

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

ИНФОРМАТИКА

INFORMATICS

TOM 20
VOL.

2 | 2023

ОТ РЕДАКЦИИ

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

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

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

Основные разделы журнала:

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

Префикс DOI: 10.37661

Условия распространения материалов:

контент доступен под лицензией Creative Commons Attribution 4.0 License

Индексирование:

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

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

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

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

Адрес редакции:

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

Editorial address:

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

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

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

THE EDITOR'S NOTE

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

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

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

The main sections of the journal:

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

DOI Prefix: 10.37661

Distribution:

content is distributed under Creative Commons Attribution 4.0 License

Indexation:

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

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

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

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

ОБЪЕДИНЕННЫЙ ИНСТИТУТ ПРОБЛЕМ ИНФОРМАТИКИ
НАЦИОНАЛЬНОЙ АКАДЕМИИ НАУК БЕЛАРУСИ

ИНФОРМАТИКА

Informatika

Том 20, № 2, апрель-июнь 2023

Ежеквартальный научный журнал

Издается с января 2004 г.

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

Г л а в н ы й р е д а к т о р

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

З а м е с т и т е л ь г л а в н о г о р е д а к т о р а

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

Р е д а к ц и о н н а я к о л л е г и я

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

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

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

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

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

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

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

Килин Сергей Яковлевич, д-р физ.-мат. наук, проф., академик НАН Беларуси, Президиум НАН
Беларуси (Минск, Беларусь)

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

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

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

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

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

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

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

Стемпковский Александр Леонидович, д-р техн. наук, проф., академик РАН, ИПИМ РАН (Москва, Россия)

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

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

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

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

Редакционный совет

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

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

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

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

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

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

Пеш Эрвин, Зигенский университет (Зиген, Германия)

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

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

Чиулла Карло, Университет Эпока (Тирана, Албания)

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

ИНФОРМАТИКА

Том 20, № 2, апрель-июнь 2023

Ответственный за выпуск *Мойсейчик Светлана Сергеевна*

Редактор *Гончаренко Галина Борисовна*

Корректор *Михайлова Анна Антоновна*

Компьютерная верстка *Бутевич Ольга Борисовна*

Сдано в набор 10.05.2023. Подписано в печать 15.06.2023. Формат 60×84 1/8. Бумага офсетная. Гарнитура Таймс. Ризография. Усл. печ. л. 13,9. Уч.-изд. л. 13,7. Тираж 40 экз. Заказ 2.

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

Свидетельство о государственной регистрации издателя, изготовителя, распространителя печатных изданий № 1/274 от 04.04.2014. ЛП № 02330/444 от 18.12.13. Ул. Сурганова, 6, 220012, Минск, Беларусь.

© Объединенный институт проблем информатики
Национальной академии наук Беларуси, 2023

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

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

INFORMATICS

Vol. 20, no. 2, April-June 2023

Published quarterly

Issued since January 2004

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

Editor-in-Chief

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

Deputy Editor-in-Chief

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

Editorial Board

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

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

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

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

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

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

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

Sergey Ya. Kilin, D. Sc. (Phys.-Math.), Prof., Academician of NASB, Presidium of NASB (Minsk, Belarus)

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

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

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

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

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

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

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

Alexander L. Stempkovsky, D. Sc. (Eng.), Prof., Academician of RAS, IPPM RAS (Moscow, Russia)

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

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

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

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

Editorial Council

Dmitry V. Efanov, Russian University of Transport (Moscow Institute of Transport Engineers) (Moscow, Russia)

Madhu Kumari, University Center for Research & Development, Chandigarh University (Mohali, Punjab, India)

Alexander A. Lazarev, V. A. Trapeznikov Institute of Control Sciences of the RAS (Moscow, Russia)

Tsung-Chyan Lai, Asia University at Taichung (The People's Republic of China, Taiwan)

Ninoslav Marina, St. Paul the Apostle University of Information Sciences and Technology (Ohrid, Macedonia)

Vazgen Sh. Melikyan, National Polytechnic University of Armenia (Yerevan, Armenia)

Erwin Pesch, University of Siegen (Siegen, Germany)

Tajinder Singh, Sant Longowal Institute of Engineering & Technology (Longowal, Punjab, India)

Maxim L. Khodachenko, Space Research Institute, Austrian Academy of Sciences (Graz, Austria)

Carlo Ciulla, Epoka University (Tirana, Albania)

Boris Steinberg, Institute of Mathematics, Mechanics and Computer Science Southern Federal University (Rostov-on-Don, Russia)

INFORMATICS

Vol. 20, no. 2, April-June 2023

Issue Head *Sviatlana S. Maiseichyk*

Editor *Halina B. Hancharenka*

Corrector *Hanna A. Mikhailava*

Computer Imposition *Volha B. Butsevich*

Sent for press 10.05.2023. Output 15.06.2023. Format 60×84 1/8. Offset paper. Headset Times. Riesography. Printed sheets 13,9. Publisher's signatures 13,7. Circulation 40 copies. Order 2.

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

Certificate on the state registration of the publisher, manufacturer, distributor of printing editions no. 1/274 dated 04.04.2014. License for the press no. 02330/444 dated 18.12.13.

6, Surganov Str., 220012, Minsk, Belarus.

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

СОДЕРЖАНИЕ

КОСМИЧЕСКИЕ ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ И ГЕОИНФОРМАТИКА

Белоконов И. В., Крот А. М., Козлов С. В., Каплярчук Е. А., Савиных И. Э., Шапкин А. С. Способ оценивания полного электронного содержания в ионосфере на основе ретрансляции сигналов глобальной навигационной спутниковой системы GPS.....7

ОБРАБОТКА СИГНАЛОВ, ИЗОБРАЖЕНИЙ, РЕЧИ, ТЕКСТА И РАСПОЗНАВАНИЕ ОБРАЗОВ

Фридман М. В., Косарева А. А., Снежко Э. В., Камлач П. В., Ковалев В. А. Методика формирования базы гистопатологических изображений папиллярного рака щитовидной железы для глубокого обучения28

ЛОГИЧЕСКОЕ ПРОЕКТИРОВАНИЕ

Бибило П. Н. Применение диаграмм решений не полностью определенных функций k -значной логики при синтезе логических схем.....39

Prihozhy A. A. Generation of shortest path search dataflow networks of actors for parallel multi-core implementation65

ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ

Липницкий С. Ф. Коррекция запросов в системе информационной поддержки принятия решений..... 85

Валевич С. В., Дик К. С., Пилецкий И. И., Крузе И., Асимов Р. М., Осипович В. С. Методика и программное средство поиска аномалий в данных телеметрии солнечной электростанции на основе анализа нормализованных значений мощности..... 96

МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ

Чуйко М. М., Королёва О. М. Решение смешанной краевой задачи для уравнения Пуассона в нерегулярных двумерных областях..... 111

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

CONTENTS

SPACE INFORMATION TECHNOLOGIES AND GEOINFORMATICS

Belokonov I. V., Krot A. M., Kozlov S. V., Kapliarchuk Y. A., Savinykh I. E., Shapkin A. S. A method for estimating the total electron content in the ionosphere based on the retransmission of signals from the global navigation satellite system GPS..... 7

SIGNAL, IMAGE, SPEECH, TEXT PROCESSING AND PATTERN RECOGNITION

Fridman M. V., Kosareva A. A., Snezhko E. V., Kamlach P. V., Kovalev V. A. Papillary thyroid carcinoma whole-slide images as a basis for deep learning..... 28

LOGICAL DESIGN

Bibilo P. N. Application of decision diagrams of incompletely specified of k -valued logic functions in the synthesis of logical circuits 39

Prihozhy A. A. Generation of shortest path search dataflow networks of actors for parallel multi-core implementation 65

INFORMATION TECHNOLOGIES

Lipnitsky S. F. Correction of requests in the information system decision support..... 85

Valevich S. V., Dzick K. S., Pilecki I. I., Kruse I., Asimov R. M., Asipovich V. S. Methods and software for anomalies searching in the telemetry data of a solar power plant based on the normalized power analysis 96

