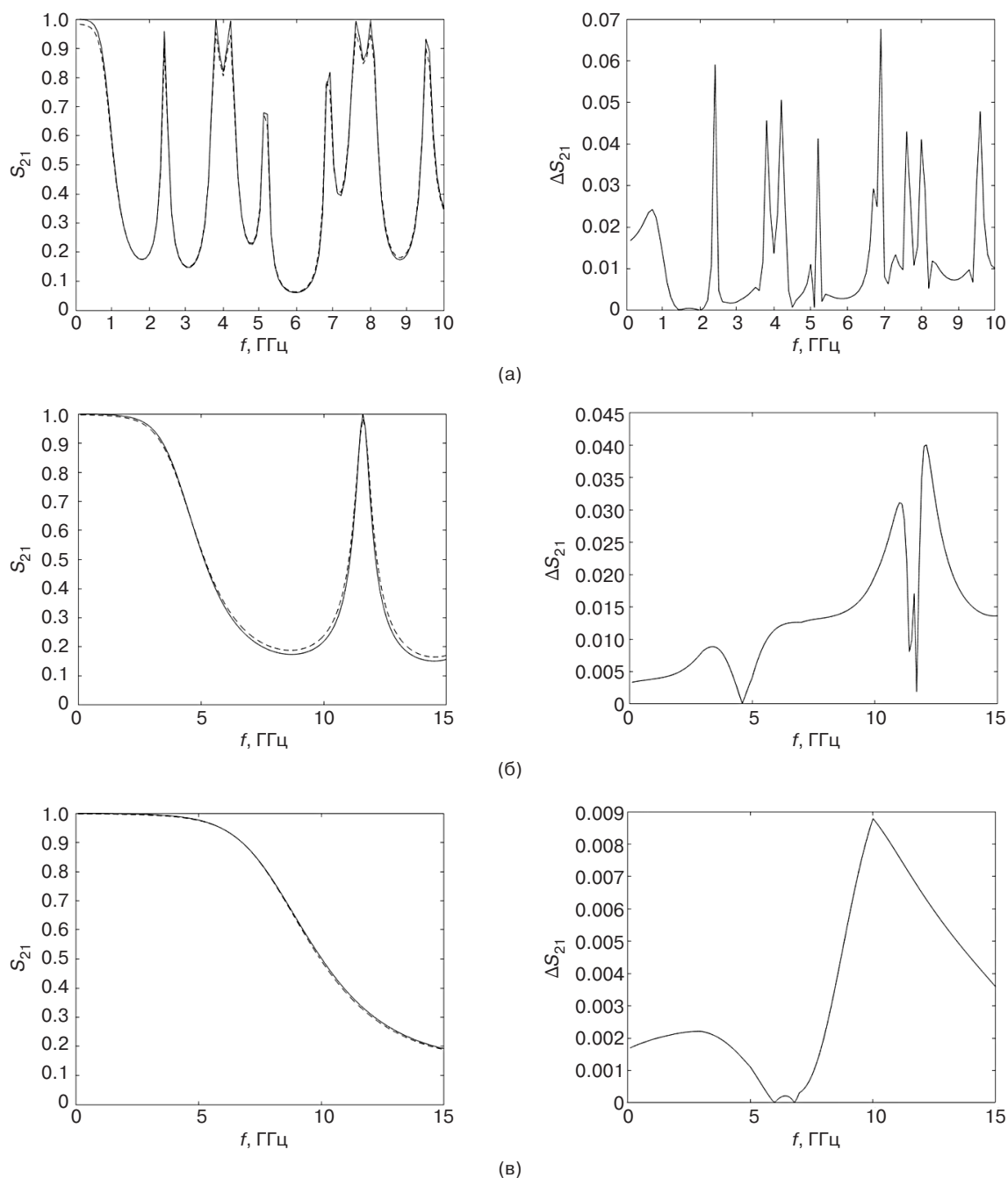


Рис. 4. Амплитудно-частотные характеристики ФНЧ при частоте среза:
(а) 1 ГГц, (б) 5 ГГц, (в) 10 ГГц

ЗАКЛЮЧЕНИЕ

На базе проекционной модели МПЛ проведена разработка микрополоскового ФНЧ и проведен расчет его матрицы рассеяния. Исследована точность полученных результатов путем сравнения коэффициента передачи разработанного фильтра с его моделью, построенной с использованием современных систем автоматизированного проектирования. На базе проведенного сравнения рассчитана абсолютная погрешность коэффициента передачи в широкой полосе частот до 15 ГГц, которая для

различных частот среза фильтра не превышает 0.08. Использование проекционного подхода позволяет значительно (в десятки раз) сократить время расчета коэффициентов отражения и передачи для каждой пары плеч многополюсника СВЧ при достижении достаточно высокой точности полученных результатов.

Вклад авторов. Все авторы в равной степени внесли свой вклад в исследовательскую работу.

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

СПИСОК ЛИТЕРАТУРЫ

1. Петров И.А. Фильтры СВЧ с использованием широкополосных согласующих структур. *Физика волновых процессов и радиотехнические системы*. 2011;14(1):51–56.
2. Ламанов Ю.А., Кудрявцева Т.О., Дроботун Н.Б. Разработка и исследование микрополоскового фильтра низких частот с высокой крутизной спада АЧХ. *Доклады Томского государственного университета систем управления и радиоэлектроники (Доклады ТУСУР)*. 2021;24(2):7–13. <https://doi.org/10.21293/1818-0442-2021-24-2-7-13>
3. *Справочник по элементам полосковой техники*; под ред. А.Л. Фельдштейна. М.: Связь; 1979. 336 с.
4. Гупта К., Гардж Р., Чадха Р. *Машинное проектирование СВЧ устройств*: пер. с англ. М.: Радио и связь; 1987. 432 с.
5. Лопатин В.В., Хворенко В.В. Моделирование и реализация микрополоскового фильтра. В сб.: *Информационные технологии в науке, промышленности и образовании: сборник трудов научно-технической конференции*. Ижевск; 2021. С. 166–176.
6. Коваленко А.Н. Проекционный метод построения электродинамических моделей полосковых линий. *Радиотехника и электроника*. 2019;64(2):108–115. <https://doi.org/10.1134/S0033849419020128>
7. Коваленко А.Н., Ярлыков А.Д. Численный анализ экранированной микрополосковой линии. В сб.: *Актуальные проблемы и перспективы развития радиотехнических и инфокоммуникационных систем (Радиоинфоком-2021): Сборник научных статей V Международной научно-практической конференции*. М.: РТУ МИРЭА; 2021. С. 331–334.
8. Коваленко А.Н., Ярлыков А.Д. Повышение эффективности проекционных моделей полосковых линий. *Радиотехника и электроника*. 2021;66(9):837–844. <https://doi.org/10.31857/S0033849421090084>
9. Коваленко А.Н., Ярлыков А.Д. Аналитические выражения для электродинамических параметров экранированной микрополосковой линии. *Russian Technological Journal*. 2021;9(4):68–76. <https://doi.org/10.32362/2500-316X-2021-9-4-68-76>
10. Фуско В. *СВЧ цепи. Анализ и автоматизированное проектирование*: пер. с англ. М.: Радио и связь; 1990. 288 с.
11. Маттей Д.Л., Янг Л., Джонс Е.М.Т. *Фильтры СВЧ, согласующие цепи и цепи связи*: пер. с англ. М.: Связь; 1972. Т. 1. 496 с.
12. Костин М.С., Ярлыков А.Д. *Устройства и модули сверхвысоких частот*. М., Вологда: Инфра-Инженерия; 2022. 400 с.
13. Малорацкий Л.Г., Явич Л.Р. *Проектирование и расчет СВЧ-элементов на полосковых линиях*. М.: Советское радио; 1972. 232 с.
14. Коваленко А.Н. Собственные волны микрополосковой линии. *Изв. вузов. Радиофизика*. 1978;21(2):188–194.

REFERENCES

1. Petrov I.A. Microwave filters with use broadband matching structures. *Fizika volnovykh protsessov i radiotekhnicheskie sistemy* = *Physics of Wave Processes and Radio Systems*. 2011;14(1):51–56 (in Russ.).
2. Lamanov Yu.A., Kudryavtseva T.O., Drobotun N.B. Design and Research Process of Microstrip Low-Pass Filters with High Slope Steepness. *Doklady Tomskogo gosudarstvennogo universiteta sistem upravleniya i radioelektroniki (Doklady TUSUR)* = *Proceedings of TUSUR University*. 2021;24(2):7–13 (in Russ.).
3. Fel'dshtein A.L. (Ed.). *Spravochnik po elementam poloskovoi tekhniki (Handbook of Elements of Strip Technology)*. Moscow: Svjaz'; 1979. 336 p. (in Russ.).
4. Gupta K.C., Garg R., Chadha R. *Mashinnoe proektirovanie SVCh ustroystv (Computer-aided Design of Microwave Circuits)*: transl. from Engl. Moscow: Radio i svjaz'; 1987. 432 p. (in Russ.).
[Gupta K.C., Garg R., Chadha R. *Computer-aided Design of Microwave Circuits*. Dedham, Mass.: Artech House Inc.; 1981. 636 p.]
5. Lopatin V.V., Khvorenko V.V. Modeling and implementation of a microstrip filter. In: *Information Technologies in Science, Industry and Education: collection of papers of the Scientific and Technical Conference*. Izhevsk; 2021. P. 166–176 (in Russ.).
6. Kovalenko A.N. Projection method for constructing full-wave models of striplines. *J. Commun. Technol. Electron.* 2019;64(2):93–99. <https://doi.org/10.1134/S1064226919020128>
[Original Russian Text: Kovalenko A.N. Projection method for constructing full-wave models of striplines. *Radiotekhnika i elektronika*. 2019;64(2):108–115 (in Russ.). <https://doi.org/10.1134/S0033849419020128>]
7. Kovalenko A.N., Yarlykov A.D. Numerical analysis of a shielded microstrip line. In: *Actual Problems and Prospects for the Development of Radio Engineering and Infocommunication Systems (Radioinfocom-2021): Proceedings of the 5th International Scientific and Practical Conference*. Moscow: RTU MIREA; 2021. P. 331–334 (in Russ.).
8. Kovalenko A.N., Yarlykov A.D. Increasing the efficiency of projection models of strip lines. *J. Commun. Technol. Electron.* 2021;66(9):997–1003. <https://doi.org/10.31857/S0033849421090084>
[Original Russian Text: Kovalenko A.N., Yarlykov A.D. Increasing the efficiency of projection models of strip lines. *Radiotekhnika i elektronika*. 2021;66(9):837–844 (in Russ.). <https://doi.org/10.31857/S0033849421090084>]
9. Kovalenko A.N., Yarlykov A.D. Analytical expressions for electrodynamic parameters of the shielded microstrip line. *Russian Technological Journal*. 2021;9(4):68–76 (in Russ.). <https://doi.org/10.32362/2500-316X-2021-9-4-68-76>
10. Fusco V. *SVCh tsepi. Analiz i avtomatizirovannoe proektirovanie (Microwave Circuits. Analysis and Computer-aided Design)*: transl. from Engl. Moscow: Radio i svjaz'; 1990. 288 p. (in Russ.).
[Fusco V.F. *Microwave Circuits. Analysis and Computer-aided Design*. Prentice Hall; 1986. 358 p.]

11. Matthaei G.L., Yong L., Jones E.M.T. *Fil'try SVCh, soglasuyushchie tsepi i tsepi svyazi (Microwave Filters, Impedance-Matching Networks, and Coupling Structures)*; transl. from Engl. Moscow: Svyaz'; 1972. V. 2. 496 p. (in Russ.).
[Matthaei G.L., Yong L., Jones E.M.T. *Microwave Filters, Impedance-Matching Networks, and Coupling Structures*. McGraw-Hill; 1964. 1096 p.]
12. Kostin M.S., Yarlykov A.D. *Ustroistva i moduli sverkhvysokikh chastot (Devices and Modules of Ultra-High Frequencies)*. Moscow, Vologda: Infra-Inzheneriya; 2022. 400 p. (in Russ.).
13. Maloratskii L.G., Yavich L.R. *Proektirovanie i raschet SVCh-elementov na poloskovykh liniyakh (Design and Calculation of Microwave Elements on Strip Lines)*. Moscow: Sovetskoe radio; 1972. 232 p. (in Russ.).
14. Kovalenko A.N. Natural modes of a microstrip line. *Radiophys. Quantum Electron.* 1978;21(2):128–133. <https://doi.org/10.1007/BF01078702>
[Original Russian Text: Kovalenko A.N. Natural modes of a microstrip line. *Izvestiya vysshikh uchebnykh zavedenii. Radiofizika*. 1978;21(2):188–194 (in Russ.).]

Об авторах

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

Демин Олег Александрович, ассистент, кафедра радиоволновых процессов и технологий, Институт радиоэлектроники и информатики, ФГБОУ ВО «МИРЭА – Российский технологический университет» (119454, Россия, Москва, пр-т Вернадского, д. 78). E-mail: demin_o@mirea.ru. <https://orcid.org/0000-0002-9864-5338>

About the Authors

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

Oleg A. Demin, Assistant, Department of Radio Wave Processes and Technologies, Institute of Radio Electronics and Informatics, MIREA – Russian Technological University (78, Vernadskogo pr., Moscow, 119454 Russia). E-mail: demin_o@mirea.ru. <https://orcid.org/0000-0002-9864-5338>

Микро- и нанoeлектроника. Физика конденсированного состояния

Micro- and nanoelectronics. Condensed matter physics

УДК 531.8; 681.518.3

<https://doi.org/10.32362/2500-316X-2025-13-3-103-121>

EDN KBLENU



ОБЗОРНАЯ СТАТЬЯ

Микроэлектромеханические системы: путь к совершенствованию гироскопов

П.С. Кузнецов [@]

АО «Государственный научно-исследовательский институт приборостроения», Москва, 129226 Россия

[@] Автор для переписки, e-mail: ps_kuznetsov@mail.ru

• Поступила: 31.10.2024 • Доработана: 13.02.2025 • Принята к опубликованию: 21.03.2025

Резюме

Цели. Микросистемная техника является одним из наиболее популярных и перспективных направлений, которые активно развиваются в настоящее время. Область применения элементов микросистемной техники весьма широка. Настоящая работа направлена на всестороннее изучение процессов разработки и создания современных гироскопов на основе микроэлектромеханических систем (МЭМС-гироскопов). Целью исследования является анализ рисков, связанных с технологическими аспектами их производства, а также определение перспективных направлений для дальнейшего развития как самих МЭМС-гироскопов, так и технологий их изготовления.

Методы. В ходе работы осуществлен детализированный анализ существующих научных публикаций, аналитических обзоров и других доступных источников, посвященных МЭМС-гироскопам и актуальным трендам в области микрооптоэлектромеханических технологий и сегнетоэлектрических пленок.

Результаты. Представлено краткое описание конструктивных решений современных МЭМС-гироскопов, а также их интеграция в мехатронные системы. Рассматриваются технологии производства МЭМС-гироскопов и специфика используемого технологического оборудования. В отдельном разделе обсуждаются аспекты настройки и калибровки этих устройств. Выделены перспективные направления развития МЭМС-гироскопов с акцентом на применение микрооптоэлектромеханических преобразователей и сегнетоэлектрических пленок.

Выводы. На основе проведенного анализа показана перспективность развития МЭМС-гироскопов, несмотря на имеющиеся технологические вызовы. Отмечено, что новые физические принципы и уникальные технологии могут способствовать появлению новых видов МЭМС-гироскопов, использующих микрооптоэлектромеханические преобразователи и сегнетоэлектрические пленки. Это, в свою очередь, открывает новые горизонты для будущих разработок в данной области. Показана необходимость разработки новых технологий производства и специализированного оборудования для повышения качества МЭМС-гироскопов.

Ключевые слова: МЭМС-гироскоп, микросистемная техника, технология создания, оборудование производства, микрооптоэлектромеханический преобразователь, оптический туннельный эффект, фотоника, сегнетоэлектричество

Для цитирования: Кузнецов П.С. Микроэлектромеханические системы: путь к совершенствованию гироскопов. *Russian Technological Journal*. 2025;13(3):103–121. <https://doi.org/10.32362/2500-316X-2025-13-3-103-121>, <https://www.elibrary.ru/KBLENU>

Прозрачность финансовой деятельности: Автор не имеет финансовой заинтересованности в представленных материалах или методах.

Автор заявляет об отсутствии конфликта интересов.

REVIEW ARTICLE

Microelectromechanical systems for improved gyroscope design

Pavel S. Kuznetsov [@]

State Scientific Research Institute of Instrument Engineering, Moscow, 129226 Russia

[@] *Corresponding author, e-mail: ps_kuznetsov@mail.ru*

• **Submitted:** 31.10.2024 • **Revised:** 13.02.2025 • **Accepted:** 21.03.2025

Abstract

Objectives. Microsystem engineering is currently receiving a great deal of research attention due to the very wide scope of application of its various elements. The present study of the development and creation of modern gyroscopes based on microelectromechanical systems (MEMS gyroscopes) analyzes the risks associated with the technological aspects of their production and identifies promising areas for further development both of MEMS gyroscopes themselves and the technologies used to manufacture them.

Methods. A detailed analysis of existing scientific publications, analytical reviews, and other available sources on MEMS gyroscopes and current trends in the field of microoptoelectromechanical technologies and ferroelectric films was carried out.

Results. A brief description of the design solutions of modern MEMS gyroscopes and their integration into mechatronic systems is presented. The production technologies of MEMS gyroscopes and specifics of the technological equipment used are considered. A separate section discusses the configuration and calibration aspects of these devices. Promising directions for the development of MEMS gyroscopes with an emphasis on the use of microoptoelectromechanical converters and ferroelectric films are highlighted.

Conclusions. Based on the analysis, the prospects for the development of MEMS gyroscopes are shown, despite the existing technological challenges. It is noted that new physical principles and unique technologies can contribute to the emergence of new types of MEMS gyroscopes using micro-optoelectromechanical converters and ferroelectric films. This, in turn, opens up new horizons for future developments in this area. The necessity of developing new production technologies and specialized equipment to improve the quality of MEMS gyroscopes is demonstrated.

Keywords: MEMS gyroscope, microsystem technology, creation technology, production equipment, microoptoelectromechanical converter, optical tunneling effect, photonics, ferroelectricity

For citation: Kuznetsov P.S. Microelectromechanical systems for improved gyroscope design. *Russian Technological Journal*. 2025;13(3):103–121. <https://doi.org/10.32362/2500-316X-2025-13-3-103-121>, <https://www.elibrary.ru/KBLENU>

Financial disclosure: The author has no financial or proprietary interest in any material or method mentioned.

The author declares no conflicts of interest.

ВВЕДЕНИЕ

Микросистемная техника является одним из наиболее популярных и перспективных направлений, которые развиваются в настоящее время. Область применения элементов микросистемной техники весьма широка. Это датчики первичной информации электрических и неэлектрических величин, микро-моторы и различные элементы авионики, а также и медицинские микроинструменты.

Микроэлектромеханические системы (МЭМС) и МЭМС-гироскопы, в частности, являются наиболее востребованным направлением микросистемной техники [1]. Назначение МЭМС-гироскопа заключается в определении параметров движения систем и приборов, в которых они установлены. Малые габаритные размеры и низкое энергопотребление делают их привлекательными для использования во многих отраслях, таких как автомобильная техника, робототехника, мобильные телефоны и многие другие. Их использование в системах специального назначения (беспилотные летательные аппараты, управляемые снаряды, инерциальные навигационные системы и т.д.) определяет повышенные требования к характеристикам и технологии создания МЭМС-гироскопов [2–6].

Целью работы является исследование процесса создания МЭМС-гироскопов, анализ его особенностей, а также определение перспективных направлений развития МЭМС-гироскопов и методов их производства.

ОСОБЕННОСТИ КОНСТРУКЦИЙ СОВРЕМЕННЫХ МЭМС

Микросистемная техника представляет собой совокупность научно-технических и технологических способов, обеспечивающих создание в объеме и (или) на поверхности твердого тела упорядоченной композиции микронных и субмикронных областей материалов с заданным составом, структурой и геометрией. Это обеспечивает реализацию функций восприятия, преобразования, хранения, обработки, трансляции информации, энергии, движения и выработки управляющих воздействий в требуемых режимах и условиях эксплуатации [7].

Микроэлектромеханическая система представляет из себя систему, в которой объединены микроэлектронные и микромеханические элементы (рис. 1). Данные устройства с помощью механических элементов преобразуют внешнее воздействие в электрический сигнал (гироскопы, акселерометры, датчики давления и т.д.) или под действием электрических сил сами совершают движения [8].



Рис. 1. Схема МЭМС [8]

МЭМС-гироскоп – это микроминиатюрная электромеханическая система, в которой энергия первичных (вынужденных) колебаний инерционной массы под воздействием внешней угловой скорости преобразуется в энергию вторичных колебаний, на основе которых получают информацию об измеряемом воздействии [9–11].

Таким образом, простейший гироскоп на основе МЭМС должен состоять из двух основных функциональных элементов: датчика угловой скорости (ДУС) и сервисной электроники, которая воспринимает, усиливает и обрабатывает сигнал с емкостного выхода датчика, а также управляет работой микромеханической структуры. Рассмотрим, что конструктивно представляют из себя эти два элемента, а также как они располагаются в сборке.

Большинство МЭМС-гироскопов относятся к гироскопам вибрационного типа. В зависимости от типа инерционной массы все конструкции микромеханических датчиков (ММД), используемых в гироскопах, можно разделить на несколько основных видов, представленных на рис. 2 [12, 13].

Первый тип – это балочные инерционные массы. Принцип их действия заключается в следующем: пьезоэлементы придают колебательные движения консольной балке в направлении оси X (рис. 3). Вращение относительно оси Z , направленной параллельно продольной оси балки, вызывает под действием силы Кориолиса колебания вдоль оси Y , которые регистрируются другими пьезоэлементами [14].

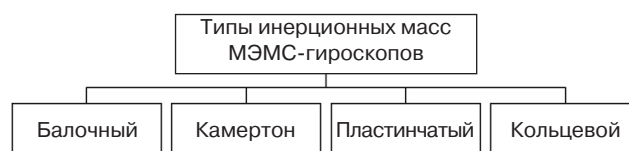


Рис. 2. Типы инерционных масс МЭМС-гироскопов

Второй тип инерционных масс – это камертоны, названные так из-за особенности конструкции резонатора. Построенные на таком принципе гироскопы работают достаточно просто (рис. 4): вращение вокруг вертикальной оси заставляет массы, колеблющиеся до этого в противофазе в одной плоскости, совершать колебания в плоскости, перпендикулярной первичным колебаниям. Детектирование вторичных колебаний осуществляется с помощью емкостных

датчиков, в результате чего получают информацию об угловой скорости [14].

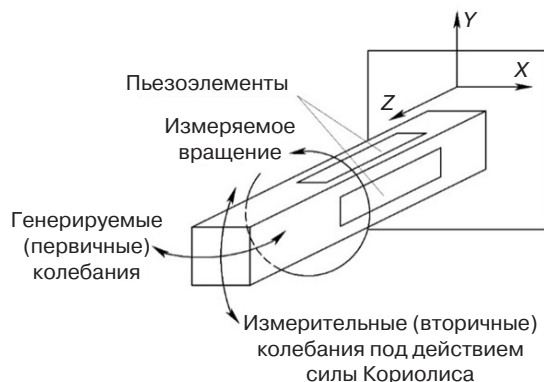


Рис. 3. Принцип действия балочного гироскопа

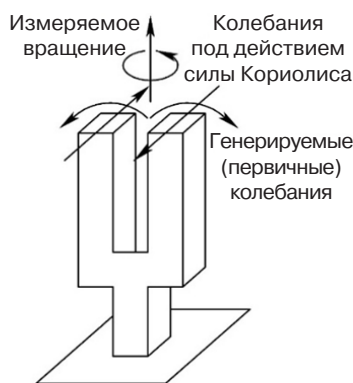


Рис. 4. Принцип действия гироскопа-камертона

Из-за наличия вертикальных колебаний данные гироскопы невозможно изготавливать с помощью планарной технологии, что мешает их массовому производству.

Следующий тип гироскопов – это гироскопы с пластинчатыми инерционными массами [9–11, 15]. В зависимости от вида движения первичных и вторичных колебаний инерциальных масс гироскопы бывают L-L-типа (linear-linear), R-R-типа (rotate-rotate) и R-L-типа (rotate-linear), причем, возможны R-L и L-R-комбинации (рис. 5) [15, 16]. Значительных успехов в области создания гироскопов L-L-типа достигла компания Analog Devices (США), которая создала технологию iMEMS (интегрированные МЭМС) [5]. МЭМС-гироскоп данного типа (рис. 6) работает следующим образом. Инерционные массы 1, подвешенные через двумерные пружины 2, раскачиваются в противоположных направлениях (противофазно), совершая первичные колебания. Пружины 2 обеспечивают перемещение инерционных масс в двух направлениях посредством электростатического датчика силы. При появлении угловой скорости возникают силы Кориолиса, которые заставляют двигаться инерционные массы в направлении, перпендикулярном направлению первичных

колебаний, также противофазно. Инерционные массы заставляют перемещаться гребенки съема, связанные с ними через двумерные пружины 2 и повешенные на одномерных пружинах 4, которые позволяют перемещаться гребенкам съема только в одном направлении. Гребенки съема связаны между собой по дифференциальной схеме, что позволяет получить на выходе сигнал, эквивалентный действующей угловой скорости.



Рис. 5. Типы МЭМС-гироскопов с инерционной массой

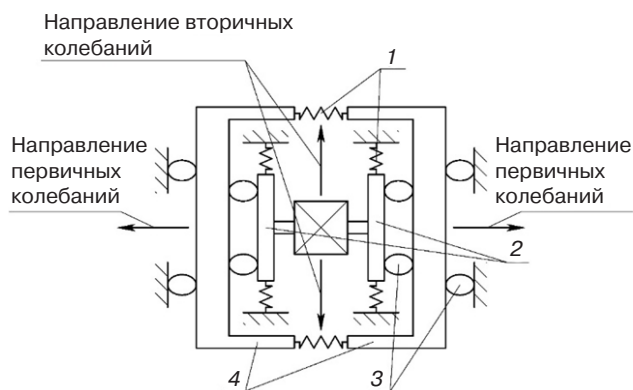


Рис. 6. Принцип действия гироскопа L-L-типа:
1 – одномерные пружины, 2 – гребенки съема, 3 – двумерные пружины, 4 – инерционные массы

Следующий тип МЭМС-гироскопов – это гироскоп R-R-типа. Его принципиальная схема представлена на рис. 7. Инерционная масса (ротор) относительно анкеров, установленных на подложке (основании), имеет подвес, включающий в себя упругие и промежуточные элементы. Электростатический привод вызывает первичные колебательные движения ротора вокруг оси Z. При появлении переносной угловой скорости Ω_x основания переменный гироскопический момент вызывает вторичные колебания ротора вокруг оси Y, что и детектируется емкостными измерителями перемещений [9–11, 17, 18].

Последний тип МЭМС-гироскопов – это гироскоп R-L-типа. По конструкции (рис. 8) он представляет из себя камертон, реализованный в виде двух инерционных масс, закрепленных с помощью упругих элементов по внешней рамке. Сама рамка связана с основанием через упругие элементы, обеспечивающие ей вращательное движение вокруг оси. Электростатический двигатель, выполняемый

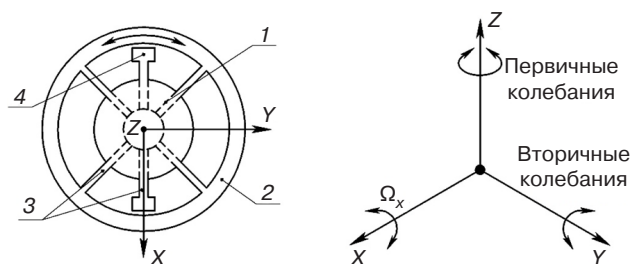


Рис. 7. Принцип действия гироскопа R-R-типа:
1 – промежуточный (кинематический) элемент подвеса, 2 – ротор, 3 – упругие элементы подвеса, 4 – анкер

в виде «гребенчатой» структуры, возбуждает противофазные поступательные колебания масс. При наличии угловой скорости Ω , вектор которой совпадает с измерительной осью вращения рамки, возникают кориолисовы силы, создающие переменный момент, приводящий к угловым колебаниям рамки вокруг оси с частотой, равной частоте двигателя. Амплитуда колебаний рамки является мерой измеряемой угловой скорости. Измерение колебаний рамки осуществляется с помощью емкостного датчика, электроды которого расположены на основании под инерционными массами.

Кольцевой МЭМС-гироскоп представляет из себя частный случай ДУС с распределенными параметрами. Кольцевой резонатор совершает колебания в направлении, соответствующем основной колебательной моде. При воздействии угловой скорости (повороте кольца) ориентация колебательной моды относительно самого кольца изменяется. Это связано со стремлением сохранить свою ориентацию под действием силы инерции, вызванной ускорением Кориолиса [11]. По сути, кольцевой МЭМС-гироскоп является одним из видов волнового твердотельного гироскопа.

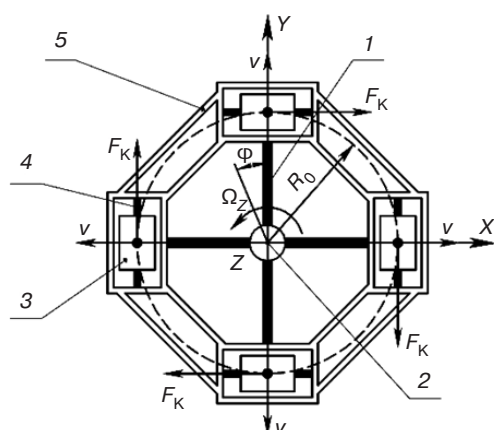


Рис. 8. Принцип действия гироскопа R-L-типа: 1 – упругие элементы подвеса вторичных колебаний, 2 – анкеры, 3 – инерционная масса, 4 – упругие элементы подвеса первичных колебаний, 5 – жесткие элементы подвеса. F_K – вектор силы Кориолиса, v – вектор скорости, ϕ – угол поворота чувствительного элемента (ЧЭ)

Среди множества существующих технологий изготовления ММД, входящих в МЭМС-гироскопы [5, 10, 11, 15, 16], рассмотрим подробнее технологию «кремний на стекле». Микромеханические датчики, изготовленные по данной технологии, представляют из себя вакуум-плотную капсулу, в которой выводы от элементов кремниевой структуры выведены герметично через металлизированные отверстия в стекле и заканчиваются на поверхности контактными площадками. Внутри капсулы находится в вакууме микромеханическая кремниевая структура, представляющая собой вибрационный микрогироскоп. Конструкция ММД в капсульном исполнении представлена на рис. 9.

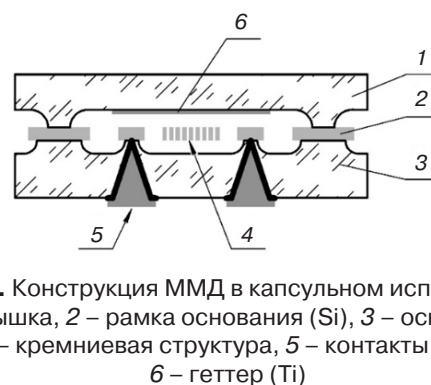
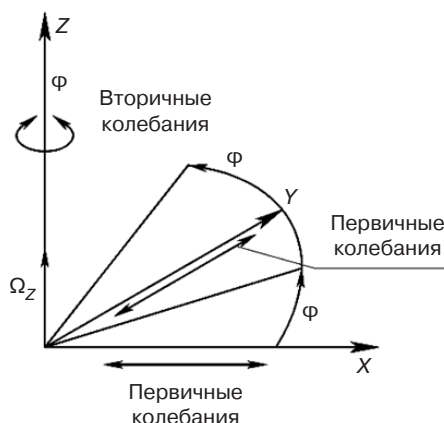


Рис. 9. Конструкция ММД в капсульном исполнении:
1 – крышка, 2 – рамка основания (Si), 3 – основание, 4 – кремниевая структура, 5 – контакты (Al), 6 – геттер (Ti)

Существует несколько вариантов конструктивного изготовления ММД и сервисной электроники. Самым малогабаритным и технологически трудоемким является вариант, при котором датчик и микросхема располагаются на одном кристалле и герметизируются в одном корпусе. Другой вариант представляет собой конструкцию, аналогичную предыдущей, в которой элементы выполняются на двух различных кристаллах. При этом получившийся МЭМС-гироскоп используется как самостоятельный



элемент. Третий вариант исполнения предполагает отдельное капсулирование датчика и отдельное корпусирование микросхемы, после чего они располагаются на коммутационной плате совместно с другими элементами системы. Четвертым вариантом, наиболее удобным с точки зрения последующего использования, является конструкция, при которой бескорпусной капсулированный датчик и интегральная схема монтируются на коммутационной плате и размещаются в одном герметизируемом корпусе как отдельные элементы.

Среди возможных вариантов конструктивного исполнения ММД технология «кремний на стекле» обладает следующими положительными чертами:

- 1) технология наиболее близка к технологии кремниевой микроэлектроники и поэтому хорошо освоена;
- 2) технология обладает возможностью группового изготовления;
- 3) используемые при производстве кремниевые и стеклянные пластины выпускаются промышленностью;
- 4) под данную технологию выпускается специализированное оборудование, которое постоянно модернизируется и совершенствуется;
- 5) технология позволяет изготавливать ММД различных типов;
- 6) готовый капсулированный элемент представляет собой самостоятельный элемент монтажа, что дает возможность проводить отдельно контроль его параметров, тем самым снижая выход бракованных МЭМС-гироскопов.

Перейдем к конструкции всего МЭМС-гироскопа в конечном исполнении, который представляет из себя интеграцию капсулированного ММД и интегральной схемы сервисной электроники.

Существует несколько вариантов взаимного расположения капсулы и микросхемы. Первый вариант – это классическое планарное расположение, т.е. микромеханический преобразователь и схема сервисной электроники расположены рядом. Второй вариант исполнения представляет из себя двухъярусное размещение, т.е. в специальном корпусе внизу располагается микросхема, а наверху – капсулированный элемент. Этот вариант имеет преимущество перед первым вариантом из-за уменьшения размеров конечного изделия при небольшом увеличении высоты. Однако для него также требуется конструирование специального корпуса, что усложняет производство.

Третий вариант предполагает установку микромеханического преобразователя непосредственно на интегральную микросхему. Недостатком этой конструкции является риск повреждения микросхемы при монтаже на нее капсулы. Другие варианты конструкции невозможны из-за необходимости размещения микросхемы на дне для отвода тепла [19].