MATHEMATICAL MODELING

Chuiko M. M., Korolyova O. M. Solution of the mixed boundary problem for the Poisson equation on two-dimensional irregular domains 111

КОСМИЧЕСКИЕ ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ И ГЕОИНФОРМАТИКА

SPACE INFORMATION TECHNOLOGIES AND GEOINFORMATICS



УДК 550.388.2
<https://doi.org/10.37661/1816-0301-2023-20-2-7-27>

Оригинальная статья
Original Paper

Способ оценивания полного электронного содержания в ионосфере на основе ретрансляции сигналов глобальной навигационной спутниковой системы GPS

И. В. Белоконов¹, А. М. Крот², С. В. Козлов^{3✉}, Е. А. Каплярчук³, И. Э. Савиных²,
А. С. Шапкин²

¹Самарский национальный исследовательский университет
имени академика С. П. Королева,
ул. Московское шоссе, 34, Самара, 443086, Россия

²Объединенный институт проблем информатики
Национальной академии наук Беларуси,
ул. Сурганова, 6, Минск, 220012, Беларусь

³Белорусский государственный университет
информатики и радиоэлектроники,
ул. П. Бровки, 6, Минск, 220012, Беларусь
✉E-mail: kozlov@bsuir.by

Аннотация

Цели. Решается задача разработки эффективного по аппаратурным затратам способа оценивания полного электронного содержания в ионосфере на основе ретрансляции сигналов глобальной навигационной спутниковой системы GPS на двух частотах L_1 , L_2 когерентных сигналов с использованием малогабаритного наноспутника-ретранслятора.

Методы. Показано, что при ретрансляции навигационных сигналов на выделенные для геофизических исследований частоты 150/400 МГц образуется когерентная многопозиционная радиолокационная система, включающая навигационные спутники (НС) – источники сигналов, наноспутник-ретранслятор (СР) и наземные приемные пункты (ПП). Время задержки и фазы четырех принимаемых сигналов содержат информацию о суммарном полном электронном содержании на трассах распространения НС – СР и СР – ПП. За счет последующей обработки сигналов возможно выделение полного электронного содержания на каждой из трасс распространения, а также нахождение координат СР.

Результаты. Определены способ и порядок оценивания полного электронного содержания по результатам обработки ретранслированных сигналов и технические требования к аппаратуре ретрансляции. Представлены характеристики точности предлагаемого метода и результаты моделирования.

Заключение. Приведенные в статье сведения могут быть полезны для специалистов и исследователей, интересующихся вопросами радиотомографического изучения ионосферы и прогнозирования опасных природных явлений.

Ключевые слова: глобальная навигационная спутниковая система, спутник-ретранслятор, навигационный сигнал, многопозиционная радиолокационная система, радиотомография, полное электронное содержание

Благодарности. Работа выполнена частично при финансовой поддержке Белорусского республиканского фонда фундаментальных исследований (проект № Ф20Р-329 «Теоретические основы исследования волновых процессов и явлений в ионосфере с использованием сигналов спутниковых радионавигационных систем») и в рамках задания 1.10.3 (Т 103) Государственной программы научных исследований «Цифровые и космические технологии, безопасность человека, общества и государства», а также в рамках договора № 220/12 «Разработать алгоритмические и программные средства обработки радиотомографических данных низкоорбитального контроля ионосферы» (4.05.2022–31.12.2025 гг.), заключенного с УП «Геоинформационные системы» по проекту «Разработать космическую систему радиометрического контроля околоземного пространства на базе малого космического аппарата и специализированных наземных средств» (мероприятия подпрограммы 6 «Исследование и использование космического пространства в мирных целях» Государственной программы «Научные технологии и техника» на 2021–2025 гг.).

Для цитирования. Способ оценивания полного электронного содержания в ионосфере на основе ретрансляции сигналов глобальной навигационной спутниковой системы GPS / И. В. Белоконов [и др.] // Информатика. – 2023. – Т. 20, № 2. – С. 7–27. <https://doi.org/10.37661/1816-0301-2023-20-2-7-27>

Конфликт интересов. Авторы заявляют об отсутствии конфликта интересов.

Поступила в редакцию | Received 24.01.2023

Подписана в печать | Accepted 24.02.2023

Опубликована | Published 29.06.2023

A method for estimating the total electron content in the ionosphere based on the retransmission of signals from the global navigation satellite system GPS

Igor V. Belokonov¹, Alexander M. Krot², Sergey V. Kozlov^{3✉}, Yauheni A. Kapliarchuk³, Irina E. Savinykh², Aliaksandr S. Shapkin²

¹*Samara National Research University named after academician S. P. Korolev
st. Moscow highway, 34, Samara, 443086, Russia*

²*The United Institute of Informatics Problems
of the National Academy of Sciences of Belarus,
st. Surganova, 6, Minsk, 220012, Belarus*

³*Belarusian State University of Informatics and Radioelectronics,
st. P. Brovki, 6, Minsk, 220012, Belarus*

✉E-mail: kozlov@bsuir.by

Abstract

Objectives. The problem of developing hardware effective method for estimating the total electron content in the ionosphere based on retransmission of the L_1 , L_2 signals of the global navigation satellite system GPS using a repeater nanosatellite is solved.

Methods. It is shown that with the retransmission of L_1 , L_2 signals at frequencies of 150/400 MHz allocated for geophysical research, a coherent multi-position radar system is formed, including navigation satellites (NS) – signal sources, repeater nanosatellite (SR) and ground receiving points (RP). The delay time and phase of the four received signals contain the information about the total TEC on the propagation paths NS – SR and SR – RP. It is shown that due to retransmission and subsequent processing, it is possible to isolate TECs on each of the propagation paths as well as determination of the coordinates of the SR.

Results. The content of the method, the procedure for evaluating TEC based on the results of processing the relayed signals, and the technical requirements for the relay equipment are determined. The accuracy characteristics of the proposed method are obtained. Simulation results are given.

Conclusion. The information presented in the article may be useful for specialists and researchers who interested in the issues of radio tomographic research of the ionosphere and forecasting hazardous natural phenomena.

Keywords: global navigation satellite system, satellite-retransmitter, navigation signal, multi-position radar system, radio tomography, total electronic content

Acknowledgements. This work has been partially financially supported by the Belarusian Republican Foundation for Fundamental Research (project no. F20R-329 "Theoretical foundations of the study of wave processes and phenomena in the ionosphere using signals from satellite radio navigation systems") and task 1.10.3 (T 103) of the State Program of Scientific Research "Digital and space technologies, security of man, society and the state", as well as within the framework of agreement no. 220/12 "Development of algorithmic and software tools for processing radio tomographic data of low-orbit ionosphere monitoring" (05.05.2022–12.31.2025) with UE "Geoinformation Systems" on the project "Develop a space system for radiometric monitoring of near-Earth space based on a small spacecraft and specialized ground facilities" (activities of subprogram 6 "Research and use of outer space for peaceful purposes" of the State Program "Science-intensive technologies and engineering" for 2021–2025).

For citation. Belokonov I. V., Krot A. M., Kozlov S. V., Kapliarchuk Y. A., Savinykh I. E., Shapkin A. S. *A method for estimating the total electron content in the ionosphere based on the retransmission of signals from the global navigation satellite system GPS*. *Informatika [Informatics]*, 2023, vol. 20, no. 2, pp. 7–27 (In Russ.). <https://doi.org/10.37661/1816-0301-2023-20-2-7-27>

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

Введение. Ионосфера существенно воздействует на распространение радиоволн, в частности сигналов глобальной навигационной спутниковой системы (ГНСС)¹ [1, 2]. Основное влияние на распространение радиоволн оказывает электронная концентрация в ионосфере. За счет электронной концентрации ионосфера становится для радиоволн диспергирующей средой, в которой скорость распространения зависит от длины волны. По этой причине происходит задержка распространения и преломление траекторий радиоволн в ионосфере, а для радиоволн большой длины – и полное отражение в ионосфере. Состояние ионосферы описывается множеством параметров: электронной концентрацией, ионным составом, ионной и электронной температурами, скоростью движения частиц и др. Среди них наиболее важным параметром является электронная концентрация. В зависимости от высоты и распределения электронной концентрации ионосферу условно разделяют на слои D (ниже 90 км), E (от 90 до 130 км), F₁ (от 130 до 200 км) и F₂ (выше 200 км). Максимальная концентрация электронов и ионов расположена в слое F₂ на высотах от 250 до 450 км [1].

На данный момент имеется большой арсенал методов, средств и систем измерения электронной концентрации в ионосфере² [1, 3–7]. К их числу следует отнести станции вертикального и наклонного зондирования ионосферы, радиозонды, средства радиозондирования ионосферы с наноспутников [4], со специализированных искусственных спутников Земли научного назначения и с использованием сигналов ГНСС. Основным методом определения электронной концентрации в настоящее время является *радиотомографический* [1]. Он предполагает измерение значений полного электронного содержания (ПЭС) как интеграла (его называют Хафовским или Радоновским образом) от электронной концентрации на заданной трассе между передатчиком и приемником с последующим использованием результатов измерений для восстановления зависимости электронной концентрации от географических координат томографическими методами. В этой связи определение ПЭС представляет собой важную задачу исследования ионосферы.

В последние годы наиболее широко применяется высокоорбитальный метод трансionoсферного зондирования на основе приема сигналов ГНСС типа ГЛОНАСС или GPS посредством двухчастотной навигационной аппаратуры потребителя, расположенной на Земле. Его суть состоит в следующем [1, 3]. При прохождении через ионосферу радиосигналы, излучаемые на несущих частотах $f_1 \approx 1,6$ и $f_2 \approx 1,2$ ГГц с навигационного спутника, испытывают различные временные задержки и, следовательно, изменения фаз. Измерение этих параметров в расположенном на Земле двухчастотном приемнике ГНСС позволяет непрерывно определять значение ПЭС в ионосфере вдоль трассы НС – ПП (от любого видимого НС) в любой момент

¹Радиотомографический комплекс для исследований ионосферы и ближнего космоса. Интеллектуальная система тематического исследования наукометрических данных [Электронный ресурс]. – Режим доступа: <https://istina.msu.ru/equipment/card/10360783/?ysclid=175pyou9a9358457142>. – Дата доступа: 18.10.2022.

²Первалова, Н. П. Исследование ионосферных возмущений методом трансionoсферного GPS-зондирования : дис. ... докт. физ.-мат. наук : 25.00.29 / Н. П. Первалова. – Иркутск, 2014. – 286 л.

времени t . В качестве аппаратуры может применяться любой двухчастотный приемник сигналов ГЛОНАСС/GPS. Недостатком метода является его достаточная «статичность», так как положение линии ПП – НС изменяется относительно медленно. Кроме того, высокоорбитальный метод не позволяет обнаруживать локальные неоднородности ионосферы.

Дополнительные данные для решения задач по реконструкции локальных значений ПЭС в ионосфере могут быть получены с использованием низкоорбитального метода на основе орбитальной группировки малых спутников в радиодиапазоне³ [7–9].

Первый вариант низкоорбитальной томографии реализуется на спутниках типа «Космос», «Транзит», FORMOSAT-3/COSMIC [5], излучающих когерентные сигналы на выделенных для геофизических исследований частотах 150/400 МГц. На основе этих сигналов фазоразностным методом измеряется ПЭС на трассах «спутник – наземный ПП», что позволяет расширять данные для радиотомографии ионосферы. *Второй вариант* заключается в размещении на спутнике двухчастотного навигационного приемника, измеряющего дальности и фазы когерентных навигационных сигналов аналогично высокоорбитальной томографии. Кроме того, организуется радиоканал для обмена данными между НС и ПП с достаточно большой пропускной способностью.

Недостатками используемых методов низкоорбитальной томографии являются:

– необходимость реализации дополнительных каналов связи с наземными ПП с достаточно высокой пропускной способностью и энергопотреблением;

– высокие технические требования к двухчастотному приемнику, который должен обеспечить измерения информационных параметров принимаемого навигационного сигнала при больших радиальных скоростях и ускорениях между НС и малым спутником, и, следовательно, достаточно большая масса и энергопотребление приемника, а также высокие требования к радиационной стойкости используемых компонентов;

– ограниченный объем исходных данных по ПЭС и их неоднозначность для существующего способа с использованием когерентных сигналов 150/400 МГц. Согласно [6] для устранения неоднозначности требуется не менее четырех когерентных немодулированных сигналов с частотами 150, 400, 1067 и 2844 МГц, формируемых из выходного сигнала общего опорного генератора.

Цель работы заключается в обосновании способа измерения ПЭС для реализации радиотомографического метода зондирования ионосферы за счет ретрансляции навигационных сигналов ГНСС GPS наноспутником формата CubeSat.

Содержание предлагаемого способа. Рассмотрим иллюстрацию предлагаемого способа (рис. 1). Один или несколько СР выводятся на орбиты высотой 150...1000 км. Общее число СР определяется требуемым темпом обновления данных для реализации радиотомографии. При равномерном распределении СР по долготе восходящего узла в зоне видимости наземного ПП будет находиться, как правило, один СР. Использование одного СР на орбите обеспечивает получение данных до пяти-шести раз в сутки.

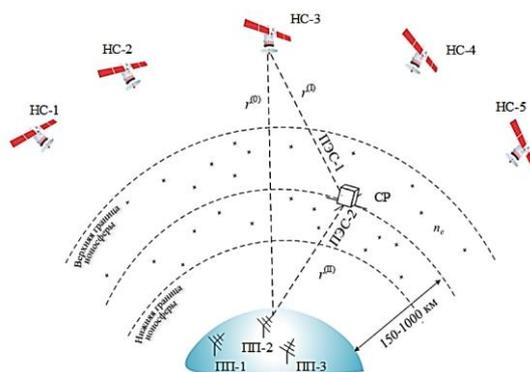


Рис. 1. Способ ретрансляции
Fig. 1. Relay method

³Ионозонд [Электронный ресурс] / Институт космических исследований Российской академии наук. – Режим доступа: <https://iki.cosmos.ru/missions/ionozond>. – Дата доступа: 18.10.2022.

Предлагаемое техническое решение заключается в приеме сигналов ГНСС спутником-ретранслятором GPS на двух рабочих частотах, фильтрации принятых сигналов, усилении, переносе спектров сигналов на частоты 150/400 МГц и излучении посредством слабонаправленной антенны на Землю (в нижнюю полусферу). Прием и обработка ретранслированных сигналов осуществляются одним или несколькими пространственно разнесенными ПП.

Так как навигационные сигналы являются когерентными с известными параметрами, включая моменты излучения и координаты НС, то предлагаемая совокупность технических средств образует многопозиционную радиолокационную систему с большим числом (до 8–10) независимо функционирующих источников когерентных зондирующих сигналов, единственной радиолокационной целью в виде СР и одним или несколькими ПП. Эту радиолокационную систему можно представить как совокупность бистатических [10, 11] радиолокационных средств, образованных каждым НС, СР и ПП, с подвижной ввиду взаимного пространственного перемещения НС, СР и ПП базой. Координаты НС и ПП известны с высокой точностью. Так, например, при использовании ультрабыстрых эфемерид⁴ точность определения координат НС составит примерно 5 см. Излучаемые навигационные сигналы содержат неизвестное навигационное сообщение. Оно может быть легко исключено из принимаемого сигнала за счет приема и дешифрации прямого навигационного сигнала на ПП с использованием стандартного навигационного приемника, так как прием прямого навигационного сигнала на ПП происходит раньше приема ретранслированного сигнала. В связи с высокой когерентностью навигационных сигналов каждого отдельного НС в этой системе могут быть реализованы методы синтеза апертуры [12] или способы обработки при длительном когерентном накоплении [13]. Аналогичный подход к использованию навигационных сигналов реализован в работе [14].