Рассмотрим подробнее второй вариант с точки зрения возможности изготовления опытных и серийных образцов МЭМС-гироскопов. Доработанный вариант (рис. 10) с добавлением промежуточной керамической платы для монтажа микромеханического преобразователя и осуществления его электрической коммутации с микросхемой имеет следующие преимущества:

- 1) возможность контроля всех составных элементов МЭМС-гироскопа перед окончательной сборкой;
- 2) возможность установки на коммутационную плату микромеханических преобразователей различных конструкций и размеров;
- 3) возможность замены капсулы датчика;
- 4) герметизация корпуса, обеспечивающая защиту от внешних воздействующих факторов микросхемы в бескорпусном исполнении и микромеханического преобразователя [20, 21].

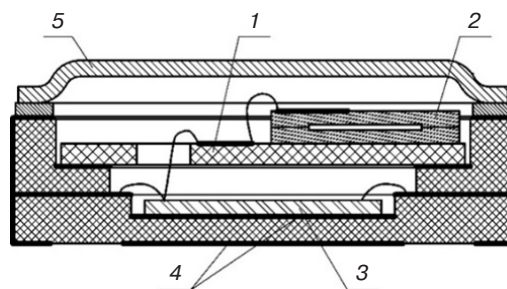


Рис. 10. Конструкция МЭМС-гироскопа:

1 – коммутационная плата,
2 – капсулированный ММД, 3 – интегральная схема,
4 – металлизация, 5 – крышка [8]

МЭМС-ГИРОСКОПЫ – КЛАСС МЕХАТРОННЫХ СИСТЕМ

Мехатроника – это область науки и техники, в основе которой лежит синергетическое объединение механики, электроники и управляющей вычислительной системы с целью проектирования и создания принципиально новых систем и модулей с интеллектуальным управлением их функциональным движением [8, 22–25]. На рис. 11 представлено схематическое изображение данного определения. По сути, МЭМС – это мехатронный узел, в котором отсутствует система управления.

Более подробное изучение конструкции МЭМС-гироскопа, в состав которого входит сервисная электроника, дает возможность говорить о нем, как о мехатронной системе. Проведем анализ изделий разработки АО «ГИРООПТИКА»¹ (Россия), представленных на рис. 12 [14, 26–31]. На рис. 13 изображена

¹ <http://gyro.ru/>. Дата обращения 22.03.2025. / Accessed March 22, 2025.

структурная схема микромеханического преобразователя угловой скорости. Как можно заметить, кроме основного ММД угловой скорости в представленный датчик еще входит дополнительный ММД линейного ускорения (акселерометр), а сервисная электроника представлена микросхемой ASIC², производства АО «ГИРООПТИКА».



Рис. 11. Схема мехатронной системы [8]

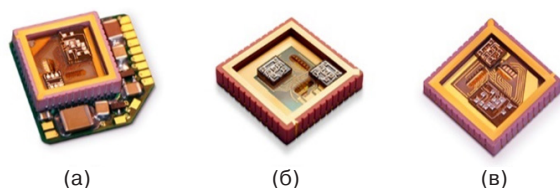


Рис. 12. Микромеханические преобразователи производства АО «ГИРООПТИКА»: (а) угловой скорости, (б) линейного ускорения, (в) комплексный преобразователь

Акселерометр в представленном МЭМС-гироскопе необходим для измерения линейного ускорения и последующей компенсации ММД угловой скорости к ускорениям.

Назначение микросхемы ASIC – это усиление, прямое цифровое преобразование

сигнала с выходов ММД угловой скорости и линейного ускорения, цифровое формирование выходных сигналов МЭМС-гироскопа и управляющих сигналов для ММД. Кроме того, в микросхему встроен процессорный блок с постоянным запоминающим устройством (ПЗУ), который обеспечивает возможность индивидуальной настройки и калибровки каждого ДУС с учетом технологического разброса параметров и их температурной зависимости. С помощью процессорного блока производится настройка ММД (подстройка частоты собственных колебаний ММД по измерительной оси относительно частоты вынужденных колебаний), коррекция нелинейности масштабного коэффициента и смещения нуля. Компенсация технологического разброса параметров и их температурной зависимости и чувствительности к перегрузке вдоль выходной оси ММД угловой скорости рассчитывается в соответствии с записанными в ПЗУ данными с учетом сигнала встроенного в микросхему термодатчика с собственным аналого-цифровым преобразователем и ММД линейного ускорения.

Комплексный микромеханический преобразователь (рис. 12в) по структуре схож с преобразователем угловой скорости, но информация о линейном ускорении используется не только для внутренней коррекции, но и выдается внешнему потребителю.

Таким образом, это вполне готовая мехатронная система, способная выполнять определенные задачи. Дальнейшее развитие микросистемной техники при использовании принципов мехатроники может привести к созданию высокоинтеллектуальных микромехатронных систем: интегральная микросхема будет управлять всей системой, микромеханические приборы станут, с одной стороны, контролировать и распознавать происходящие вокруг процессы, а, с другой стороны, станут микроминиатюрными исполнительными механизмами. Первые образцы микромехатронных роботов уже существуют [8, 25, 32, 33].

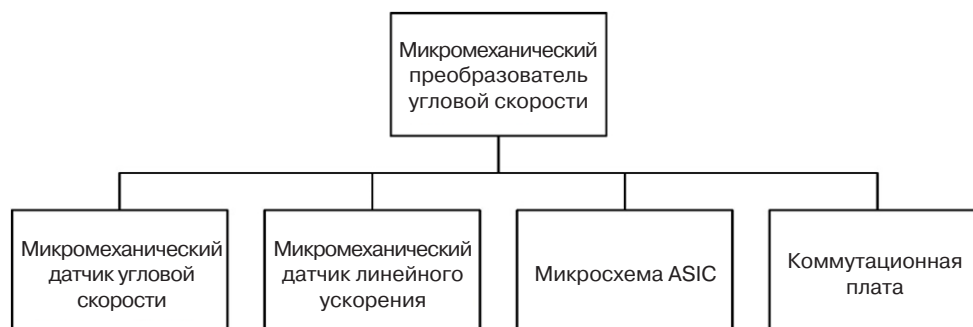


Рис. 13. Структурная схема микромеханического преобразователя угловой скорости

² Application-specific integrated circuit – интегральная схема для конкретного применения.

ОСНОВЫ ТЕХНОЛОГИИ ПРОИЗВОДСТВА МЭМС-ГИРОСКОПОВ

Производство МЭМС-гироскопа можно разделить на 4 основных процесса:

- 1) изготовление ММД в капсульном исполнении;
- 2) изготовление интегральной микросхемы, которая осуществляет обработку сигнала и управление ММД;
- 3) изготовление коммутационной платы;
- 4) сборка готового изделия.

Определяющими и наиболее сложными процессами при создании МЭМС-гироскопов являются изготовление ММД и управляющей интегральной микросхемы. Все это требует наличия специального оборудования. Процесс изготовления ASIC технологически отработан и является во многом типовым. Остановимся более подробно на производстве капсулированного ММД [34].

В основе технологии его изготовления лежит групповая технология объемной микромеханики с глубоким плазмохимическим травлением кремния и анодным соединением кремниевых и стеклянных пластин, по причине того, что конструкция требует герметичного соединения между ними. Исходными материалами в этой технологии служат кремниевые пластины с двухсторонней полировкой и стеклянные пластины такого же диаметра.

Стеклянная пластина подвергается предварительной обработке, в результате которой методом микроабразивной обработки в ней создаются сквозные отверстия для контактов в нижней пластине, а также формируются выемки глубиной порядка 50 мкм в верхней пластине. После этого на верхнюю пластину напыляют геттер, необходимый для поддержания вакуума во внутреннем объеме ММД.

Кремниевые пластины также проходят предварительную подготовку. Конструктивно необходимая толщина кремния составляет от 50 до 70 мкм. Промышленным способом такие пластины не изготавливаются из-за нестандартного размера, кроме того, требуется большая осторожность при работе с ними. В связи с этим обычно используют или стандартные кремниевые пластины (толщина 100–500 мкм для диаметра 100 мм), или пластины «кремний на изоляторе» с толщиной рабочего слоя 70 мкм и опорного кремния 500 мкм. Это позволяет использовать стандартное оборудование при обработке кремния.

На первом этапе технологического цикла производства на кремниевых пластинах проводят операции фотолитографии и глубокого плазмохимического травления кремния с целью формирования на пластине двухсторонних знаков совмещения. Следующим шагом является плазмохимическое травление полостей в кремнии. В этом процессе

в качестве маски используется оксид кремния, по которому делают фотолитографию с последующим травлением. Далее проводится плазмохимическое травление кремния на глубину 20 мкм. И хотя линейные размеры в данном рабочем слое не являются критическими, важно получить хорошую равномерность по глубине в процессе травления. После формирования структуры производится удаление оксида кремния и очистка поверхности перед операцией анодного соединения. В результате последнего процесса соединяется стеклянная пластина с отверстиями и пластина кремния. На данном этапе важно контролировать давление газа в полости, поскольку и низкий вакуум, и высокое избыточное давление могут привести к разрушению слоя кремния при последующей операции его утонения.

Суть технологического процесса утонения кремния заключается в формировании из исходной кремниевой пластины слоя кремния общей толщиной 70 мкм. После процесса утонения кремния следует операция проекционной фотолитографии. На этом шаге происходит формирование рисунка со структурой в слое кремния.

Следующая после проекционной фотолитографии операция глубокого плазмохимического травления также является ключевой, поскольку в ходе нее и формируется большая часть структуры ММД. Глубина травления составляет 50 мкм, а минимальный зазор – 2 мкм, что определяет максимальное аспектное соотношение как 1:25. Неравномерность травления, уход геометрических размеров в одну или другую сторону, наклон и шероховатость стенок – все это сказывается на качестве функционирования ММД, а большие отклонения от заданных величин могут привести к значительному ухудшению его характеристик вплоть до брака.

В процессе травления происходит формирование всех основных структур ММД, включая подвижные части и упругих элементов. С этого момента из технологического процесса должны быть исключены любые операции, которые могут повредить структуру, включая операции нанесения фоторезиста и жидкостной обработки пластин.

Следующая операция технологического процесса заключается в анодном соединении верхней стеклянной пластины с кремниевым основанием. На этом этапе происходит герметизация структуры на уровне пластины. При этом необходимо поддерживать заданный уровень вакуума в объеме ММД, что достигается за счет термической активации тонкопленочного геттера, напыленного на стекло. В процессе герметизации изделия путем анодного соединения происходят два противоположных процесса: выделение кислорода из стекла и его поглощение геттером при повышенной температуре.

Последние операции технологического процесса, которые проводятся на уровне пластин, заключаются в создании внешней металлизации. На этом этапе происходит напыление толстого слоя алюминия, который покрывает кремниевые контактные площадки на дне сквозных отверстий в стекле, а также боковые стенки этих отверстий с выводом металлизации на поверхность стекла.

Следующим этапом является резка пластин на кристаллы, которая производится при помощи установки дисковой резки пластин в два прохода. После этого пластина передается на функциональный контроль. Те чипы, которые успешно проходят тест на функционирование, передаются на следующие этапы для установки их в корпус и дальнейшей сборки преобразователя.

Рассматривая такую технологию, важно отметить, что процессы анодного соединения с предварительным совмещением соединяемых пластин (двухстороннее совмещение) и процесс сухого или глубокого плазмохимического травления кремния являются процессами, которые невозможно выполнить на типовом оборудовании производств, осуществляющих выпуск интегральных микросхем и/или полупроводниковых приборов. Остальные процессы, в принципе, возможно провести на стандартном оборудовании с соответствующими изменениями в режимах и применяемых материалах. Это касается процессов химических обработок, фотолитографических процессов, вакуумных нанесений покрытий, разделения пластин на кристаллы и др. Этому также способствует использование для изготовления микромеханических элементов стеклянных и кремниевых пластин, имеющих стандартные для полупроводниковых производств размеры (толщины и диаметр).

Изготовленные капсулированные ММД после разделения на отдельные кристаллы кроме визуального контроля проверяются на резонансные частоты и добротность. Это дает возможность провести отбраковку не только по разности выходной и входной частоты, но и по наличию необходимого вакуума внутри капсулированного элемента. Такой контроль позволяет значительно сократить трудоемкость и повысить процент выхода годных изделий на операциях сборки и настройки МЭМС-гироскопа. Здесь следует отметить, что полностью исключить плохо работающие ММД угловой скорости на ранних стадиях изготовления невозможно, т.к. требуется подключение ММД к обрабатываемой электронике и его настройка, включающая механические воздействия в виде вращений и поворотов.

Сборка МЭМС-гироскопа осуществляется путем 3D-интеграции бескорпусной интегральной микросхемы и капсулированного ММД с использованием

керамической коммутационной платы в специальный металлокерамический корпус. Интеграция осуществляется последовательным монтажом элементов в корпус с последующим их взаимными соединениями микропроволокой по методу «шарик-клин».

Монтаж кристалла ASIC, коммутационной платы и капсулированного элемента осуществляется с помощью проводящего клея, применяемого в микроэлектронике.

Герметизация корпуса осуществляется пайкой керамической крышки к основанию корпуса, проверка герметичности проводится с помощью гелиевого течеискателя по методике и критериям, применяемым для микросхем в металлокерамических корпусах.

Операции настройки и калибровки преобразователей проводятся для того, чтобы обеспечить работоспособность изделия с заданными техническими характеристиками.

ОБОРУДОВАНИЕ ПРОИЗВОДСТВА МЭМС-ГИРОСКОПОВ

Выбор в пользу технологий «кремний на стекле» или «кремний на изоляторе», сделанный в предыдущих разделах, основывался, в т.ч. и на возможности использования в технологии изготовления гироскопов промышленного оборудования. Так как выбранные технологии можно разделить на две части: технологии, перешедшие из микроэлектроники и технологии, которые присущи только изготовлению микромеханических устройств, то оборудование, обеспечивающее соответствующие технологические процессы, делится на две группы. Оборудование первой группы выпускалось промышленностью СССР, в дальнейшем – странами СНГ. Сегодня на этом рынке доминируют иностранные производители из различных стран и регионов.

Существуют широкие возможности для подбора ручного, полуавтоматического или автоматического оборудования для стандартных процессов микроэлектроники, таких как вакуумное напыление, операций фотолитографии, химических обработок, термического окисления и т.д. Особого внимания заслуживает оборудование, предназначенное для специальных процессов в рамках выбранной технологии объемной микромеханики и технологий «кремний на стекле». К таким процессам следует отнести:

- глубокое плазмохимическое травление кремния и стекла;
- двухстороннее соединение кремниевых и стеклянных пластин;
- анодное соединение кремниевых и стеклянных пластин без потери точности совмещения;
- утонение кремния на стекле.

Здесь следует заметить несколько особенностей оборудования для МЭМС:

- специализированное оборудование для операций микромеханики является высокоточным и очень дорогостоящим и выпускается производителями только под заказ и под конкретную технологию и требования заказчика;
- одно и то же в своей основе оборудование, как правило, выпускается в двух модификациях. Первая – это вариант ручного или полуавтоматического оборудования, предназначенного для научно-исследовательских и опытно-конструкторских работ, мелкосерийного производства или пилотных производств. Вторая модификация – автоматическое оборудование с загрузкой через кассеты, предназначенное для производства;
- оборудование выпускается также в кластерном варианте для объединения с установками, осуществляющими смежные операции, т.е. для создания кластера, осуществляющего автоматически целый цикл операций;
- большинство производителей в последнее время вместе с оборудованием стали предлагать и технологию, и все производители в стоимость включают пуско-наладочные работы и обучение персонала [35].

Независимо от осуществляемого технологического процесса с его конкретными особенностями, основным требованием к оборудованию производства элементов и устройств микроэлектроники и микромеханики является его способность производить с наименьшим процентом брака.

Требования к уровню привносимых загрязнений и составу остаточной газовой среды внутри рабочей камеры играют важную роль как при производстве микроэлектроники (рис. 14а – замыкание соседних дорожек), так и при производстве микромеханики (рис. 14б – микрочастица блокирует движение). На рис. 15 представлена структура оборудования высоких технологий, где средства откачки и элементы ввода движения в вакуум позволяют создавать и поддерживать сверхчистую вакуумную среду. Кроме того, они имеют способность защищать технологический объем от частиц и загрязнений, создаваемых другими элементами вакуумной системы. Связано это в первую очередь с тем, что криогенные средства откачки вообще не имеют движущихся элементов, а устройства с бесконтактным магнитным взаимодействием не имеют трущихся элементов. Принципиальные их особенности и основные свойства представлены в многочисленной специальной литературе [36–41].