Отметим, что при ретрансляции из принимаемых на ПП сигналов можно извлечь ту же самую информацию, что и при приеме прямых навигационных сигналов в двухчастотном приемнике: время задержки и разность фаз сигналов на частотах f_1 и f_2 . Для сохранения когерентности перенос ретранслируемых навигационных сигналов необходимо проводить с использованием двух гетеродинов на базе синтезаторов частот, работающих от одного высокостабильного опорного генератора. Кроме того, по измерениям суммарных дальностей НС – СР – ПП от множества видимых НС возникает возможность определения координат и параметров движения СР.

В общем случае СР может быть выведен на высоты от 150 до 1000 км. Результаты анализа преимуществ и недостатков высот, на которые потенциально может быть выведен СР, приведены в таблице.

Сравнительная таблица высот, на которые выводится СР
Comparative table of the heights to which the repeater satellite is launched

Высота орбиты СР, км <i>Orbit height of SR, km</i>	Преимущества <i>Advantages</i>	Недостатки <i>Disadvantages</i>
150...200	Измерение ПЭС на трассах НС – СР при достаточно высокой угловой скорости линии НС – СР, возможность детального исследования ПЭС в слоях D, E, F ₁	Необходимость оснащения СР двигателем для обеспечения достаточного времени активного существования
300...600	Возможность детального исследования слоя F ₂ с увеличением объема данных для радиотомографии за счет измерения ПЭС на трассах НС – СР и СР – ПП, достаточное время активного существования СР без двигателя поддержания орбиты	Усложнение аппаратуры ретрансляции и обработки
> 600 (аналогично FORMOSAT-3/ COSMIC)	Возможность измерения ПЭС на трассах НС – СР выше слоя F ₂ при аномальных возмущениях в ионосфере. Высокая угловая скорость на линии СР – ПП при небольшой дальности, что позволяет выявлять локальные и кратковременно существующие неоднородности ионосферы. Максимальное время активного существования	Невозможность детального исследования слоя F ₂

⁴Эфемериды бортовые и точные // Orient Systems – Российский производитель ГНСС оборудования [Электронный ресурс]. – Режим доступа: <https://orsyst.ru/blog10>. – Дата доступа: 18.10.2022.

Основные преимущества предлагаемого способа определяются расширением объема данных ПЭС для радиотомографии ионосферы; существенным упрощением аппаратуры, расположенной на спутнике, за счет отказа от размещения на нем прецизионного двухчастотного многоканального навигационного приемника с высокой радиационной стойкостью и большим энергопотреблением; отказом от использования аппаратуры связи с достаточно высокой пропускной способностью; возможностью измерения ПЭС путем обработки принятой реализации одновременно по большому числу НС; возможностью оценивания координат СР и прогнозирования его положения на следующем витке, а также использования существующей приемной аппаратуры 150/400 МГц после ее модернизации в части реализации методов цифровой обработки принимаемых сигналов.

Обоснование способа и вывод основных соотношений. Осуществим ретрансляцию принимаемого аппаратурой СР навигационного сигнала системы GPS на частоте $f_1=1575,42$ МГц одновременно на частоты $f_{p1}=150$ и $f_{p2}=400$ МГц, выделенные в соответствии с регламентом радиосвязи для геофизических исследований на трассах «космос – Земля». Аналогичные операции выполним для навигационного сигнала GPS на частоте $f_2=1227,6$ МГц. Получим четыре ретранслированных сигнала, принимаемых на наземном ПП:

$$\begin{aligned} s_{f_1, f_{p1}}(t) : f_1 \rightarrow f_{p1}; \quad s_{f_1, f_{p2}}(t) : f_1 \rightarrow f_{p2}; \\ s_{f_2, f_{p1}}(t) : f_2 \rightarrow f_{p1}; \quad s_{f_2, f_{p2}}(t) : f_2 \rightarrow f_{p2}. \end{aligned} \quad (1)$$

Для ГНСС GPS на частоте f_1 используется сигнал C/A, а на частоте f_2 – сигнал L2C, оба эти сигнала характеризуются фазовой модуляцией и кодовым разделением. Сигнал L2C введен для НС, начиная с IIR-M, в настоящее время имеются 24 спутника из 32 в орбитальной группировке⁵. Дальномерные коды (законы модуляции) этих сигналов имеют одинаковую ширину спектра (1,023 МГц), но разную длительность (1 и 20 мс) и являются квазиортогональными по законам модуляции, что обеспечивает возможность их разделения при согласованной фильтрации на ПП.

Будем считать, что плоская радиоволна распространяется от СР, который находится на времязависимом расстоянии $L(t)$ до ПП, вдоль оси z . В этом случае напряженность электрического поля определяется формулой [15]

$$E(L, t) = E_0 \cos \left(2\pi f \left[t - \frac{1}{c} \int_0^{L(t)} \bar{n}(z) dz \right] \right), \quad (2)$$

где \bar{n} – показатель преломления, c – скорость света.

С учетом того что ионосферный слой расположен в интервале $[L_0(t), L(t)]$, а атмосферный и тропосферные слои – в интервале $[0, L_0(t)]$, оценим время запаздывания радиосигнала (2):

$$\tau = \frac{1}{c} \int_0^{L_0(t)} \bar{n}(z) dz + \frac{1}{c} \int_{L_0(t)}^{L(t)} \bar{n}(z) dz = \frac{1}{c} \int_0^{L_0(t)} \bar{n}_{tr}(z) dz + \frac{1}{c} \int_{L_0(t)}^{L(t)} \bar{n}_f(z) dz, \quad (3)$$

где \bar{n}_{tr} , \bar{n}_f – тропосферный и ионосферный показатели преломления. Если принять во внимание, что ионосферный и тропосферный показатели преломления равны соответственно [15, 16]

$$\bar{n}_f(z) \approx 1 - \frac{e^2 n_e(z)}{8\pi^2 \epsilon_0 m_e f^2}; \quad (4)$$

$$\bar{n}_{tr}(z) \approx 1 + \Delta \bar{n}_{tr}(z), \quad (5)$$

⁵Современные ГНСС. Прикладной потребительский центр ГЛОНАСС [Электронный ресурс]. – Режим доступа: <https://www.glonass-iac.ru>. – Дата доступа: 18.10.2022.

то фазовый путь радиоволны с частотой f (псевдодальность d_ϕ) определяется соотношением

$$d_f = L(t) + \int_0^{L_0(t)} \Delta \bar{n}_{tr}(z) dz - \frac{A}{f^2} \int_{L_0(t)}^{L(t)} n_e(z) dz, \quad (6)$$

где $A = \frac{e^2}{8\pi^2 \epsilon_0 m_e} \approx 40,308 \text{ (м}^3/\text{с}^2)$, e – заряд электрона, ϵ_0 – диэлектрическая проницаемость свободного пространства, m_e – масса электрона. Если фазы и соответствующие псевдодальности измеряются на двух разных частотах f_1 и f_2 , то согласно выражению (6) нетрудно найти ПЭС ионосферного слоя:

$$TEC = \int_{L_0(t)}^{L(t)} n_e(z) dz = \frac{1}{A} \cdot \frac{d_{f_2} - d_{f_1}}{\frac{1}{f_1^2} - \frac{1}{f_2^2}}. \quad (7)$$

В случае распространения не монохроматической плоской волны с напряженностью (2), а модулированного навигационного сигнала $s(t)$ его можно выразить через обратное преобразование Фурье [2]:

$$s(t) = \text{Re} \left(e^{j2\pi ft} \frac{1}{2\pi} \int_{-\Delta f/2}^{\Delta f/2} \dot{S}(\nu) e^{j2\pi \nu t} d\nu \right),$$

где $\dot{S}(\nu)$ – спектральная плотность комплексной огибающей навигационного сигнала, j – мнимая единица, Δf – ширина спектра сигнала, и представить навигационный сигнал в виде волнового пакета из плоских волн с длинами $\lambda \in \left[\frac{c}{f + \Delta f/2}, \frac{c}{f - \Delta f/2} \right]$. Различные частотные компоненты сигнала испытывают разные задержки, поэтому необходимо использовать групповую задержку, а не фазовую. Групповая задержка определяется аналогично (3):

$$\tau = \frac{1}{c} \int_0^{L_0(t)} \bar{n}^{gp}(z) dz + \frac{1}{c} \int_{L_0(t)}^{L(t)} \bar{n}^{gp}(z) dz = \frac{1}{c} \int_0^{L_0(t)} \bar{n}_{tr}^{gp}(z) dz + \frac{1}{c} \int_{L_0(t)}^{L(t)} \bar{n}_f^{gp}(z) dz, \quad (8)$$

где $\bar{n}_{tr}^{gp}, \bar{n}_f^{gp}$ – тропосферный и ионосферный групповые показатели преломления.

Тропосферу считаем недиспергирующей средой, поэтому для нее групповой показатель преломления находится по формуле

$$\bar{n}_{tr}^{gp}(z) = \bar{n}_{tr}(z) \approx 1 + \Delta \bar{n}_{tr}(z). \quad (9)$$

Для ионосферы групповой показатель преломления определяется соотношением [16]

$$\bar{n}_f^{gp}(z) \approx 1 + \frac{e^2 n_e(z)}{8\pi^2 \epsilon_0 m_e f^2}. \quad (10)$$

Тогда псевдодальность, измеренная по времени задержки огибающей модулированного навигационного сигнала, будет вычисляться по аналогии с выражением (6), но со знаком + для последнего слагаемого, отражающего влияние ионосферы:

$$d_f = L(t) + \int_0^{L_0(t)} \Delta \bar{n}_{tr}(z) dz + \frac{A}{f^2} \int_{L_0(t)}^{L(t)} n_e(z) dz = L(t) + \int_0^{L_0(t)} \Delta \bar{n}_{tr}(z) dz + \frac{A}{f^2} TEC. \quad (11)$$

Учитывая в соответствии с рис. 1 двукратное распространение сигнала на разных центральных частотах и при различных значениях ПЭС на трассах распространения, а также опускающая тропосферную ошибку дальности, одинаковую для всех комбинаций «частота навигационного сигнала/частота ретрансляции», запишем суммарные псевдодальности, измеряемые по задержкам огибающих (при использовании групповой скорости) принимаемых сигналов:

$$\begin{aligned}d_{f_1, f_{p1}} &= r^{(I)} + \frac{A \cdot TEC^{(I)}}{f_1^2} + r^{(II)} + \frac{A \cdot TEC^{(II)}}{f_{p1}^2}, \\d_{f_1, f_{p2}} &= r^{(I)} + \frac{A \cdot TEC^{(I)}}{f_1^2} + r^{(II)} + \frac{A \cdot TEC^{(II)}}{f_{p2}^2}, \\d_{f_2, f_{p1}} &= r^{(I)} + \frac{A \cdot TEC^{(I)}}{f_2^2} + r^{(II)} + \frac{A \cdot TEC^{(II)}}{f_{p1}^2}, \\d_{f_2, f_{p2}} &= r^{(I)} + \frac{A \cdot TEC^{(I)}}{f_2^2} + r^{(II)} + \frac{A \cdot TEC^{(II)}}{f_{p2}^2},\end{aligned}\tag{12}$$

где $r^{(I)}$, $r^{(II)}$ – геометрические длины трасс НС – СР и СР – ПП соответственно; $TEC^{(I)}$, $TEC^{(II)}$ – значения ПЭС на указанных трассах. Для сокращения изложения в (12) не приведены другие составляющие измерения суммарных псевдодальностей, такие как ошибка часов НС [2], задержка сигналов в ретрансляции и др.

Оценка ПЭС на трассе СР – ПП. Из выражений (7) и (12) следует, что, имея результаты измерения $\hat{d}_{f_1, f_{p1}}$, $\hat{d}_{f_1, f_{p2}}$, $\hat{d}_{f_2, f_{p1}}$, $\hat{d}_{f_2, f_{p2}}$ суммарных псевдодальностей для одного из видимых НС, нетрудно получить две независимые оценки ПЭС $TEC^{(II)}$ (обозначены нижним индексом) на трассе СР – ПП на частотах f_1 и f_2 :

$$TEC_{f_1}^{(II)} = \frac{\hat{d}_{f_1, f_{p1}} - \hat{d}_{f_1, f_{p2}}}{A \left(\frac{1}{f_{p1}^2} - \frac{1}{f_{p2}^2} \right)}, \quad TEC_{f_2}^{(II)} = \frac{\hat{d}_{f_2, f_{p1}} - \hat{d}_{f_2, f_{p2}}}{A \left(\frac{1}{f_{p1}^2} - \frac{1}{f_{p2}^2} \right)}.\tag{13}$$

При использовании перекрестной ретрансляции на частоты 150/400 МГц примечательными являются следующие обстоятельства:

1. За счет ретрансляции одного и того же навигационного сигнала с частоты f_1 или f_2 на две другие частоты 150/400 МГц и последующего вычитания в выражении (13) оценок суммарной дальности в результирующем выражении взаимно компенсируются как истинные (геометрически) дальности $r^{(I)}$, $r^{(II)}$ между элементами системы, так и ионосферная ошибка, обусловленная ПЭС на трассе НС – СР. Кроме того, исключаются ошибки часов НС и практически полностью тропосферная составляющая ошибки.

2. За счет ретрансляции на более низкие частоты резко увеличивается чувствительность метода определения ПЭС по дальностным измерениям, так как величина ПЭС $1 \text{ TECu} = 10^{16} \text{ 1/М}^2$ будет соответствовать разности суммарных псевдодальностей $d_{f_1, f_{p1}} - d_{f_1, f_{p2}} = d_{f_2, f_{p1}} - d_{f_2, f_{p2}} = 15,4 \text{ м}$, в то время как при использовании классического метода на основе приема навигационных сигналов на частотах f_1, f_2 при ПЭС 1 TECu разность дальностей будет соответствовать всего $0,1 \text{ м}$, что весьма сложно для измерения.

3. Так как измерения $TEC_{f_1,n}^{(II)}$, $TEC_{f_2,n}^{(II)}$ по сигналам $n = \overline{1, N_{GPS}}$ НС, которые одновременно наблюдаются на СР и ПП, являются независимыми, можно усреднить:

$$\overline{TEC}^{(II)} = \frac{1}{2} \left(\frac{1}{N_{GPS}} \sum_{n=1}^{N_{GPS}} TEC_{f_1,n}^{(II)} + \frac{1}{N_{GPS}} \sum_{n=1}^{N_{GPS}} TEC_{f_2,n}^{(II)} \right). \quad (14)$$

Нижние оценки дисперсии измерения времени задержки ретранслированного сигнала на ПП определяются модификацией формулы Вудворда для узкополосного радиосигнала со случайной начальной фазой [17]:

$$\sigma_{\tau}^2 = \frac{1}{(2E / N_0) 4\pi^2 \beta_f^2}, \quad (15)$$

где $2E / N_0$ – отношение удвоенной энергии принимаемого на ПП ретранслированного навигационного сигнала к спектральной плотности шума;

$$\beta_f^2 = \left(\int_{-\infty}^{\infty} f^2 |S(f)|^2 df - \int_{-\infty}^{\infty} f |S(f)|^2 df \right) / \int_{-\infty}^{\infty} |S(f)|^2 df - \text{нормированный второй момент}$$

квадрата модуля спектра принимаемого сигнала $|S(f)|^2$ относительно несущей частоты [17, с. 197].