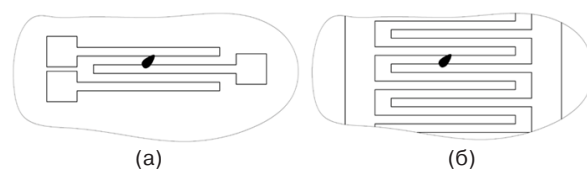


Рис. 14. Попадание микрочастиц на поверхность изделий:
(а) замыкание соседних дорожек микросхемы,
(б) блокировка движения гребенок микромеханики

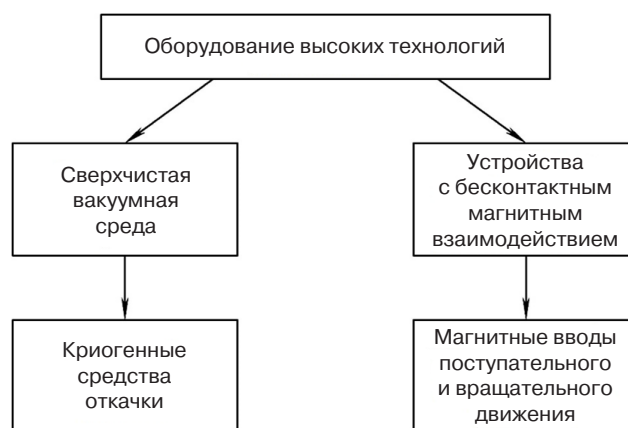


Рис. 15. Структура экологичного оборудования высоких технологий

ОПРЕДЕЛЕНИЕ ПАРАМЕТРОВ ПРИ ПРОИЗВОДСТВЕ МЭМС-ГИРОСКОПОВ

Функциональное назначение МЭМС определения угловой скорости (МЭМС-гироскопа) – преобразование неэлектрических физических величин (угловой скорости) в электрический измерительный сигнал, содержащий количественную информацию о воздействующей угловой скорости.

Основными параметрами, определяющими функциональное назначение и область применения, являются:

- диапазон измерения угловых скоростей;
- разрешающая способность;
- нелинейность масштабного коэффициента.

Основной технической характеристикой МЭМС-гироскопа является выходная (преобразовательная) характеристика, т.е. зависимость выходного сигнала от значений определяемых угловых скоростей в пределах диапазона измерения. Выходная характеристика используется в преобразовательном канале МЭМС-гироскопа, который является информационным каналом, обеспечивающим выработку информации о проекциях угловой скорости на оси чувствительности ДУС и передачу этой информации потребителю в соответствии с протоколом информационного обмена.

Поэтому необходимо учитывать погрешности выходных характеристик МЭМС-гироскопов,

которые могут возникнуть при их производстве. Погрешности делятся на две категории: основные погрешности и дополнительные погрешности. Основные погрешности определяются при нормальных условиях, т.е. при отсутствии внешних воздействующих факторов. К ним относятся нелинейность и нестабильность выходных характеристик.

Нестабильность выходной характеристики включает нестабильность смещения нуля и нестабильность масштабного коэффициента МЭМС-гироскопа.

Дополнительные погрешности возникают под воздействием внешних факторов, таких как температура окружающей среды, механические воздействия и т.д. Наибольшее влияние оказывает температурная погрешность выходной характеристики. Причина этого в том, что ММД МЭМС-гироскопов представляют собой сложную трехслойную структуру и имеют температурную зависимость своих параметров. Это, прежде всего, связано с тем, что измерительные зазоры в конденсаторах, выполненных из кремния, составляют значения 2–3 мкм, а регистрируемые минимальные перемещения имеют значения менее нанометра. При столь малых значениях и сложности микромеханической структуры даже применение дифференциальных методов измерений не позволяет исключить влияние температуры [42, 43].

Вообще температурной зависимостью обладает не только микромеханический элемент, но и электроника, обрабатывающая сигнал с емкостных датчиков и управляющая работой гироскопа. Поэтому возникает необходимость проведения настройки и калибровки МЭМС-гироскопа.

Настройка МЭМС-гироскопа является необходимой операцией для получения стабильных параметров выходной характеристики. Заключается она в задании и стабилизации в температурном диапазоне полосы пропускания и масштабного коэффициента МЭМС-гироскопа. Кроме этого, в процессе настройки определяется и компенсируется температурный дрейф смещения нуля.

Настройка проводится в климатической камере на опорно-поворотном стенде. С помощью климатической камеры задается температура в соответствии с требованиями. Опорно-поворотный стенд автоматически обрабатывает заданный набор угловых скоростей для определения масштабных коэффициентов. Для измерения температуры используется встроенный в МЭМС-гироскоп термодатчик.

Результатом настройки являются зависимости коэффициентов, отвечающих за полосу пропускания и масштабный коэффициент от показаний встроенного термодатчика. Эти зависимости, представленные в табличном (матричном) виде, используются программой управления в алгоритме кусочно-линейной аппроксимации, который вычисляет значения

коэффициентов для любого показания встроенного термодатчика.

После того, как в программу управления добавлена зависимость коэффициентов от температуры, определяется температурный дрейф смещения нуля.

Настройка МЭМС-гироскопа всегда предшествует градуировке и предназначена для обеспечения работоспособности его в диапазоне рабочих температур и угловых скоростей. В результате выполнения настройки МЭМС-гироскоп имеет технические параметры, близкие к требуемым. Окончательная доводка параметров осуществляется при градуировке.

Калибровка МЭМС-гироскопа проводится с целью определения выходных характеристик преобразовательных каналов угловой скорости в нормальных условиях (основные погрешности) и при воздействии внешних факторов (дополнительные погрешности). Точная калибровка во всем температурном диапазоне чаще всего проводится во время окончательной настройки и калибровки инерциального измерительного блока (ИИБ), в который МЭМС-гироскопы устанавливаются. Поскольку контроллер ИИБ обычно намного мощнее интегральной схемы управляющей электроники датчика, в нем осуществляется алгоритмическая компенсация всех погрешностей МЭМС-гироскопов и ИИБ.

АЛГОРИТМИЧЕСКАЯ КОМПЕНСАЦИЯ ПОГРЕШНОСТЕЙ МЭМС-ГИРОСКОПОВ

Компенсация погрешностей датчиков для нормальных условий и для каждой из температур рабочего диапазона, при которых проводится калибровка, осуществляется в контроллере ИИБ с помощью специальной программы управления, в которой используются алгоритмы компенсации погрешностей, определенные при проведении калибровки. Эти алгоритмы основаны на использовании температурных зависимостей характеристик МЭМС-гироскопов, оформленных в виде таблиц, полученных в процессе проведения калибровки.

Окончательный вариант таблицы характеристик датчиков получается при моделировании и контроле этих характеристик в специальной программе при проверке выходных характеристик ИИБ. Моделирование проводится с использованием записанных при калибровке файлов и дополнительных замеров, выполненных после калибровки. На основе полученных физических величин из выходных данных блока и их погрешностей при моделировании делается вывод о выполнении или невыполнении требований к параметрам конечного ИИБ.

Компенсация нелинейности, нестабильности и несимметрии выходных характеристик ДУС осуществляется с использованием калибровочных

характеристик, определенных при градуировке в климатической камере в диапазоне рабочих температур ИИБ и во всем диапазоне измерения угловой скорости от ± 0.01 град/с до максимального значения согласно документации.

Выходные характеристики ДУС после градуировки в климатической камере представляются в виде кусочно-линейной аппроксимации реальной зависимости выходного сигнала от задаваемого значения угловой скорости для сформированного ряда температур, устанавливаемого в климатической камере.

В результате реализации указанных алгоритмов в выходном сигнале ДУС убирается нелинейность, нестабильность и несимметрия. Тем самым масштабный коэффициент и смещение нуля делаются одинаковыми на всех температурах и при всех угловых скоростях из рабочего диапазона [44–48].

Теперь перейдем к компенсации погрешностей МЭМС-гироскопов, вызываемых вибрацией. При установке ММД на вибрирующее основание на подвижные массы действуют инерционные силы, обуславливаемые виброускорением.

У ММД увеличивается шум выходного сигнала вследствие чувствительности гироскопов к линейной перегрузке из-за конечной жесткости подвеса в синфазном направлении движения подвижных масс и технологической несимметрии подвеса. При несимметричной чувствительности гироскопов к линейному ускорению в случае вибрации может образовываться паразитное смещение их нулевого сигнала. Нелинейность чувствительности к линейному ускорению при действии постоянного ускорения (измеряемого или ускорения свободного падения) приводит к появлению несимметрии чувствительности к наложенному переменному ускорению и, соответственно, к постоянному смещению нуля гироскопов.

Дополнительная компенсация остаточных погрешностей производится алгоритмически. Влияние чувствительности ММД к линейному ускорению снижается путем их градуировки с учетом действующего гравитационного ускорения $1g$ и введения поправочных коэффициентов с привязкой к внешним акселерометрам.

ПЕРСПЕКТИВНЫЕ НАПРАВЛЕНИЯ РАЗВИТИЯ МЭМС-ГИРОСКОПОВ

Кроме положительных качеств МЭМС-гироскопов, таких как невысокая стоимость и малые габаритные размеры, есть еще негативные стороны. Данный вид преобразователей характеризуется высокой нестабильностью параметров от запуска к запуску. Смещение нуля МЭМС-гироскопа может

достигать значений, примерно 70 град/ч. Данные особенности, присущие как российским, так и зарубежным образцам, требуют проведения периодических испытаний и перекалибровок, в т.ч. обеспечения возможности проведения самокалибровки каналов в процессе эксплуатации. Все это ограничивает возможность использования МЭМС-гироскопов в аппаратуре специального назначения, требующей высокой точности. Использование их в аппаратуре другого класса может привести к ее усложнению и повышению конечной стоимости.

Точность и стабильность параметров МЭМС-гироскопов, в первую очередь, зависят от того, как происходит детектирование микроперемещений ММД. Чаще всего применяется емкостной съем информации, т.е. при перемещении изменяется емкость между неподвижными и подвижными частями (электродами) ММД, предназначенными для этого. При этом происходит взаимовлияние цепей управления и детектирования полезного сигнала датчика.

Также на точность выходных параметров МЭМС-гироскопа существенное влияние оказывает отношение сигнал/шум. Попытки ликвидировать данную проблему, доработав конструкцию ММД, приводят к возникновению противоречий. Чтобы помехоустойчивость ММД стала выше, необходимо увеличить начальную емкость. Это приводит к увеличению площади электродов и уменьшения зазора между ними, что приводит к увеличению демпфирования подвижных узлов. Для того чтобы это компенсировать, необходимо провести перфорацию кремниевой структуры ММД, что в свою очередь приведет к уменьшению площади электродов, а, следовательно, уменьшится начальная емкость.

Улучшение параметров ММД возможно, если для детектирования микроперемещений использовать оптические технологии. Совместное использование МЭМС и микрооптики может привести к синергетическому эффекту, способному решить многие проблемы. Микрооптоэлектромеханический (МОЭМ) преобразователь представляет собой миниатюрное устройство, которое осуществляет измерение и последующую обработку оптического сигнала при возникновении перемещений инерционной массы.

В последнее время в измерительных устройствах для съема информации все чаще стали применять возможности оптического туннельного эффекта. В его основе лежит процесс, при котором происходит проникновение света из оптически более плотной среды в оптически менее плотную среду при условии полного внутреннего отражения от границы раздела сред. При этом в оптически менее плотной

среде возникает электромагнитное поле, экспоненциально затухающее по нормали к границе раздела на расстояние, равное длине волны источника излучения. Приборы, построенные на оптическом туннельном эффекте, имеют высокую разрешающую способность, низкую температурную погрешность, высокую помехозащищенность [49–53].

Принцип действия первичного преобразователя, выполненного из плавленного кварца, основан на зависимости коэффициента отражения света структурой «среда-зазор-среда» от величины зазора [54]. Угол падения света на границу между первой средой и зазором (воздухом) выбирается таким, что при большой величине зазора происходит полное внутреннее отражение. Если величина зазора сравнима с длиной волны излучения, то его часть проходит (туннелирует) через зазор во вторую среду, и коэффициент отражения структуры «среда-зазор-среда» уменьшается. Таким образом, мощность оптического излучения, отраженного от структуры «среда-зазор-среда», несет информацию о величине зазора и, соответственно, о характере движения объекта.

На рис. 16 представлена принципиальная схема МОЭМ-детектора перемещения. Главным ЧЭ этого устройства является тонкая пластина из кварцевого стекла, на которой с помощью лазера вырезана своего рода «балка». В торец этой пластины направлен лазерный луч, который распространяясь по пластине, передает часть своей энергии на фотоприемники, которые установлены на некотором небольшом расстоянии сверху и снизу от пластины. Под действием внешних сил «балка» изгибается и появляется некоторая разность между величинами энергии, которая передается на фотоприемники, поскольку расстояние до датчиков начинает отличаться (рис. 17). По этой разности и определяют смещение балки [55].

В настоящее время развиваются технологии создания оптических компьютеров и процессоров. Оптические или фотонные вычислители – гипотетические вычислительные устройства, вычисления в которых производятся с помощью фотонов, излучаемыми лазерами или светодиодами. Большинство современных исследований направлено на замену традиционных (электронных) компонентов компьютера на их оптические эквиваленты. Важно, что частота световой волны на несколько порядков выше частоты электрических сигналов и волн, используемых в кремниевой технологии. Кроме того, поскольку длина световой волны мала, то имеется возможность обработки информации с необычайно высокой скоростью.

В большинстве работ по оптическим вычислителям [56–61] для начала обработки информации

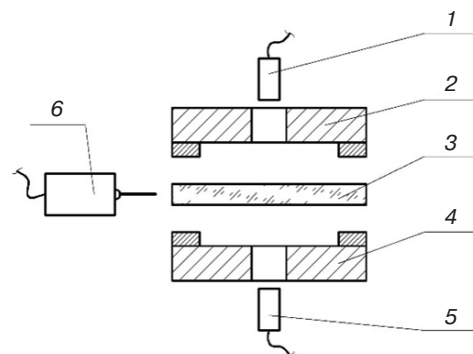


Рис. 16. Принципиальная схема МОЭМ-детектора:
1 – фотоприемник Φ_1 , 2 – крышка корпуса,
3 – кварцевая пластина (ЧЭ), 4 – основание корпуса,
5 – фотоприемник Φ_2 , 6 – лазерный источник [55]

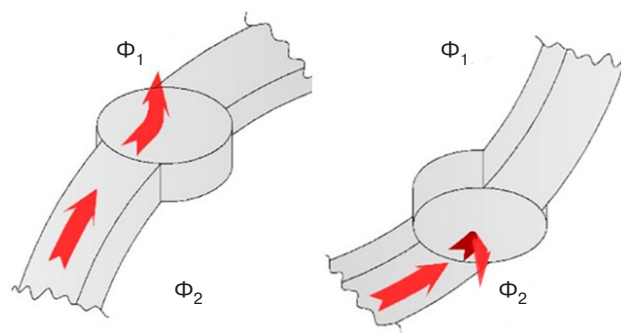


Рис. 17. Снятие показаний с ЧЭ датчика [55]

требуется ее перевод в оптический сигнал. В конструкции МОЭМ-детектора, представленного выше (рис. 16), есть два фотоприемника для съема информации и преобразования в электрический сигнал. Если же лазерное излучение напрямую посылать в приемник оптического сигнала фотонного вычислителя, то получится, что информация о колебаниях ЧЭ до процесса обработки никак не преобразуется, а, следовательно, никак не искажается, что важно для аппаратуры специального назначения.

Использование сегнетоэлектриков и сегнетоэлектрических пленок становится все популярнее в устройствах микроэлектроники, датчиках, исполнительных механизмах и т.д. Большое внимание уделяется применению сегнетоэлектрических структур и в МЭМС [62–66].

Сегнетоэлектриком называется кристаллический диэлектрик, имеющий два или большее число стабильных (или нестабильных) состояний с различной ненулевой электрической поляризацией при нулевом внешнем воздействии (электрическом поле, температуре и т.д.); эту поляризацию называют спонтанной поляризацией [67].

Применение в МЭМС-гироскопах при детектировании перемещений сегнетоэлектрических пленок имеет ряд преимуществ по сравнению

с классическими тензорезистивными и емкостными способами, что может позволить качественно расширить возможности датчиков. Пороговая чувствительность сенсоров динамической деформации на основе сегнетоэлектрических пленок уменьшается до $(\Delta/I) \approx 10^{-9}$. Их использование позволит увеличить чувствительность сенсоров на два порядка по сравнению с существующими аналогами. При этом такие датчики обладают долговременной стабильностью, им не требуется источник стабилизированного напряжения, т.к. они – сенсоры генераторного типа.

Создание подобных сенсоров связано с решением определенных технологических задач. Первая заключается в совмещении технологии создания сегнетоэлектрических пленок с технологией создания кремниевых механических структур. Для решения второй необходимо разработать способы создания устойчивого поляризованного состояния в пленке [63, 68].

ЗАКЛЮЧЕНИЕ

Постоянно растущие потребности в производстве МЭМС-гироскопов и других ММД способствуют бурному развитию микросистемной техники. Весь процесс создания конкретного изделия требует постоянного управления, начиная с разработки конструкции, где необходимо точно рассчитать все входящие элементы, и заканчивая настройкой и калибровкой.

Кроме этого, немаловажным фактором при производстве является грамотная организация технологического процесса создания ЧЭ, что включает в себя сами операции, выбор и эксплуатацию специального вакуумного оборудования.

Несмотря на все имеющиеся сложности, появляется все больше разнообразных типов МЭМС-гироскопов, работающих на новых принципах. Это требует разработки прогрессивных технологий их производства, нового специализированного оборудования и методов настройки.