Квадрат модуля спектра принимаемого сигнала $|S(f)|^2$ определяется квадратом модуля спектра фазоманипулированного навигационного сигнала $|S_{\text{фм}}(f)|^2 = \sin^2(\pi f \tau_0) / (\pi f)^2$ и ограничением его полосы при излучении с НС, ретрансляции и приеме ПП. С учетом указанных факторов будем полагать, что полоса частот принимаемого сигнала составляет $\Delta f_p = 1 / \tau_0 = K_{sh} / T_0$, где $\tau_0 = T_0 / K_{sh}$ – длительность символа дальномерного кода навигационного сигнала (для открытых навигационных сигналов GPS $\tau_0 = 0,978$ мкс независимо от частоты), K_{sh} – коэффициент сжатия навигационного сигнала, T_0 – длительность одного периода навигационного сигнала. Для указанных выше условий нетрудно вычислить $\beta_f^2 = 0,0655 / \tau_0^2 = 6,85 \cdot 10^{10}$ Гц².

Используя равенство (15) и взаимосвязь времени задержки и дальности $r = c\tau$, где c – скорость света, для дисперсии измерений суммарных дальностей запишем

$$\sigma_{f_{1(2)}, f_{p1(2)}}^2 = \frac{c^2}{(2\pi)^2 2\beta_f^2 \rho_{f_{1(2)}, f_{p1(2)}}} = \frac{B}{\rho_{f_{1(2)}, f_{p1(2)}}}, \quad (16)$$

где $B = \frac{c^2}{(2\pi)^2 2\beta_f^2} = 1,66 \cdot 10^4$ м⁴; $\rho_{f_{1(2)}, f_{p1(2)}} = E_{f_{1(2)}, f_{p1(2)}} / N_0$ – отношения сигнал/шум для соответствующих комбинаций «частота навигационного сигнала/частота ретрансляции».

Из выражений (13) и (16) следует, что дисперсии $D_{f_1}^{(II)}$, $D_{f_2}^{(II)}$ оценивания ПЭС на трассе НС – ПП составят

$$\begin{aligned} D_{f_{1(2)}}^{(II)} &= \frac{\sigma_{f_{1(2)}, f_{p1}}^2 + \sigma_{f_{1(2)}, f_{p2}}^2}{\left(A \left(\frac{1}{f_{p1}^2} - \frac{1}{f_{p2}^2} \right) \right)^2} = 7 \cdot 10^{33} \cdot \left(\frac{1}{\rho_{f_{1(2)}, f_{p1}}} + \frac{1}{\rho_{f_{1(2)}, f_{p2}}} \right) [\text{м}^{-4}] = \\ &= 70 \left(\frac{1}{\rho_{f_{1(2)}, f_{p1}}} + \frac{1}{\rho_{f_{1(2)}, f_{p2}}} \right) [\text{TECu}^2]. \end{aligned} \quad (17)$$

Для $\rho_{f_1, f_{p1}} = \rho_{f_1, f_{p2}} = 30$ дБ (как показано ниже, такое значение вполне достижимо для системы GPS при времени наблюдения порядка $0,5 \dots 1,0$ с) получим среднеквадратическую ошибку (СКО) оценивания ПЭС с использованием сигнала одного НС $\sqrt{D_{f_1}^{(II)}} = 0,374$ TECu.

Отметим, что достаточной для определения ПЭС на трассе CP – ПП является ретрансляция только одной из частот GPS. Поскольку эквивалентная изотропная излучаемая мощность на частоте f_2 GPS на $1,5 \dots 6$ дБ меньше, чем на частоте f_1 (за исключением НС GPS III) [2], при фиксированном коэффициенте ретрансляции СКО оценки ПЭС на ретранслированной частоте f_2 окажется больше примерно в два раза. В связи с этим вклад ретранслированного сигнала на частоте f_2 при весовом усреднении результатов на двух частотах окажется меньше, поэтому указанную частоту можно не ретранслировать. Результирующее значение СКО оценивания ПЭС на трассе CP – ПП с учетом ретрансляции только частоты f_1 и усреднения (14) по N_{GPS} видимым НС при $\rho = \rho_{f_1, f_{p1}} = \rho_{f_1, f_{p2}}$ определяется выражением

$$\sigma_{TEC_2} = \frac{11,8}{\sqrt{N_{GPS}\rho}} [\text{TECu}]. \quad (18)$$

В частности, при $\rho = 30$ дБ и $N_{GPS} = 6$ результирующая СКО составит $0,15$ TECu, что вполне достаточно для реализации радиотомографических методов. Отметим, что такой результат получен ввиду высокой точности измерения разности суммарных дальностей с применением модулированного сигнала при низких частотах ретрансляции в сравнении с частотами навигационных сигналов и усреднения результата по всем видимым НС.

Оценка ПЭС на трассе НС – CP. Как следует из выражений (12), при использовании дальностных измерений оценки ПЭС на трассе НС – CP аналогично случаю с трассой CP – ПП будут равны величинам

$$TEC_{f_{p1}}^{(I)} = \frac{\hat{d}_{f_1, f_{p1}} - \hat{d}_{f_2, f_{p1}}}{A \left(\frac{1}{f_1^2} - \frac{1}{f_2^2} \right)}, \quad TEC_{f_{p2}}^{(I)} = \frac{\hat{d}_{f_1, f_{p2}} - \hat{d}_{f_2, f_{p2}}}{A \left(\frac{1}{f_1^2} - \frac{1}{f_2^2} \right)}, \quad (19)$$

т. е. в данном случае результаты измерений на разных частотах комбинируются таким образом, чтобы исключить трассы CP – ПП. При этом можно ограничиться ретрансляцией частот f_1, f_2 только на одной из частот ретрансляции $f_{p1(2)}$. Аналогично (17) дисперсия измерений ПЭС составит

$$D_{f_{p1(2)}}^{(I)} = \frac{\sigma_{f_1, f_{p1(2)}}^2 + \sigma_{f_2, f_{p1(2)}}^2}{\left(A \left(\frac{1}{f_1^2} - \frac{1}{f_2^2} \right) \right)^2} = 1,5 \cdot 10^{38} \cdot \left(\frac{1}{\rho_{f_1, f_{p1(2)}}} + \frac{1}{\rho_{f_2, f_{p1(2)}}} \right) [\text{M}^{-4}] =$$

$$= 1,5 \cdot 10^6 \cdot \left(\frac{1}{\rho_{f_{1(2)}, f_{p1}}} + \frac{1}{\rho_{f_{1(2)}, f_{p2}}} \right) [\text{TECu}^2]. \quad (20)$$

При наличии $m=1 \dots M_{\text{пп}}$ ПП результаты оценивания ПЭС $TEC_{f_{p1}, m}^{(I)}$ на любой из трасс НС – CP могут быть усреднены. При использовании только первой частоты ретрансляции получаем равенство

$$\overline{TEC}^{(I)} = \frac{1}{M_{\text{пп}}} \sum_{m=1}^{M_{\text{пп}}} TEC_{f_{p1}, m}^{(I)}. \quad (21)$$

Из выражений (20), (21) следует, что для достижения высокой точности определения ПЭС необходимы большие значения отношения сигнал/шум (40...50 дБ) в единичных измерениях и использование большого числа ПП. Кроме того, возможно последующее усреднение на ограниченном временном интервале аналогично методу измерения ПЭС в двухчастотном навигационном приемнике при реализации высокоорбитальной томографии. Однако из-за высокой угловой скорости перемещения СР относительно НС и возможной неоднородности электронной концентрации в ионосфере время указанного усреднения меньше, чем при высокоорбитальной томографии.

Расчет ПЭС на трассе НС – СР на основе фазовых измерений по ретранслированным сигналам. Рассмотрим распространяющуюся плоскую радиоволну (2). С учетом выражений (4) и (5) определим фазу радиосигнала (2) следующим образом [15]:

$$\varphi(t) = 2\pi ft - \frac{2\pi f}{c} L(t) + \frac{2\pi A}{c} \frac{1}{f} \int_{L_0(t)}^{L(t)} n_e(z) dz - \frac{2\pi f}{c} \int_0^{L_0(t)} \Delta \bar{n}_r dz. \quad (22)$$

С целью упрощения дальнейших выкладок считаем, что на рис. 1 НС-3, СР и ПП-2 располагаются на одной линии. В случае ретрансляции сигнал (2) можно представить в виде двух последовательно излученных плоских электромагнитных волн, фазы которых будут соответственно определяться выражениями

$$\varphi^{(I)}(t) = 2\pi ft - \frac{2\pi f}{c} (L(t) - L_p(t)) + \frac{2\pi A}{c} \frac{1}{f} \int_{L_p(t)}^{L(t)} n_e(z) dz, \quad (23)$$

$$\varphi^{(II)}(t) = 2\pi f_p t - \frac{2\pi f_p}{c} L_p(t) + \frac{2\pi A}{c} \frac{1}{f_p} \int_{L_0(t)}^{L_p(t)} n_e(z) dz - \frac{2\pi f_p}{c} \int_0^{L_0(t)} \Delta \bar{n}_r dz + \Delta \varphi^{(I)}(t), \quad (24)$$

где $L(t), L_p(t)$ – расстояния от ПП до НС и СР; $\Delta \varphi^{(I)}(t) = \varphi(t) - 2\pi ft$ – сдвиг фазы ретранслированного сигнала, обусловленный распространением радиоволны по трассе НС – СР.

В соответствии со схемой (1) и выражением (23) принимаемая фаза (24) имеет вид

$$\varphi_{f_1, f_{p1}} = 2\pi f_{p1} t - \frac{2\pi f_1}{c} r^{(I)}(t) + \frac{2\pi A}{c f_1} TEC^{(I)}(t) - \frac{2\pi f_{p1}}{c} r^{(II)}(t) + \frac{2\pi A}{c f_{p1}} TEC^{(II)}(t) - \frac{2\pi f_{p1}}{c} \int_0^{L_0(t)} \Delta \bar{n}_r dz, \quad (25)$$

$$\varphi_{f_2, f_{p1}} = 2\pi f_{p1} t - \frac{2\pi f_2}{c} r^{(I)}(t) + \frac{2\pi A}{c f_2} TEC^{(I)}(t) - \frac{2\pi f_{p1}}{c} r^{(II)}(t) + \frac{2\pi A}{c f_{p1}} TEC^{(II)}(t) - \frac{2\pi f_{p1}}{c} \int_0^{L_0(t)} \Delta \bar{n}_r dz, \quad (26)$$

$$\varphi_{f_1, f_{p2}} = 2\pi f_{p2} t - \frac{2\pi f_1}{c} r^{(I)}(t) + \frac{2\pi A}{c f_1} TEC^{(I)}(t) - \frac{2\pi f_{p2}}{c} r^{(II)}(t) + \frac{2\pi A}{c f_{p2}} TEC^{(II)}(t) - \frac{2\pi f_{p2}}{c} \int_0^{L_0(t)} \Delta \bar{n}_r dz, \quad (27)$$

$$\varphi_{f_2, f_{p2}} = 2\pi f_{p2} t - \frac{2\pi f_2}{c} r^{(I)}(t) + \frac{2\pi A}{c f_2} TEC^{(I)}(t) - \frac{2\pi f_{p2}}{c} r^{(II)}(t) + \frac{2\pi A}{c f_{p2}} TEC^{(II)}(t) - \frac{2\pi f_{p2}}{c} \int_0^{L_0(t)} \Delta \bar{n}_r dz, \quad (28)$$

где $TEC^{(I)}(t) = \int_{L_p(t)}^{L(t)} n_e(z) dz$, $TEC^{(II)}(t) = \int_{L_0(t)}^{L(t)} n_e(z) dz$, $r^{(I)}(t) = L(t) - L_p(t)$, $r^{(II)}(t) = L_p(t)$. Вычитая из уравнения (25) уравнение (26), а из уравнения (27) уравнение (28), получим

$$\Delta\varphi_{f_{p1}}(t) = \varphi_{f_1, f_{p1}} - \varphi_{f_2, f_{p1}} = -\frac{2\pi(f_1 - f_2)}{c} \cdot r^{(1)}(t) + \frac{2\pi A}{c} \left(\frac{1}{f_1} - \frac{1}{f_2} \right) TEC^{(1)}(t), \quad (29)$$

$$\Delta\varphi_{f_{p2}}(t) = \varphi_{f_1, f_{p2}} - \varphi_{f_2, f_{p2}} = -\frac{2\pi(f_1 - f_2)}{c} \cdot r^{(1)}(t) + \frac{2\pi A}{c} \left(\frac{1}{f_1} - \frac{1}{f_2} \right) TEC^{(1)}(t). \quad (30)$$

Из равенства (29) нетрудно найти оценку ПЭС на трассе НС – СР:

$$TEC_{f_{p1}}^{(1)}(t) = \frac{\Delta\varphi_{f_{p1}}(t) + \frac{2\pi(f_1 - f_2)}{c} \cdot r^{(1)}(t)}{\frac{2\pi A}{c} \left(\frac{1}{f_1} - \frac{1}{f_2} \right)} = \frac{\widehat{\varphi}_{f_1, f_{p1}}(t) - \widehat{\varphi}_{f_2, f_{p1}}(t)}{\frac{2\pi A}{c} \left(\frac{1}{f_1} - \frac{1}{f_2} \right)} - \frac{f_1 f_2}{A} r^{(1)}(t), \quad (31)$$

где $\widehat{\varphi}_{f_1, f_{p1}}(t)$, $\widehat{\varphi}_{f_2, f_{p1}}(t)$ – оценки фаз соответствующих принимаемых сигналов. Так как точно определить $\widehat{\varphi}_{f_1, f_{p1}}(t)$, $\widehat{\varphi}_{f_2, f_{p1}}(t)$ с учетом фазовой неоднозначности практически невозможно, перейдем по аналогии с работами [1, 3] к фазоразностному методу, выполнив дифференцирование (31), которое убирает постоянную на интервале наблюдения неоднозначность фаз⁶:

$$\frac{d}{dt} TEC_{f_{p1}}^{(1)}(t) = \frac{\frac{d}{dt} \widehat{\varphi}_{f_1, f_{p1}}(t) - \frac{d}{dt} \widehat{\varphi}_{f_2, f_{p1}}(t)}{\frac{2\pi A}{c} \left(\frac{1}{f_1} - \frac{1}{f_2} \right)} - \frac{f_1 f_2}{A} \frac{d}{dt} r^{(1)}(t). \quad (32)$$

Подставляя значения констант и заданных частот в выражение (32), получим

$$\frac{d}{dt} TEC_{f_{p1}}^{(1)} = -6,6 \cdot 10^{15} \cdot \left(\frac{d}{dt} \widehat{\varphi}_{f_1, f_{p1}}(t) - \frac{d}{dt} \widehat{\varphi}_{f_2, f_{p1}}(t) \right) - 4,8 \cdot 10^{16} \cdot \frac{d}{dt} r^{(1)}(t) \text{ [TECу/с]}. \quad (33)$$

Приведенные выше соотношения для разности фаз принимаемых на двух частотах сигналов базируются на исходном выражении излучаемой плоской волны (2) и принимаемой плоской волны в точке приема. Они позволяют установить влияние ПЭС на разность фаз принимаемых сигналов с учетом ретрансляции. Однако навигационные сигналы ГНСС являются модулированными. Для демодуляции принимаемого сигнала на ПП необходимо умножить его на совмещенный по времени комплексно-сопряженный опорный сигнал с фазовой манипуляцией дальномерного кода навигационного спутника и навигационного сообщения и с медленной фазовой модуляцией, обусловленной взаимным пространственным перемещением НС, СР и ПП. Для этого требуется более детальная оценка фаз принимаемых сигналов и их разностей.

Запишем фазу модулированного сигнала, принимаемого на ПП. Преобразование фазы сигнала на любой из частот f навигационного сигнала и на любой из ретрансляционных частот f_p в процессе ретрансляции показано на рис. 2.