СПИСОК ЛИТЕРАТУРЫ

1. Мальцев П.П., Телец В.А. Лекция 13. Микросистемная техника. В кн.: *Базовые лекции по электронике*: в 2-х т. Т. II. *Твердотельная электроника*. М.: Техносфера; 2009. С. 499–575.
2. Peshekhonov V.G. The Outlook for Gyroscopy. *Gyroscopy Navig.* 2020;11:193–197. <https://doi.org/10.1134/S2075108720030062>
3. Damianos D., Mouly J., Delbos P. *Status of the MEMS Industry 2021*. Market & Technology Report; 2021. 10 p.
4. Синельников А.О., Тихменев Н.В., Ушанов А.А., Медведев В.М. Современное состояние и тенденции развития инерциальных навигационных систем на кольцевых лазерных гироскопах. *Фотоника*. 2024;18(6):450–466. <http://doi.org/10.22184/1993-7296.FRos.2024.18.6.450.466>
5. Robin L., Perlmutter M. Gyroscopes and IMUs for Defense Aerospace and Industrial. *Report by Yole Development*. 2012. 317 p.
6. Пыжова Е.М. Микромеханические гироскопы в системах навигации. В сб.: *Прикладная электродинамика, фотоника и живые системы – 2019: материалы конференции*. Казань: ИП Сагеева А.Р.; 2019. С. 669–671. <https://www.elibrary.ru/agslgx>
7. Вернер В.Д., Сауров А.Н., Иванов А.А., Телец В.А., Коломенская Н.Г., Лучинин В.В., Мальцев П.П., Попова И.В. Изделия микросистемной техники – основные понятия и термины. *Нано- и микросистемная техника*. 2007;12:2–5.
8. Кузнецов П.С. Вопросы и перспективы развития мехатроники и микросистемной техники. *Нано- и микросистемная техника*. 2024;26(4):159–169. <https://doi.org/10.17587/nmst.25.159-169>
9. Распопов В.Я., Никулин А.В., Лихошерст В.В. Классификация конструкций микромеханических гироскопов. *Известия высших учебных заведений. Приборостроение*. 2005;48(8):5–8.
10. Вавилов В.Д., Тимошенко С.П., Тимошенко А.С. *Микросистемные датчики физических величин*: монография в двух частях. М.: ТЕХНОСФЕРА; 2018. 550 с.
11. Тимошенко С.П., Михеев А.В., Каменский А.М., Артемов Е.И., Полушкин В.М., Петрова Н.А., Боев Л.Р., Пузиков В.В. Инерциальные микроэлектромеханические системы (акселерометры и гироскопы). *Наноиндустрия*. 2020;(S96-2):471–474. <https://doi.org/10.22184/1993-8578.2020.13.3s.471.474>
12. Asar C., Shkel A. *MEMS Vibratory Gyroscopes. Structural Approaches to Improve Robustness*. Springer; 2009. 260 p. <https://doi.org/10.1007/978-0-387-09536-3>
13. Лукьянов Д.П., Распопов В.Я., Филатов Ю.В. *Прикладная теория гироскопов*. СПб.: ЦНИИ «Электроприбор»; 2015. С. 42–46.
14. Шахнович И.В. МЭМС-гироскопы – единство выбора. *Электроника: Наука, Технология, Бизнес*. 2007;1:76–85.
15. Xinfeng Z., Jingjing F., Zhipeng L., Mengtong Z. MEMS Gyroscopes Development and Application Overview on Intelligent Vehicles. In: *2020 Chinese Control and Decision Conference (CCDC)*. Hefei, China. 2020. P. 53–59. <https://doi.org/10.1109/CCDC49329.2020.9164093>
16. Zhanshe G., Fucheng C., Boyu L., et al. Research development of silicon MEMS gyroscopes: a review. *Microsyst. Technol.* 2015;21(10):2053–2066. <https://doi.org/10.1007/s00542-015-2645-x>

17. Коновалов С.Ф., Пономарев Ю.А., Майоров Д.В. Магнитная компенсация нулевого сигнала в гибридном двухкоординатном МЭМС-гироскопе R-R-R-типа. *Вестник Московского государственного технического университета им. Н.Э. Баумана. Серия Приборостроение*. 2013;4(93):122–131.
18. Каликанов А.В. Расчет чувствительного элемента микромеханического гироскопа RR-типа. *Молодой ученый*. 2020;1(291):28–33.
19. Бочаров Л.Ю., Мальцев П.П. Состояние и перспективы развития микроэлектромеханических систем за рубежом. *Микросистемная техника*. 1999;1:41–46.
20. Попова И.В., Лестев А.М., Семенов А.А., Иванов В.А., Ракитянский О.И. *Микромеханический гироскоп-акселерометр*: пат. 84542 РФ. Заявка № 20009105376/22; заявл. 16.02.2009; опубл. 10.07.2009.
21. Попова И.В., Лестев А.М., Семенов А.А., Иванов В.А., Ракитянский О.И., Бурцев В.А. Капсулированные микромеханические гироскопы и акселерометры для систем навигации и управления. *Гироскопия и навигация*. 2008;3(62):27–36.
22. Подураев Ю.В. Актуальные проблемы мехатроники. *Мехатроника, автоматизация, управление*. 2007;4:50–53.
23. Филимонов Н.Б. Предмет современной мехатроники и ее место в системе наук. *Многоядерные процессоры, параллельное программирование, ПЛИС, системы обработки сигналов*. 2017;1(7):151–159.
24. Егоров О.Д., Подураев Ю.В. Анализ и синтез интегрированных мехатронных модулей. *Мехатроника, автоматизация, управление*. 2005;10:3–8.
25. Bradley D. Mechatronics – More questions than answers. *Mechatronics*. 2010;20(8):827–841. <https://doi.org/10.1016/j.mechatronics.2010.07.011>
26. Bogolyubov V., Bakhtieva L. Astatic Gyrocompass Based on a Hybrid Micromechanical Gyroscope. In: *IEEE East-West Design and Test Symposium, (EWDTS) – Proceedings*. 2021. P. 1–5. <https://doi.org/10.1109/EWDTS52692.2021.9580982>
27. Евстифеев М.И. Основные этапы разработки отечественных микромеханических гироскопов. *Известия высших учебных заведений. Приборостроение*. 2011;54(6):75–80.
28. Евстифеев М.И. Опыт разработки микромеханических гироскопов. В сб.: *Навигация и управление движением: материалы докладов X Юбилейной конференции молодых ученых*. СПб.: ЦНИИ «Электроприбор»; 2008. С. 9–20.
29. Моисеев Н.В., Некрасов Я.А. Анализ систем управления вторичными колебаниями современных микромеханических гироскопов. *Известия Тульского государственного университета. Технические науки*. 2012;7:115–126.
30. Бекмачев А., Попова Н. Инерциальные МЭМС-датчики производства АО «ГИРООПТИКА». *Современная электроника*. 2018;5:4–6.
31. Евстафьев С.Д., Ракитянский О.И., Северов Л.А., Семенов А.А. Калибровка информационных характеристик микромеханического гироскопа. *Известия Тульского государственного университета. Технические науки*. 2012;7:167–172.
32. Новиков В.Ю., Прилуцкий И.С. Перспективы развития микромеханических и наномеханических устройств. *Электронные средства и системы управления*. Материалы докладов IV Международной научно-практической конференции, 13–16 октября 2010 г. Томск: В-Спектр; 2011;2:109–113.
33. Теряев Е.Д., Филимонов Н.Б. Наномехатроника: состояние, проблемы, перспективы. *Мехатроника, автоматизация, управление*. 2010;1:2–14.
34. Кузнецов П.С. К вопросу технологии создания твердотельных микромеханических систем. *Естественные и технические науки*. 2024;3(190):136–139. <https://doi.org/10.25633/ETN.2024.03.09>
35. Ивашов Е.Н., Кузнецов П.С., Федотов К.Д. Оптимизация управления параметрами метрологического обеспечения при производстве микромеханических гироскопов. *Вестник машиностроения*. 2015;5:40–45.
36. Васин В.А., Ивашов Е.Н., Кузнецов П.С., Степанчиков С.В. Монография 6. Микроинженерия магнитных устройств. *Микро- и наноинженерия в электронном машиностроении: Серия из 7 монографий*. Ивanteevka Моск. обл.: Изд-во НИИ предельных технологий; 2013. 205 с.
37. Васин В.А., Ивашов Е.Н., Степанчиков С.В. Повышение равномерности нанесения тонких пленок в вакууме. *Технология машиностроения*. 2011;6:27–30.
38. Васин В.А., Ивашов Е.Н., Кузнецов П.С., Степанчиков С.В. Устройства с бесконтактным магнитным взаимодействием для специального технологического оборудования. *Технология машиностроения*. 2011;2:47–51.
39. Васин В.А., Ивашов Е.Н., Кузнецов П.С., Степанчиков С.В. Особенности применения устройств с бесконтактным магнитным взаимодействием в современном сверхвысоковакуумном контрольно-диагностическом и технологическом оборудовании. *Контроль. Диагностика*. 2011;2:44–48.
40. Васин В.А., Ивашов Е.Н., Степанчиков С.В. Контрольно-навигационная система современного сверхвысоковакуумного аналитико-технологического комплекса. *Контроль. Диагностика*. 2012;5:71–74.
41. Ивашов Е.Н., Лучников А.П., Сигов А.С., Степанчиков С.В. *Проектирование элементов и устройств технологических систем электронной техники*. М.: Энергоатомиздат; 2008. 288 с.
42. Саленко Д.С. Особенности моделирования MEMS-гироскопа. *Автоматика и программная инженерия*. 2015;2(12):109–115.
43. Кузнецов П.С. Определение параметров при производстве МЭМС-гироскопов. *Инновации на основе информационных и коммуникационных технологий*. 2014;1:456–460. <https://www.elibrary.ru/ukrxnl>
44. Кузнецов П.С. Компенсация нелинейности МЭМС-гироскопов с помощью математического моделирования. *Инновации на основе информационных и коммуникационных технологий*. 2014;1:460–462. <https://www.elibrary.ru/ukrxnv>
45. Бойков И.В., Кривулин Н.П. Идентификация дискретных динамических систем с распределенными параметрами. *Известия высших учебных заведений. Поволжский регион. Физико-математические науки*. 2014;2(30):34–48.

46. Крылов А.А., Корниук Д.В. Технологические подходы к устранению смещения нуля МЭМС гироскопов в составе гироскопического блока. *Труды МАИ*. 2018;103:18.
47. Крылов А.А., Кузнецов П.С. Устранение смещения нуля МЭМС-гироскопов при различной температурной динамике. *Вестник Концерна ВКО Алмаз-Антей*. 2019;2(29):34–39.
48. Крылов А.А. Исследование неустойчивости дрейфа нуля МЭМС-гироскопов и способов ее учета при калибровке. *Известия Тульского государственного университета. Технические науки*. 2020;1:64–69.
49. Бусурин В.И., Васецкий С.О., Казарьян А.В. Компенсационный микрооптоэлектромеханический датчик угловой скорости: пат. 2806242 РФ. Заявка № 2023123201А; заявл. 06.09.2023; опубл. 30.10.2023.
50. Xia D., Huang L., Zhao L. A New Design of an MOEMS Gyroscope Based on a WGM Microdisk Resonator. *Sensors*. 2019;19(12):2798. <https://doi.org/10.3390/s19122798>
51. Barbin E., Nesterenko T., Koleda A., Shesterikov E., Kulinich I., Kokolov A. An Optical Measuring Transducer for a Micro-Opto-Electro-Mechanical Micro-g Accelerometer Based on the Optical Tunneling Effect. *Micromachines*. 2023;14(4):802. <https://doi.org/10.3390/mi14040802>
52. Бусурин В.И., Казарьян А.В., Штек С.Г., Жеглов М.А., Васецкий С.О., Чжи П.Л. Рамочный микрооптоэлектромеханический преобразователь угловой скорости с узлами оптического считывания на основе оптического туннельного эффекта. *Измерительная техника*. 2022;5:50–55. <https://doi.org/10.32446/0368-1025it.2022-5-50-55>
53. Бусурин В.И., Васецкий С.О., Штек С.Г., Жеглов М.А. Микрооптоэлектромеханический датчик угловой скорости: пат. 2790042 RF. Заявка № 2022129761; заявл. 16.11.2022; опубл. 14.02.2023.
54. Barbin E., Nesterenko T., Koleda A., Shesterikov E., Kulinich I., Kokolov A., Perin A. The Design, Modeling and Experimental Investigation of a Micro-G Microoptoelectromechanical Accelerometer with an Optical Tunneling Measuring Transducer. *Sensors*. 2024;24(3):765. <https://doi.org/10.3390/s24030765>
55. Shtek S.G., Zheglov M.A., Belyakov V.V., Andreasyan O.G., Vasetskiy S.O., Kuznetsov P.S. Development of a Sensitive Element of a Micro-Opto-Electromechanical Accelerometer. In: *2023 30th Saint Petersburg International Conference on Integrated Navigation Systems (ICINS)*. <https://doi.org/10.23919/ICINS51816.2023.10168511>
56. Шмидт С.П., Иванов В.В. Перспективы создания и развития полностью оптического компьютера. *Инновационная наука*. 2015;8-2(8):83–85.
57. Молодяков С.А. Оптоэлектронные процессоры с ПЗС-фотоприемниками. Конвейерная обработка сигналов. *Информационно-управляющие системы*. 2008;6:2–8.
58. Степаненко С.А. Фотонный компьютер: структура и алгоритмы, оценка параметров. *Фотоника*. 2017;7(67):72–83. <https://doi.org/10.22184/1993-7296.2017.67.7.72.83>
59. Умарова У.Б. Оптические компьютеры и их достоинства. В сб.: *Современные инструментальные системы, информационные технологии и инновации: сборник научных трудов XI Международной научно-практической конференции*: в 4-х т. Курск: ЗАО «Университетская книга»; 2014. С. 219–220.
60. Гордеев А., Войтович В., Святец Г. Перспективные фотонные и фононные отечественные технологии для терагерцевых микропроцессоров, ОЗУ и интерфейса со сверхнизким энергопотреблением. *Современная электроника*. 2022;2:65–72.
61. Rawat U., Anderson J.D., Weinstein D. Design and Applications of Integrated Transducers in Commercial CMOS Technology. *Front. Mech. Eng.* 2022;8:902421. <https://doi.org/10.3389/fmech.2022.902421>
62. Таишев С.Р., Крайнова К.Ю. Оптимизация условий поляризации в наноразмерных сегнетоэлектриках с целью дальнейшего применения в датчиках и МЭМС. *Молодой ученый*. 2016;1(105):224–227.
63. Воротилов К.А., Мухортов В.М., Сигов А.С. *Интегрированные сегнетоэлектрические устройства*. М.: Энергоатомиздат; 2011. 175 с.
64. Li S., Wang Y., Yang M., et al. Ferroelectric thin films: performance modulation and application. *Mater. Adv.* 2022;3(14):5735–5752. <https://doi.org/10.1039/D2MA00381C>
65. Tadigadapa S. Piezoelectric microelectromechanical systems – challenges and opportunities. *Procedia Eng.* 2010;5:468–471. <https://doi.org/10.1016/j.proeng.2010.09.148>
66. Мухортов В.М., Головки Ю.И., Бирюков С.В., Масычев С.И., Павленко А.В., Стрюков Д.В., Зинченко С.П., Ковтун А.П., Толмачев Г.Н. Наноразмерные сегнетоэлектрические пленки – новая активная среда для микроэлектроники. *Наука Юга России*. 2022;18(4):33–43. <https://doi.org/10.7868/S25000640220404>
67. *Физика сегнетоэлектриков: современный взгляд*; под ред. Рабе К.М., Ана Ч.Г., Трискона Ж.М.: пер. с англ. М.: Лаборатория знаний; 2020. 443 с. ISBN 978-5-00101-827-8
68. Воротилов К.А., Сигов А.С., Романов А.А., Машевич П.Р. Формирование сегнетоэлектрических наноструктур для нового поколения устройств микро- и наноэлектроники. *Наноматериалы и наноструктуры – XXI век*. 2010;1(1):45–53.

REFERENCES

1. Maltsev P.P., Telets V.A. Lecture 13. Microsystem technology. In: *Bazovye lektsii po elektronike (Basic Lectures on Electronics)*: in 2 v. V. II. *Solid-State Electronics*. Moscow: Tekhnosfera; 2009. P. 499–575 (in Russ.).
2. Peshekhonov V.G. The Outlook for Gyroscopy. *Gyroscopy Navig.* 2020;11:193–197. <https://doi.org/10.1134/S2075108720030062>
3. Damianos D., Mouly J., Delbos P. *Status of the MEMS Industry 2021*. Market & Technology Report; 2021. 10 p.

4. Sinelnikov A.O., Tikhmenev N.V., Ushanov A.A., Medvedev V.M. State-of-the-Art and Development Trends of Inertial Navigation Systems Based on the Ring Laser Gyroscopes. *Fotonika = Photonics Russia*. 2024;8(6):450–466. <http://doi.org/10.22184/1993-7296.FRos.2024.18.6.450.466>
5. Robin L., Perlmutter M. Gyroscopes and IMUs for Defense Aerospace and Industrial. *Report by Yole Development*. 2012. 317 p.
6. Pyzhova E.M. Micromechanical gyroscopes in navigation systems. In: *Prikladnaya elektrodinamika, fotonika i zhivye sistemy – 2019 (Applied Electrodynamics, Photonics and Living Systems – 2019). Proceedings*. Kazan: Individual Entrepreneur Sagieva A.R.; 2019. P. 669–671 (in Russ.). <https://www.elibrary.ru/agslgx>
7. Verner V.D., Ivanov A.A., Kolomenskaya N.G., Luchinin V.V., Maljtcev P.P., Popova I.V., Saurov A.N., Teletc V.A. The Produces of Microsystems Techniques – Fundamental Ideals and Terms. *Nano- i mikrosistemnaja tehnika = Nano- and Microsystem Technique*. 2007;12:2–5 (in Russ.).
8. Kuznetsov P.S. Issues and prospects for the development of mechatronics and microsystem technology. *Nano- i mikrosistemnaja tehnika = Nano- and Microsystem Technique*. 2024;26(4):159–169 (in Russ.). <https://doi.org/10.17587/nmst.25.159-169>
9. Raspopov V.Ya., Nikulin A.V., Likhoshurst V.V. Classification of designs of micromechanical gyros. *Izvestiya vysshikh uchebnykh zavedenii. Priborostroenie = J. Instrument. Eng.* 2005;48(8):5–8 (in Russ.).
10. Vavilov V.D., Timoshenkov S.P., Timoshenkov A.S. *Mikrosistemnye datchiki fizicheskikh velichin (Microsystem Sensors of Physical Quantities)*. Monograph in two parts. Moscow: TEKhNOSFERA; 2018. 550 p. (in Russ.).
11. Timoshenkov S.P., Mikheev A.V., Kamenskii A.M., Artemov E.I., Polushkin V.M., Petrova N.A., Boev L.R., Puzikov V.V. Inertial microelectromechanical systems (accelerometers and gyroscopes). *Nanoindustriya = Nanoindustry*. 2020;(S96-2): 471–474 (in Russ.). <https://doi.org/10.22184/1993-8578.2020.13.3s.471.474>
12. Asar C., Shkel A. *MEMS Vibratory Gyroscopes. Structural Approaches to Improve Robustness*. Springer; 2009. 260 p. <https://doi.org/10.1007/978-0-387-09536-3>
13. Lukyanov D.P., Raspopov V.Ya., Filatov Yu.V. *Prikladnaya teoriya giroskopov (Applied Theory of Gyroscopes)*. St. Petersburg: Electropribor; 2015. P. 42–46 (in Russ.).
14. Shakhnovich I.V. MEMS-gyroscopes – unity of choice. *Elektronika: Nauka, Tekhnologiya, Biznes = Electronics: Science, Technology, Business*. 2007;1:76–85 (in Russ.).
15. Xinfeng Z., Jingjing F., Zhipeng L., Mengtong Z. MEMS Gyroscopes Development and Application Overview on Intelligent Vehicles. In: *2020 Chinese Control and Decision Conference (CCDC)*. Hefei, China. 2020. P. 53–59. <https://doi.org/10.1109/CCDC49329.2020.9164093>
16. Zhanshe G., Fucheng C., Boyu L., et al. Research development of silicon MEMS gyroscopes: a review. *Microsyst. Technol.* 2015;21(10):2053–2066. <https://doi.org/10.1007/s00542-015-2645-x>
17. Konovalov S.F., Ponomarev Yu.A., Mayorov D.V. Magnetic compensation of the zero signal in a two-axis hybrid R-R-R type MEMS gyro. *Vestnik Moskovskogo gosudarstvennogo tekhnicheskogo universiteta im. N.E. Baumana. Seriya Priborostroenie = Herald of the Bauman Moscow State Technical University. Series Instrument Engineering*. 2013;4(93):122–131 (in Russ.).
18. Kalikanov A.V. Calculation of the sensing element of an RR-type micromechanical gyroscope. *Molodoi uchenyi = Young Scientist*. 2020;1(291):28–33 (in Russ.).
19. Bocharov L.Yu., Maltsev P.P. Status and development prospects of microelectromechanical systems abroad. *Mikrosistemnaya tekhnika = Microsystem Technology*. 1999;1:41–46 (in Russ.).
20. Popova I.V., Lestev A.M., Semenov A.A., Ivanov V.A., Rakityanskii O.I. *Micromechanical Gyroscope-Accelerometer: RF Pat. 84542*. Publ. 10.07.2009 (in Russ.).
21. Popova I.V., Lestev A.M., Semenov A.A., Ivanov V.A., Rakityanskii O.I., Burtsev V.A. Encapsulated micromechanical gyroscopes and accelerometers for navigation and control systems. *Giroskopiya i navigatsiya = Gyroscopy and Navigation*. 2008;3(62):27–36 (in Russ.).
22. Poduraev Yu.V. Actual problems of mechatronics. *Mekhatronika, avtomatizatsiya, upravlenie = Mechatronics, Automation, Control*. 2007;4:50–53 (in Russ.).
23. Filimonov N.B. The subject of modern mechatronics and its place in the system of sciences. *Mnogoyadernye protsessory, parallel'noe programmirovaniye, PLIS, sistemy obrabotki signalov = Multicore Processors, Parallel Programming, FPGAs, Signal Processing Systems*. 2017;1(7):151–159 (in Russ.).
24. Egorov O.D., Poduraev Yu.V. Analysis and synthesis of integrated mechatronic modules. *Mekhatronika, avtomatizatsiya, upravlenie = Mechatronics, Automation, Control*. 2005;10:3–8 (in Russ.).
25. Bradley D. Mechatronics – More questions than answers. *Mechatronics*. 2010;20(8):827–841. <https://doi.org/10.1016/j.mechatronics.2010.07.011>
26. Bogolyubov V., Bakhtieva L. Astatic Gyrocompass Based on a Hybrid Micromechanical Gyroscope. In: *IEEE East-West Design and Test Symposium, (EWDTS) – Proceedings*. 2021. P. 1–5. <https://doi.org/10.1109/EWDTS52692.2021.9580982>
27. Evstifeev M.I. Principal stages of development of domestic micromechanical gyroscopes. *Izvestiya vysshikh uchebnykh zavedenii. Priborostroenie = J. Instrument. Eng.* 2011;54(6):75–80 (in Russ.).
28. Evstifeev M.I. Experience in the development of micromechanical gyroscopes. In: *Navigation and Motion Control: Materials of the Reports of the 10th Anniversary Conference of Young Scientists*. St. Petersburg: Electropribor; 2008. P. 9–20 (in Russ.).
29. Moiseev N.V., Nekrasov Ya.A. Analysis of control systems of mems gyroscope channel. *Izvestiya Tul'skogo gosudarstvennogo universiteta. Tekhnicheskie nauki = News of the Tula State University. Technical Sciences*. 2012;7:115–126 (in Russ.).

30. Bekmachev A., Popova N. Inertial MEMS sensors manufactured by JSC GYROOPTIKA. *Sovremennaya elektronika = Modern Electronics*. 2018;5:4–6 (in Russ.).
31. Evstafiev S.D., Rakityanskii O.I., Severov L.A., Semenov A.A. Calibration of information characteristics MMG LL type. *Izvestiya Tul'skogo gosudarstvennogo universiteta. Tekhnicheskie nauki = News of the Tula State University. Technical Sciences*. 2012;7:167–172 (in Russ.).
32. Novikov V.Yu., Prilutskii I.S. Prospects for the development of micromechanical and nanomechanical devices. *Elektronnyye sredstva i sistemy upravleniya (Electronic Means and Control Systems)*. Materials of the Reports of the 4th International Scientific and Practical Conference. October 13–16, 2010. Tomsk: V-Spektr; 2011;2:109–113 (in Russ.).
33. Teryaev E.D., Filimonov N.B. Nanomechatronics: state, problems, prospects. *Mekhatronika, avtomatizatsiya, upravlenie = Mechatronics, Automation, Control*. 2010;1:2–14 (in Russ.).
34. Kuznetsov P.S. On the issue of technology for creating solid-state micromechanical systems. *Estestvennye i tekhnicheskie nauki = Natural and Technical Sciences*. 2024;3(190):136–139 (in Russ.). <https://doi.org/10.25633/ETN.2024.03.09>
35. Ivashov E.N., Kuznetsov P.S., Fedotov K.D. Optimization of control of metrological support parameters in the production of micromechanical gyroscopes. *Vestnik Mashinostroeniya*. 2015;5:40–45 (in Russ.).
36. Vasin V.A., Ivashov E.N., Kuznetsov P.S., Stepanchikov S.V. Monograph 6. Microengineering of magnetic devices. In: *Mikro- i nanoinzheneriya v elektronnom mashinostroenii (Micro- and Nanoengineering in Electronic Engineering): A series of 7 monographs*. Ivanteevka: Research Institute of Extreme Technologies; 2013. 205 p. (in Russ.).
37. Vasin V.A., Ivashov E.N., Stepanchikov S.V. Increasing the uniformity of thin film deposition in vacuum. *Tekhnologiya Mashinostroeniya*. 2011;6:27–30 (in Russ.).
38. Vasin V.A., Ivashov E.N., Kuznetsov P.S., Stepanchikov S.V. Devices with contactless magnetic interaction for special technological equipment. *Tekhnologiya Mashinostroeniya*. 2011;2:47–51 (in Russ.).
39. Vasin V.A., Ivashov E.N., Kuznetsov P.S., Stepanchikov S.V. The features of application of devices with contactless magnetic interaction in modern ultrahigh vacuum control and diagnostic and technological equipment. *Kontrol'. Diagnostika = Testing. Diagnostics*. 2011;2:44–48 (in Russ.).
40. Vasin V.A., Ivashov E.N., Stepanchikov S.V. Control and navigation system of the modern ultrahigh vacuum analytical and processing complex. *Kontrol'. Diagnostika = Testing. Diagnostics*. 2012;5:71–74 (in Russ.).
41. Ivashov E.N., Luchnikov A.P., Sigov A.S., Stepanchikov S.V. *Proektirovanie elementov i ustroystv tekhnologicheskikh sistem elektronnoi tekhniki (Design of Elements and Devices of Technological Systems of Electronic Equipment)*. Moscow: Energoatomizdat; 2008. 288 p. (in Russ.).
42. Salenko D.S. Features of MEMS gyroscope modeling. *Avtomatika i programmnaya inzheneriya = Automatics & Software Engineering*. 2015;2(12):109–115 (in Russ.).
43. Kuznetsov P.S. Determination of parameters in the production of MEMS gyroscopes. *Innovatsii na osnove informatsionnykh i kommunikatsionnykh tekhnologii = Innovations Based on Information and Communication Technologies*. 2014;1:457–460 (in Russ.). <https://www.elibrary.ru/ukrxnv>
44. Kuznetsov P.S. Compensation of the nonlinearity of MEMS gyroscopes using mathematical modeling. *Innovatsii na osnove informatsionnykh i kommunikatsionnykh tekhnologii = Innovations Based on Information and Communication Technologies*. 2014;1:460–462 (in Russ.). <https://www.elibrary.ru/ukrxnv>
45. Boikov I.V., Krivulin N.P. Identification of discrete dynamical systems with distributed parameters. *Izvestiya vysshikh uchebnykh zavedenii. Povolzhskii region. Fiziko-matematicheskie nauk = University Proceedings. Volga region. Physical and Mathematical Sciences*. 2014;2(30):34–48 (in Russ.).
46. Krylov A.A., Korniyuk D.V. Technological approaches to zero offset compensation in MEMS gyroscopes being a part of the inertial measurement unit. *Trudy MAI*. 2018;103:18 (in Russ.).
47. Krylov A.A., Kuznetsov P.S. MEMS gyroscope zero drift elimination at different temperature dynamics. *Vestnik Kontserna VKO Almaz-Antei*. 2019;2(29):34–39 (in Russ.).
48. Krylov A.A. Research of MEMS gyroscopes zero drift instability and ways of its accounting at calibration. *Izvestiya Tul'skogo gosudarstvennogo universiteta. Tekhnicheskie nauki = News of the Tula State University. Technical Sciences*. 2020;1:64–69 (in Russ.).
49. Busurin V.I., Vasetsky S.O., Kazaryan A.V. *Compensating Micro-optoelectromechanical Angular Velocity Sensor*: RF Pat. 2806242. Publ. 30.10.2023 (in Russ.).
50. Xia D., Huang L., Zhao L. A New Design of an MOEMS Gyroscope Based on a WGM Microdisk Resonator. *Sensors*. 2019;19(12):2798. <https://doi.org/10.3390/s19122798>
51. Barbin E., Nesterenko T., Koleda A., Shesterikov E., Kulinich I., Kokolov A. An Optical Measuring Transducer for a Micro-Opto-Electro-Mechanical Micro-g Accelerometer Based on the Optical Tunneling Effect. *Micromachines*. 2023;14(4):802. <https://doi.org/10.3390/mi14040802>
52. Busurin V.I., Kazaryan A.V., Shtek S.G., Zheglov M.A., Vasetskiy S.O., Ky P.L. Frame microoptoelectromechanical angular velocity transducer with optical readout units based on optical tunneling effect. *Izmeritel'naya Tekhnika*. 2022;5:50–55 (in Russ.). <https://doi.org/10.32446/0368-1025it.2022-5-50-55>
53. Busurin V.I., Vasetskii S.O., Shtek S.G., Zheglov M.A. *Micro-optoelectromechanical Angular Velocity Sensor*: RF Pat. 2790042. Publ. 14.02.2023 (in Russ.).
54. Barbin E., Nesterenko T., Koleda A., Shesterikov E., Kulinich I., Kokolov A., Perin A. The Design, Modeling and Experimental Investigation of a Micro-G Microoptoelectromechanical Accelerometer with an Optical Tunneling Measuring Transducer. *Sensors*. 2024;24(3):765. <https://doi.org/10.3390/s24030765>

55. Shtek S.G., Zheglov M.A., Belyakov V.V., Andreasyan O.G., Vasetskiy S.O., Kuznetsov P.S. Development of a Sensitive Element of a Micro-Opto-Electromechanical Accelerometer. In: *2023 30th Saint Petersburg International Conference on Integrated Navigation Systems (ICINS)*. <https://doi.org/10.23919/ICINS51816.2023.10168511>
56. Schmidt S.P., Ivanov V.V. Prospects for the creation and development of a fully optical computer. *Innovatsionnaya nauka = Innovative Science*. 2015;8-2(8):83–85 (in Russ.).
57. Molodyakov S.A. Optoelectronic processors with CCD photodetectors. Pipeline signal processing. *Informatsionno-upravlyayushchie sistemy = Information and Control systems*. 2008;6:2–8 (in Russ.).
58. Stepanenko S.A. Photonic computer: Structure and algorithms. Estimations of parameters. *Fotonika = Photonics*. 2017;7(67):72–3 (in Russ.). <https://doi.org/10.22184/1993-7296.2017.67.7.72.83>
59. Umarova U.B. Optical computers and their advantages. In: *Sovremennye instrumental'nye sistemy, informatsionnye tekhnologii i innovatsii (Modern Instrumental Systems, Information Technologies and Innovations)*. Collection of Scientific Papers of the 11th International Scientific and Practical Conference: in 4 v. Kursk: Universitetskaya kniga; 2014. P. 219–220 (in Russ.).
60. Gordeev A., Voitovich V., Svyatets G. Promising photonic and phonon domestic technologies for terahertz microprocessors, RAM and interface with ultra-low power consumption. *Sovremennaya elektronika = Modern Electronics*. 2022;2:65–72 (in Russ.).
61. Rawat U., Anderson J.D., Weinstein D. Design and Applications of Integrated Transducers in Commercial CMOS Technology. *Front. Mech. Eng.* 2022;8:902421. <https://doi.org/10.3389/fmech.2022.902421>
62. Taishev S.R., Krainova K.Yu. Optimization of polarization conditions in nanoscale ferroelectrics for further application in sensors and MEMS. *Molodoi uchenyi = Young Scientist*. 2016;1(105):224–227 (in Russ.).
63. Vorotilov K.A., Mukhortov V.M., Sigov A.S. *Integrirovannye segnetoelektricheskie ustroystva (Integrated Ferroelectric Devices)*. Moscow: Energoatomizdat; 2011. 175 p. (in Russ.).
64. Li S., Wang Y., Yang M., et al. Ferroelectric thin films: performance modulation and application. *Mater. Adv.* 2022;3(14): 5735–5752. <https://doi.org/10.1039/D2MA00381C>
65. Tadigadapa S. Piezoelectric microelectromechanical systems – challenges and opportunities. *Procedia Eng.* 2010;5:468–471. <https://doi.org/10.1016/j.proeng.2010.09.148>
66. Mukhortov V.M., Golovko Yu.I., Biryukov S.V., Masychev S.I., Pavlenko A.V., Stryukov D.V., Zinchenko S.P., Kovtun A.P., Tolmachev G.N. Nanoscale ferroelectric films – a new active medium for microelectronics. *Nauka Yuga Rossii = Science in the South of Russia*. 2022;18(4):33–43 (in Russ.). <https://doi.org/10.7868/S25000640220404>
67. Rabe K.M., Ahn C.H., Triscone J.M. (Eds.). *Fizika segnetoelektrikov: sovremenniy vzglyad (Physics of Ferroelectrics: A Modern Perspective)*: transl. from Engl. Moscow: Laboratoriya znaniy; 2020. 443 p. (in Russ.). ISBN 978-5-00101-827-8 [Rabe K.M., Ahn C.H., Triscone J.M. (Eds.). *Physics of Ferroelectrics. A Modern Perspective*. Berlin, Heidelberg: Springer; 2007. 402 p.]
68. Vorotilov K.A., Sigov A.S., Romanov A.A., Mashevich P.R. Nanomaterials and structures in electronics. *Nanomaterialy i nanostruktury – 21 vek = Nanomaterials and Nanostructures – 21st Century*. 2010;1(1):45–53 (in Russ.).

Об авторе

Кузнецов Павел Сергеевич, к.т.н., заместитель начальника экспериментального комплекса микроэлектроники и микромеханических систем, Акционерное общество «Государственный научно-исследовательский институт приборостроения» (АО «ГосНИИП») (129226, Россия, Москва, пр-т Мира, д. 125). E-mail: ps_kuznetsov@mail.ru. SPIN-код РИНЦ 6564-9540, <https://orcid.org/0000-0001-5459-7883>

About the Author

Pavel S. Kuznetsov, Cand. Sci. (Eng.), Deputy Head of the Experimental Complex of Microelectronics and Micromechanical Systems, State Scientific Research Institute of Instrument Engineering (GosNIIP) (125, Mira pr., Moscow, 129226 Russia). E-mail: ps_kuznetsov@mail.ru. RSCI SPIN-code 6564-9540, <https://orcid.org/0000-0001-5459-7883>

Микро- и нанoeлектроника. Физика конденсированного состояния
Micro- and nanoelectronics. Condensed matter physics

УДК 537.632
<https://doi.org/10.32362/2500-316X-2025-13-3-122-128>
EDN IJRLMF



НАУЧНАЯ СТАТЬЯ

Особенности магниторефрактивного эффекта в нанокompозитах Co–Si

А.Н. Юрасов¹,
Р. Кулгунина¹,
М.М. Яшин^{1, 2, @},
М.А. Симдянова¹

¹ МИРЭА – Российский технологический университет, Москва, 119454 Россия

² МГТУ им. Н.Э. Баумана, Москва, 105005 Россия

@ Автор для переписки, e-mail: ihkamax@mail.ru

• Поступила: 14.11.2024 • Доработана: 03.02.2025 • Принята к опубликованию: 20.03.2025

Резюме

Цели. Целью работы является исследование спектров магниторефрактивного эффекта (МРЭ) в нанокompозитах «кобальт–кремний» (Co–Si) с учетом вклада размерного эффекта, а также сравнение полученных результатов при изменении параметров размерного эффекта. Данное исследование является важным для практического применения бесконтактных методов, т.к. оно направлено на расширение их возможностей и создание новых подходов к неразрушающему контролю и исследованию магнитооптических свойств нанокompозитов, что может значительно повысить эффективность их использования в различных областях, включая спинтронику и оптику.

Методы. Применялось компьютерное моделирование в рамках перспективного метода эффективной среды – приближения Бруггемана, согласно которому исследуемая структура заменяется средой с эффективными свойствами.

Результаты. В рамках моделирования получены спектры МРЭ в диапазоне 0.5–3.5 эВ. При этом моделирование проводилось для МРЭ без учета и с учетом квазиклассического размерного эффекта. Конечным результатом стало моделирование спектральных зависимостей МРЭ на примере нанокompозита Co–Si при различных значениях размера частиц и форм-фактора кобальта. Показано влияние размерных эффектов на вид спектров МРЭ. Достоверность методик хорошо подтверждается сравнением полученных результатов с эмпирическими данными, а ценность полученных результатов обусловлена тем, что все рассчитанные параметры обсуждаемого нанокompозита и форма спектральных зависимостей МРЭ хорошо согласуются с результатами различных экспериментов.

Выводы. В рамках моделирования показано, что учет размеров и форм-фактора гранул оказывает значительное влияние на вид спектров МРЭ, демонстрируя перспективные свойства нанокompозита при определенных размерах частиц. Представленные результаты подчеркивают возможность оптимизации характеристик материала для улучшения чувствительности в магнитных сенсорах и устройствах бесконтактного исследования наноструктур.

Ключевые слова: нанокompозиты, теория эффективной среды, магниторефрактивный эффект, ферромагнетик, размерные эффекты

Для цитирования: Юрасов А.Н., Кулгунина Р., Яшин М.М., Симдянова М.А. Особенности магниторефрактивного эффекта в нанокompозитах Co–Si. *Russian Technological Journal*. 2025;13(3):122–128. <https://doi.org/10.32362/2500-316X-2025-13-3-122-128>, <https://www.elibrary.ru/IJRLMF>

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

Авторы заявляют об отсутствии конфликта интересов.

RESEARCH ARTICLE

Features of the magnetorefractive effect in Co–Si nanocomposites

Alexey N. Yurasov ¹,
Regina Kulgunina ¹,
Maxim M. Yashin ^{1, 2, @},
Marina A. Simdyanova ¹

¹ MIREA – Russian Technological University, Moscow, 119454 Russia

² Bauman Moscow State Technical University, Moscow, 105005 Russia

@ Corresponding author, e-mail: ihkamax@mail.ru

• Submitted: 14.11.2024 • Revised: 03.02.2025 • Accepted: 20.03.2025

Abstract

Objectives. The work set out to study the spectra of the magnetorefractive effect (MRE) in the cobalt–silicon (Co–Si) nanocomposite, taking into account the contribution of the size effect (SE), and to compare the results obtained by varying the parameters of the SE. The presented approaches to investigating the magneto-optical properties of nanocomposites, which are relevant for the practical application of nondestructive testing methods, have the potential to significantly increase the efficiency of their use in various fields, including spintronics and optics.

Methods. Computer modeling approaches based on the Bruggeman approximation are used to model the examined structure as a medium with effective properties.

Results. MRE spectra obtained within the framework of the modeling fell within the range of 0.5–3.5 eV. The modeling was carried out for MRE both with and without taking into account the semiclassical size effect. The resultant modeling of the spectral dependencies of the MRE is based on the example of a Co–Si nanocomposite at different cobalt particle sizes and form factors. The influence of size effects on the form of the MRE spectra is confirmed. The reliability of the methods is confirmed by a comparison of the obtained results with empirical data. The value of the obtained results consists in the good agreement of all the calculated parameters of the discussed nanocomposite and the form of the spectral dependencies of the MRE with the results of various experiments.

Conclusions. The confirmation that both the size and form factor of granules have a significant impact on the appearance of the MRE spectra raises the prospect of developing promising nanocomposite properties at particular particle sizes. The presented results highlight the possibility of optimizing the material characteristics to improve sensitivity in magnetic sensors and noncontact devices for studying nanostructures.

Keywords: nanocomposites, effective medium theory, magnetorefractive effect, ferromagnetic, size effects

For citation: Yurasov A.N., Kulgunina R., Yashin M.M., Simdyanova M.A. Features of the magnetorefractive effect in Co–Si nanocomposites. *Russian Technological Journal*. 2025;13(3):122–128. <https://doi.org/10.32362/2500-316X-2025-13-3-122-128>, <https://www.elibrary.ru/IJRLMF>

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

The authors declare no conflicts of interest.

ВВЕДЕНИЕ

Изучение магниторефрактивного эффекта (МРЭ) в нанокompозитах «кобальт–кремний» (Co–Si) представляет собой важную задачу в контексте разработки современных магнитных и оптических устройств. Понимание вклада размерных эффектов в данное явление является ключевым для раскрытия магнитных и оптических свойств наноматериалов, что расширяет возможности для использования бесконтактных методов исследования [1–3].

С уменьшением размеров зерен в нанокompозитах существенно изменяются их физико-химические свойства, что может привести к значительному усилению МРЭ. Нанокompозиты на основе кобальта и кремния представляют особый интерес благодаря своей способности демонстрировать уникальные магнитооптические характеристики. Моделирование МРЭ с учетом размерного эффекта позволяет глубже понять механизмы, управляющие взаимодействием света и магнитного поля в таких системах [4, 5].

В связи с этим изучение свойств перспективных наноструктур сегодня является актуальной задачей с учетом возможного значительного усиления важных с практической точки зрения эффектов, таких как магнитосопротивление, квантовые эффекты Холла, МРЭ и многие другие [6]. Интересным примером наноструктуры служит нанокompозит Co–Si, а моделирование наблюдаемых оптических и магнитооптических эффектов позволяет бесконтактным способом оценивать различные характерные параметры исследуемых образцов [7, 8].

МАТЕМАТИЧЕСКАЯ МОДЕЛЬ И МЕТОДИКА РАСЧЕТА

В рамках данного исследования для анализа МРЭ в композитных материалах, содержащих наночастицы кобальта в кремниевой матрице, была разработана математическая модель, основанная на теории эффективной среды. Основная цель расчетов заключалась в исследовании влияния размерного эффекта и форм-фактора частиц на спектры МРЭ.

Магниторефрактивный эффект описывает влияние магнитного поля на комплексный показатель преломления нанокompозита, что выражается через изменение диэлектрической проницаемости ε под

воздействием магнитного поля [9]. В данном исследовании было выбрано приближение эффективной среды с использованием модели Бруггемана (ЕМА¹) для расчета эффективной диэлектрической проницаемости ε^{EMA} , на примере нанокompозита Co–Si. Объемная доля металлических частиц (Co) в данной структуре составляет $X = 0.5$.

Расчеты проводились для частиц диаметром от 2 до 8 нм с различными значениями форм-фактора L , что позволило изучить, как варьирование размеров и формы частиц влияет на спектры МРЭ.

Магниторефрактивный эффект рассчитывался как изменение коэффициента отражения R нанокompозита [10]:

$$\frac{\Delta R}{R} = -(1 - R) \left(\frac{\Delta \rho}{\rho} k^2 \left[\frac{3n^2 - k^2 - 1}{(n^2 + k^2)((1 - n)^2 + k^2)} \right] \right), \quad (1)$$

где $\frac{\Delta \rho}{\rho}$ – магнитосопротивление; k, n – коэффициенты экстинкции и рефракции соответственно.

Ключевыми параметрами модели являются диагональные и недиагональные комплексные компоненты тензора диэлектрической проницаемости (ТДП):

$$\gamma = \gamma_1 - i\gamma_2, \quad \varepsilon = \varepsilon_{01} - i\varepsilon_{02}, \quad (2)$$

где ε_{01} и γ_1 – действительные части диагональных и недиагональных компонент ТДП, а ε_{02} и γ_2 – мнимые части компонент ТДП, соответственно.

Эти параметры зависят от квазиклассических размерных эффектов, которые учитываются в данной работе как вклад формы и размера частиц, что отражается на спектральной зависимости МРЭ.

Размерный эффект учитывается через добавочные члены в диагональных и недиагональных компонентах ТДП на основе модели Друде – Лоренца. Диэлектрическая проницаемость и коэффициент поглощения частиц рассчитываются с учетом времени свободного пробега τ и концентрации частиц. Оптимальной при описании спектральных зависимостей наноструктур и нанокompозитов, в частности, является теория эффективной среды [11]. Эффективная среда описывается с помощью

¹ Effective medium approximations – приближение эффективной среды.

формулы Бруггемана, где учитывается вклад магнитной компоненты материала, объемной концентрации кобальта и формы наночастиц:

$$X \frac{\varepsilon_1 - \varepsilon^{\text{EMA}}}{\varepsilon^{\text{EMA}} + L(\varepsilon_1 - \varepsilon^{\text{EMA}})} + (1 - X) \frac{\varepsilon_0 - \varepsilon^{\text{EMA}}}{\varepsilon^{\text{EMA}} + L(\varepsilon_0 - \varepsilon^{\text{EMA}})} = 0, \quad (3)$$

где ε_0 и ε_1 – диэлектрические проницаемости компоненты среды, L – форм-фактор частиц среды.

Размерные эффекты учитываются варьированием форм-факторов частиц L и добавками в диагональные и недиагональные компоненты ТДП ферромагнитной компоненты нанокompозита, что связано с рассеянием электронов на поверхностях гранул. Окончательно с учетом вклада в ТДП размерных эффектов комплексные компоненты ТДП ε_{mod} и γ_{mod} согласно модели Друде – Лоренца представляются в виде [11, 12]:

$$\varepsilon_{\text{mod}} = \varepsilon_{\text{Co}} + \frac{\omega_p^2}{\omega(\omega + i/\tau_{\text{bulk}})} - \frac{\omega_p^2}{\omega(\omega + i/\tau_{\text{gr}})}, \quad (4)$$

$$\gamma_{\text{mod}} = \gamma_{\text{Co}} + \frac{4\pi\sigma_{xy}^{\text{bulk}}/\tau_{\text{bulk}}^2}{\omega(\omega + i/\tau_{\text{bulk}})^2} - \frac{4\pi\sigma_{xy}^{\text{gr}}/\tau_{\text{gr}}^2}{\omega(\omega + i/\tau_{\text{gr}})^2},$$

где ε_{Co} и γ_{Co} – диагональные и недиагональные компоненты ТДП ферромагнетика, в данном случае кобальта; ω_p – плазменная частота; ω – частота электромагнитной волны; $\sigma_{xy}^{\text{bulk}} = 4\pi M_s R_{\text{bulk}} / \rho_{\text{bulk}}^2$; $\sigma_{xy}^{\text{gr}} = 4\pi M_s R_{\text{gr}} / \rho_{\text{gr}}^2$; M_s – намагниченность насыщения ферромагнетика; R_{bulk} и R_{gr} – коэффициент аномального эффекта Холла массивного образца и гранулы соответственно; ρ_{bulk} и ρ_{gr} – удельное сопротивление массивного образца и гранулы соответственно; τ_{bulk} , τ_{gr} – время свободного пробега электронов в массивном образце и грануле соответственно.

Размерный эффект проявляется как в параметре аномального эффекта Холла, так и в удельном сопротивлении:

$$R_{\text{gr}} = R_{\text{bulk}} + 0.2R \frac{l}{r_0} \left(1 + \frac{l}{r_0} \right), \quad (5)$$

$$\rho_{\text{gr}} = \rho_{\text{bulk}} \left(1 + \frac{l}{r_0} \right), \quad (6)$$

где R – значение параметра аномального эффекта Холла материала поверхности гранул, r_0 – размер частиц нанокompозита и l – длина свободного пробега электрона.

РЕЗУЛЬТАТЫ МОДЕЛИРОВАНИЯ

В рамках перспективного метода эффективной среды – приближения Бруггемана по формулам (1)–(3) получены значения параметра МРЭ ($\Delta R/R$) без учета размерного эффекта, а затем проведен анализ влияния квазиклассического размерного эффекта на спектры при различной форме частиц L (форм-фактор) и диаметре частиц кобальта d . В качестве образца был выбран нанокompозит с объемной долей кобальта $X = 0.5$. Выбор определялся близостью к порогу перколяции, что может существенно менять и усиливать физические эффекты.

Как видно из рис. 1, наибольшее изменение МРЭ наблюдаются в ближней инфракрасной (ИК) области спектра, что связано с внутризонными переходами, при учете размерного эффекта и диаметре частиц $d = 2$ нм. На рис. 2 видно, что только начиная с размера гранул 4 нм вклад размерного эффекта становится заметным.

На рис. 3 произведен учет квазиклассического размерного эффекта для различных форм-факторов частиц. Наибольшее усиление эффекта соответствует $L = 0.2$. Полученные результаты по порядку величины хорошо согласуются с известными экспериментальными данными для нанокompозитов (например, [3, 4]) и дают общую тенденцию: при уменьшении размера гранул и форм-фактора можно наблюдать усиление МРЭ.

Полученные результаты открывают широкую область возможностей перспективного применения нанокompозитов в современной электронике [13–15], которые основаны на различных оптических, магнитооптических, гальваномагнитных и других эффектах, наблюдаемых в обсуждаемых наноструктурах.

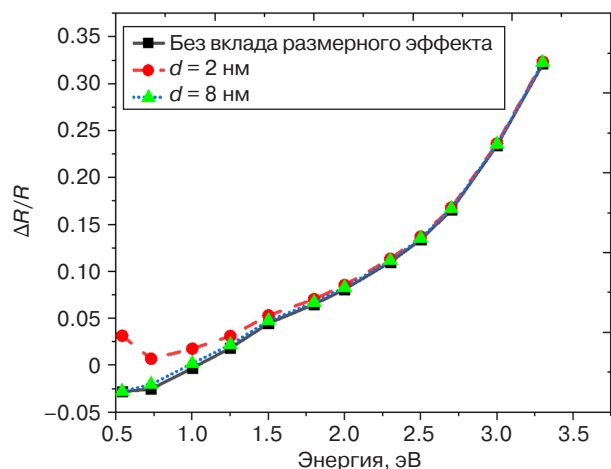


Рис. 1. Зависимость МРЭ от энергии падающей электромагнитной волны без учета вклада размерного эффекта (сплошная) и с вкладом размерного эффекта с размерами частиц Co $d = 2$ нм (точки) и $d = 8$ нм (пунктир)

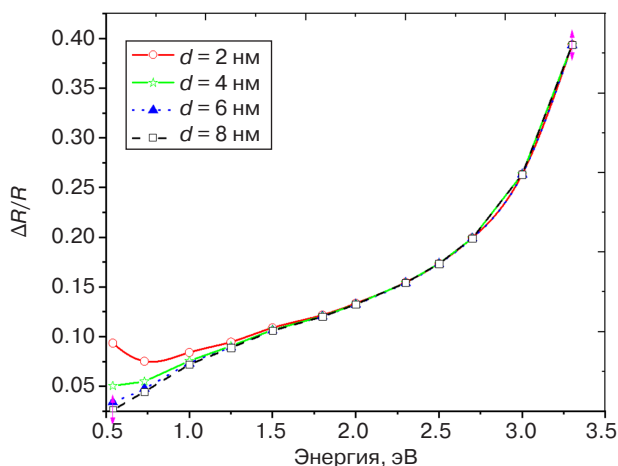


Рис. 2. Зависимость МРЭ с учетом вклада размерного эффекта от энергии падающей электромагнитной волны для нанокompозита Co–Si при разном диаметре частиц $d = 2, 4, 6$ и 8 нм

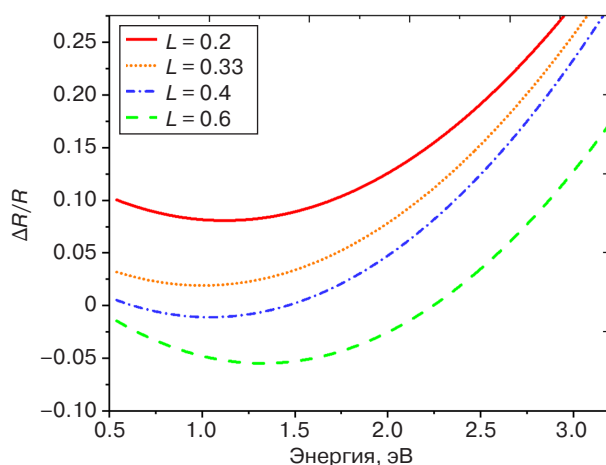


Рис. 3. Зависимость МРЭ с учетом вклада размерного эффекта от энергии падающей электромагнитной волны для нанокompозита Co–Si при $L = 0.2, 0.33, 0.4$ и 0.6

ЗАКЛЮЧЕНИЕ

В результате проведенной работы получены модельные спектры МРЭ на примере нанокompозитов Co–Si с учетом вклада размерного эффекта.

Показана важность учета вклада форм-фактора частиц, а также квазиклассического размерного эффекта в спектральные зависимости МРЭ. Все полученные результаты согласуются по порядку величины с известными данными для похожих наноструктур.

Важно отметить, что полученные результаты могут быть использованы для расширения возможностей бесконтактных методов исследования и разработки высокочувствительных датчиков и систем памяти на основе широкого спектра наноструктур, например, на нанокompозите Co–Si.

БЛАГОДАРНОСТИ

Работа выполнена при поддержке программы «Акселератор РТУ МИРЭА».

ACKNOWLEDGMENTS

The study was supported by the “RTU MIREA Accelerator” program.

Вклад авторов

А.Н. Юрасов – моделирование, обсуждение результатов, написание статьи, редактирование статьи.

Р. Кулгунина – обработка литературных источников, компьютерное моделирование, обсуждение результатов, написание статьи.

М.М. Яшин – разработка модели, компьютерное моделирование, обсуждение результатов, написание статьи.

М.А. Симдянова – обработка литературных источников, компьютерное моделирование, обсуждение результатов, написание статьи.

Authors' contributions

A.N. Yurasov – computer simulation, discussion of results, writing and editing the text of the article.

R. Kulgunina – literature review, computer simulation, discussion of results, writing the text of the article.

M.M. Yashin – model development, computer simulation, discussion of results, writing the text of the article.

M.A. Simdyanova – literature review, computer simulation, discussion of results, writing the text of the article.

СПИСОК ЛИТЕРАТУРЫ

- Shkurdoda Yu.O., Dekhtyaruk L.V., Basov A.G., Chornous A.M., Shabelnyk Yu.M., Kharchenko A.P., Shabelnyk T.M. The giant magnetoresistance effect in Co/Cu/Co three-layer films. *J. Magn. Magn. Mater.* 2019;477:88–91. <https://doi.org/10.1016/j.jmmm.2019.01.040>
- Jacquet J.C., Valet T. A New Magneto-optical Effect Discovered on Magnetic Multilayers: The Magnetorefractive Effect. In: *Magnetic Ultrathin Films, Multilayers and Surfaces: MRS Symposium Proc.* 1995. V. 384. P. 477–490. <https://doi.org/10.1557/PROC-384-477>
- Granovsky A., Sukhorukov Yu., Gan'shina E., Telegin A. Magnetorefractive effect in magnetoresistive materials. In: *Magnetophotonics: From Theory to Applications*. Berlin Heidelberg: Springer; 2013. P. 107–133. http://doi.org/10.1007/978-3-642-35509-7_5

4. Kelley C.S., Naughton J., Benson E., Bradley R.C., Lazarov V.K., Thompson S.M., Matthew J.A. Investigating the magnetic field-dependent conductivity in magnetite thin films by modelling the magnetorefractive effect. *J. Phys.: Condens. Matter*. 2014;26(3):036002. <http://doi.org/10.1088/0953-8984/26/3/036002>
5. Кринчик Г.С., Артемьев В.А. Магнитооптические свойства Ni, Co, и Fe в ультрафиолетовой, видимой и инфракрасной областях спектра. *Журнал экспериментальной и теоретической физики*. 1967;53(6):1901–1912.
6. Юрасов А.Н., Сайфулина Д.А., Бахвалова Т.Н. Магниторефрактивный эффект в металлических наноструктурах Co/Pt. *Russian Technological Journal*. 2024;12(2):57–66. <https://doi.org/10.32362/2500-316X-2024-12-2-57-66>
7. Ganshina E., Granovsky A., Gushin V., Kuzmichov M., Podrugin P., Kravetz A., Shipil E. Optical and magneto-optical spectra of magnetic granular alloys. *Physica A: Statistical Mechanics Application*. 1997;241(1–2):45–51. [https://doi.org/10.1016/S0378-4371\(97\)00057-5](https://doi.org/10.1016/S0378-4371(97)00057-5)
8. Domashevskaya E.P., Ivkov S.A., Sitnikov A.V., et al. Influence of the relative content of the metal component in the dielectric matrix on the formation and size of cobalt nanocrystals in $\text{Co}_x(\text{MgF}_2)_{100-x}$ film composites. *Phys. Solid State*. 2019;61(2):71–79. <https://doi.org/10.1134/S1063783419020112>
9. Reig C., Cardoso de Freitas S., Mukhopadhyay S.C. *Giant Magnetoresistance (GMR) Sensors. From Basis to State-of-the-Art Applications. In Series: Smart Sensors, Measurement and Instrumentation*. Berlin, Heidelberg: Springer; 2013. V. 6. 301 p. ISBN 978-3-642-37172-1
10. Юрасов А.Н. Магниторефрактивный эффект как бесконтактный метод исследования функциональных материалов. *Материаловедение*. 2014;6:32–37.
11. Юрасов А.Н., Яшин М.М. Методы эффективной среды как оптимальные методы моделирования физических свойств наноструктур. *Russian Technological Journal*. 2020;8(5):68–77. <https://doi.org/10.32362/2500-316X-2020-8-5-68-77>
12. Johnson P.B., Christy R.W. Optical constants of transition metals: Ti, V, Cr, Mn, Fe, Co, Ni, and Pd. *Phys. Rev. B*. 1974;9:5056–5070. <https://doi.org/10.1103/PhysRevB.9.5056>
13. Головань Л.А., Тимошенко В.Ю., Кашкаров П.К. Оптические свойства нанокompозитов на основе пористых систем. *УФН*. 2007;177(6):619–638. <https://doi.org/10.3367/UFNr.0177.200706b.0619>
14. Вахрушев А.В., Федотов А.Ю., Северюхина О.Ю., Сидоренко А.С. Исследование влияния структуры кобальта на магнитные свойства нанопленок. *Chemical Physics and Mesoscopy*. 2022;24(4):436–453. <https://doi.org/10.15350/17270529.2022.4.36>
15. Demirel F.E., Lavrijsen R., Koopmans B. An investigation of the interface and bulk contributions to the magneto-optic activity in Co/Pt multi-layered thin films. *J. Appl. Phys.* 2021;129(16):163904. <https://doi.org/10.1063/5.0047093>

REFERENCES

1. Shkurdoda Yu.O., Dekhtyaruk L.V., Basov A.G., Chornous A.M., Shabelnyk Yu.M., Kharchenko A.P., Shabelnyk T.M. The giant magnetoresistance effect in Co/Cu/Co three-layer films. *J. Magn. Magn. Mater.* 2019;477:88–91. <https://doi.org/10.1016/j.jmmm.2019.01.040>
2. Jacquet J.C., Valet T. A New Magneto-optical Effect Discovered on Magnetic Multilayers: The Magnetorefractive Effect. In: *Magnetic Ultrathin Films, Multilayers and Surfaces: MRS Symposium Proc.* 1995. V. 384. P. 477–490. <https://doi.org/10.1557/PROC-384-477>
3. Granovsky A., Sukhorukov Yu., Gan'shina E., Telegin A. Magnetorefractive effect in magnetoresistive materials. In: *Magnetophotonics: From Theory to Applications*. Berlin Heidelberg: Springer; 2013. P. 107–133. http://doi.org/10.1007/978-3-642-35509-7_5
4. Kelley C.S., Naughton J., Benson E., Bradley R.C., Lazarov V.K., Thompson S.M., Matthew J.A. Investigating the magnetic field-dependent conductivity in magnetite thin films by modelling the magnetorefractive effect. *J. Phys.: Condens. Matter*. 2014;26(3):036002. <http://doi.org/10.1088/0953-8984/26/3/036002>
5. Krinchik G.S., Artem'ev V.A. Magneto-optical properties of Ni, Co and Fe in the ultraviolet visible and infrared parts of the spectrum. *Journal of Experimental and Theoretical Physics (JETP)*. 1968;26(6):1080–1085. [Original Russian Text: Krinchik G.S., Artem'ev V.A. Magneto-optical properties of Ni, Co and Fe in the ultraviolet visible and infrared parts of the spectrum. *Zhurnal Eksperimental'noi i Teoreticheskoi Fiziki*. 1967;53(6):1901–1912 (in Russ.).]
6. Yurasov A.N., Sayfulina D.A., Bakhvalova T.N. Magnetorefractive effect in metallic Co/Pt nanostructures. *Russian Technological Journal*. 2024;12(2):57–66. <https://doi.org/10.32362/2500-316X-2024-12-2-57-66>
7. Ganshina E., Granovsky A., Gushin V., Kuzmichov M., Podrugin P., Kravetz A., Shipil E. Optical and magneto-optical spectra of magnetic granular alloys. *Physica A: Statistical Mechanics Application*. 1997;241(1–2):45–51. [https://doi.org/10.1016/S0378-4371\(97\)00057-5](https://doi.org/10.1016/S0378-4371(97)00057-5)
8. Domashevskaya E.P., Ivkov S.A., Sitnikov A.V., et al. Influence of the relative content of the metal component in the dielectric matrix on the formation and size of cobalt nanocrystals in $\text{Co}_x(\text{MgF}_2)_{100-x}$ film composites. *Phys. Solid State*. 2019;61(2):71–79. <https://doi.org/10.1134/S1063783419020112>
9. Reig C., Cardoso de Freitas S., Mukhopadhyay S.C. *Giant Magnetoresistance (GMR) Sensors. From Basis to State-of-the-Art Applications. In Series: Smart Sensors, Measurement and Instrumentation*. Berlin, Heidelberg: Springer; 2013. V. 6. 301 p. ISBN 978-3-642-37172-1
10. Yurasov A.N. Magnetorefractive effect as a contactless method for investigation of functional materials. *Materialovedenie = Material Science*. 2014;6:32–37 (in Russ.).

11. Yurasov A.N., Yashin M.M. Methods of effective media as optimal methods for modeling the physical properties of nanostructures. *Russian Technological Journal*. 2020;8(5):68–77 (in Russ.). <https://doi.org/10.32362/2500-316X-2020-8-5-68-77>
12. Johnson P.B., Christy R.W. Optical constants of transition metals: Ti, V, Cr, Mn, Fe, Co, Ni, and Pd. *Phys. Rev. B*. 1974;9: 5056–5070. <https://doi.org/10.1103/PhysRevB.9.5056>
13. Golovan L.A., Timoshenko V.Yu., Kashkarov P.K. Optical properties of porous-system-based nanocomposites. *Phys. Usp.* 2007;50(6):595–612. <https://doi.org/10.1070/PU2007v050n06ABEH006257>
[Original Russian Text: Golovan L.A., Timoshenko V.Yu., Kashkarov P.K. Optical properties of porous-system-based nanocomposites. *Uspekhi fizicheskikh nauk*. 2007;177(6):619–638 (in Russ.). <https://doi.org/10.3367/UFNr.0177.200706b.0619>]
14. Vakhrushev A.V., Fedotov A.Yu., Severyukhina O.Yu., Sidorenko A.S. Study of the influence of the cobalt structure on the magnetic properties of nanofilms. *Chemical Physics and Mesoscopy*. 2022;24(4):436–453 (in Russ.). <https://doi.org/10.15350/17270529.2022.4.36>
15. Demirer F.E., Lavrijsen R., Koopmans B. An investigation of the interface and bulk contributions to the magneto-optic activity in Co/Pt multi-layered thin films. *J. Appl. Phys.* 2021;129(16):163904. <https://doi.org/10.1063/5.0047093>

Об авторах

Юрасов Алексей Николаевич, д.ф.-м.н., профессор, кафедра наноэлектроники, Институт перспективных технологий и промышленного программирования, ФГБОУ ВО «МИРЭА – Российский технологический университет» (119454, Россия, Москва, пр-т Вернадского, д. 78). E-mail: alexey_yurasov@mail.ru. ResearcherID M-3113-2016, Scopus Author ID 6602974416, SPIN-код РИНЦ 4259-8885, <https://orcid.org/0000-0002-9104-3529>

Кулгунина Регина, бакалавр, Институт перспективных технологий и промышленного программирования, ФГБОУ ВО «МИРЭА – Российский технологический университет» (119454, Россия, Москва, пр-т Вернадского, д. 78). E-mail: yummy2002@mail.ru. <https://orcid.org/0009-0004-1864-9155>

Яшин Максим Михайлович, к.ф.-м.н., доцент, кафедра наноэлектроники, Институт перспективных технологий и промышленного программирования, ФГБОУ ВО «МИРЭА – Российский технологический университет» (119454, Россия, Москва, пр-т Вернадского, д. 78); доцент, кафедра физики, ФГАОУ ВО «Московский государственный технический университет имени Н.Э. Баумана (национальный исследовательский университет)» (МГТУ им. Н.Э. Баумана) (105005, Россия, Москва, ул. 2-я Бауманская, д. 5). E-mail: ihkamax@mail.ru. ResearcherID G-6809-2017, Scopus Author ID 57210607470, SPIN-код РИНЦ 2438-6135, <https://orcid.org/0000-0001-8022-9355>

Симдянова Марина Александровна, ассистент, кафедра наноэлектроники, Институт перспективных технологий и промышленного программирования, ФГБОУ ВО «МИРЭА – Российский технологический университет» (119454, Россия, Москва, пр-т Вернадского, д. 78). E-mail: marina.simdyanova3103@mail.ru. Scopus Author ID 58532241200, SPIN-код РИНЦ 1447-1140, <https://orcid.org/0009-0009-8418-6896>

About the Authors

Alexey N. Yurasov, Dr. Sci. (Phys.-Math.), Professor, Department of Nanoelectronics, Institute for Advanced Technologies and Industrial Programming, MIREA – Russian Technological University (78, Vernadskogo pr., Moscow, 119454 Russia). E-mail: alexey_yurasov@mail.ru. ResearcherID M-3113-2016, Scopus Author ID 6602974416, RSCI SPIN-code 4259-8885, <https://orcid.org/0000-0002-9104-3529>

Regina Kulgunina, Student, Institute for Advanced Technologies and Industrial Programming, MIREA – Russian Technological University (78, Vernadskogo pr., Moscow, 119454 Russia). E-mail: yummy2002@mail.ru. <https://orcid.org/0009-0004-1864-9155>

Maxim M. Yashin, Cand. Sci. (Phys.–Math.), Associate Professor, Department of Nanoelectronics, Institute for Advanced Technologies and Industrial Programming, MIREA – Russian Technological University (78, Vernadskogo pr., Moscow, 119454 Russia); Associate Professor, Department of Physics, Bauman Moscow State Technical University (5, 2-ya Baumanskaya ul., Moscow, 105005 Russia). E-mail: ihkamax@mail.ru. ResearcherID G-6809-2017, Scopus Author ID 57210607470, RSCI SPIN-code 2438-6135, <https://orcid.org/0000-0001-8022-9355>

Marina A. Simdyanova, Assistant, Department of Nanoelectronics, Institute for Advanced Technologies and Industrial Programming, MIREA – Russian Technological University (78, Vernadskogo pr., Moscow, 119454 Russia). E-mail: marina.simdyanova3103@mail.ru. Scopus Author ID 58532241200, RSCI SPIN-code 1447-1140, <https://orcid.org/0009-0009-8418-6896>

Отпечатано в «МИРЭА – Российский технологический университет».

119454, РФ, г. Москва, пр-т Вернадского, д. 78.

Подписано в печать 29.05.2025 г.

Формат 60 × 90/8. Печать цифровая.

Уч.-изд. листов 16.00.

Тираж 50 экз. Заказ № 2760.

Подписку на печатную версию
Russian Technological Journal можно оформить
через ООО «Агентство «Книга-Сервис», www.akc.ru.
Подписной индекс: 79641.

Printed in MIREA – Russian Technological University.
78, Vernadskogo pr., Moscow, 119454 Russian
Federation.

Signed to print May 29, 2025.

Format 60 × 90/8. Digital print.

C.p.l. 16.00.

50 copies. Order No. 2760.

Subscription to the *Russian Technological
Journal* printed version can be made through
the *Kniga-Servis* Agency, www.akc.ru.
Subscription index: 79641.