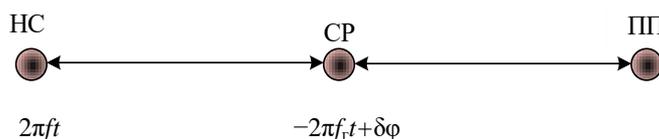


Рис. 2. Преобразование фазы сигнала при ретрансляции

Fig. 2. Signal phase conversion during relaying

⁶Смирнов, В. М. Определение электронного содержания ионосферы Земли по данным дальномерных и фазовых измерений / В. М. Смирнов, В. Н. Марчук // Исследовано в России. – Режим доступа: <https://cyberleninka.ru/article/n/opredelenie-elektronnogo-soderzhaniya-ionosfery-zemli-po-dannym-dalnomernyh-i-fazovyh-izmereniy>. – Дата доступа: 18.10.2022.

Фаза $\varphi_{\text{III}}(t)$ принимаемого на ПП в момент времени t сигнала (без учета ПЭС на трассе) составит [2, 18] $\varphi_{\text{III}}(t) = \varphi_{\text{CP}}\left(t - \frac{r^{(\text{II})}(t)}{c}\right)$, где $\varphi_{\text{CP}}(\tau)$ – фаза модулированного сигнала, излучаемого СР. В свою очередь, для фазы сигнала, излучаемого СР, можно записать

$$\begin{aligned}\varphi_{\text{CP}}(\tau) &= \varphi_{\text{НС}}\left(\tau - \frac{r^{(\text{I})}(\tau)}{c}\right) - 2\pi f_{\text{r}}\tau + \delta\varphi = 2\pi f\left(\tau - \frac{r^{(\text{I})}(\tau)}{c}\right) - 2\pi f_{\text{r}}\tau + \delta\varphi = \\ &= 2\pi(f - f_{\text{r}})\tau - 2\pi\frac{f}{c}r^{(\text{I})}(\tau) + \delta\varphi = 2\pi f_{\text{p}}\tau - 2\pi\frac{f}{c}r^{(\text{I})}(\tau) + \delta\varphi,\end{aligned}\quad (34)$$

где $f_{\text{p}} = f - f_{\text{r}}$ – частота ретранслированного сигнала, $\delta\varphi$ – фазовый сдвиг в ретрансляторе, f_{r} – частота гетеродина.

Объединив этапы на трассах (I) и (II), получим

$$\begin{aligned}\varphi_{\text{III}}(t) &= 2\pi f\left(t - \frac{r^{(\text{II})}(t)}{c} - \frac{1}{c}r^{(\text{I})}\left(t - \frac{r^{(\text{II})}(t)}{c}\right)\right) - 2\pi f_{\text{r}}\left(t - \frac{1}{c}r^{(\text{II})}(t)\right) + \delta\varphi = \\ &= 2\pi f_{\text{p}}t - 2\pi\frac{f_{\text{p}}}{c}r^{(\text{II})}(t) - 2\pi\frac{f}{c}r^{(\text{I})}\left(t - \frac{r^{(\text{II})}(t)}{c}\right) + \delta\varphi.\end{aligned}\quad (35)$$

Выражение (35) определяет закон фазовой модуляции сигнала, обусловленной взаимным перемещением НС, СР и ПП. Конкретизируем его относительно принимаемых сигналов и изменяющихся во времени ПЭС $TEC^{(\text{I})}(t)$ и $TEC^{(\text{II})}(t)$ на трассах распространения при ретрансляции частот f_1, f_2 на частоту ретрансляции f_{p1} :

$$\begin{aligned}\varphi_{f_1, f_{\text{p1}}}(t) &= 2\pi f_{\text{p1}}t - 2\pi\frac{f_{\text{p1}}}{c}r^{(\text{II})}(t) - 2\pi\frac{f_1}{c}r^{(\text{I})}\left(t - \frac{r^{(\text{II})}(t)}{c}\right) + \\ &+ \frac{2\pi A}{c}\left(\frac{TEC^{(\text{I})}(t)}{f_1} + \frac{TEC^{(\text{II})}(t)}{f_{\text{p1}}}\right) + \delta\varphi_{f_1, f_{\text{p1}}}; \\ \varphi_{f_2, f_{\text{p1}}}(t) &= 2\pi f_{\text{p1}}t - 2\pi\frac{f_{\text{p1}}}{c}r^{(\text{II})}(t) - 2\pi\frac{f_2}{c}r^{(\text{I})}\left(t - \frac{r^{(\text{II})}(t)}{c}\right) + \\ &+ \frac{2\pi A}{c}\left(\frac{TEC^{(\text{I})}(t)}{f_2} + \frac{TEC^{(\text{II})}(t)}{f_{\text{p1}}}\right) + \delta\varphi_{f_2, f_{\text{p1}}},\end{aligned}\quad (36)$$

где $\delta\varphi_{f_1(2), f_{\text{p1}}(2)}$ – фазовые сдвиги для различных комбинаций «частота навигационного сигнала / частота ретрансляции» в трактах ретранслятора. Из выражений (36) найдем

$$\begin{aligned}\varphi_{f_1, f_{\text{p1}}}(t) - \varphi_{f_2, f_{\text{p1}}}(t) &= \delta\varphi_{f_1, f_{\text{p1}}} - \delta\varphi_{f_2, f_{\text{p1}}} - 2\pi\frac{f_1 - f_2}{c}r^{(\text{I})}\left(t - \frac{r^{(\text{II})}(t)}{c}\right) + \\ &+ \frac{2\pi A}{c} \cdot TEC^{(\text{I})}(t) \cdot \left(\frac{1}{f_1} - \frac{1}{f_2}\right),\end{aligned}\quad (37)$$

откуда следует

$$TEC_{f_{p1}}^{(I)} = \frac{\widehat{\varphi}_{f_1, f_{p1}}(t) - \widehat{\varphi}_{f_2, f_{p1}}(t) - (\delta\varphi_{f_1, f_{p1}} - \delta\varphi_{f_2, f_{p1}}) + 2\pi \frac{f_1 - f_2}{c} r^{(I)} \left(t - \frac{r^{(II)}(t)}{c} \right)}{\frac{2\pi A}{c} \cdot \left(\frac{1}{f_1} - \frac{1}{f_2} \right)}, \quad (38)$$

где $\widehat{\varphi}_{f_1, f_{p1}}(t)$, $\widehat{\varphi}_{f_2, f_{p1}}(t)$ – оценки фаз соответствующих принимаемых сигналов, формируемые в аппаратуре ПП.

Так как получение $\widehat{\varphi}_{f_1, f_{p1}}(t)$, $\widehat{\varphi}_{f_2, f_{p1}}(t)$ с учетом фазовой неоднозначности практически невозможно, перейдем аналогично [1, 3] к фазоразностному методу, выполнив дифференцирование (38), убирающее постоянную на интервале наблюдения неоднозначность фаз:

$$\begin{aligned} \frac{d}{dt} TEC_{f_{p1}}^{(I)} &= \frac{\frac{d}{dt} \widehat{\varphi}_{f_1, f_{p1}}(t) - \frac{d}{dt} \widehat{\varphi}_{f_2, f_{p1}}(t)}{\frac{2\pi A}{c} \cdot \left(\frac{1}{f_1} - \frac{1}{f_2} \right)} - \frac{f_1 f_2}{A} \frac{d}{dt} \left(r^{(I)} \left(t - \frac{r^{(II)}(t)}{c} \right) \right) [\text{м}^{-4}] = \\ &= -6,6 \cdot 10^{15} \cdot \left(\frac{d}{dt} \widehat{\varphi}_{f_1, f_{p1}}(t) - \frac{d}{dt} \widehat{\varphi}_{f_2, f_{p1}}(t) \right) - 4,8 \cdot 10^{16} \frac{d}{dt} \left(r^{(I)} \left(t - \frac{r^{(II)}(t)}{c} \right) \right) [\text{ТЕСм}/\text{с}]. \quad (39) \end{aligned}$$

Как видим, выражения (38) и (39) схожи с (31)–(33), но получены с помощью разных подходов, что связано с учетом модуляции и конечной скорости распространения волн.

Скорости изменения фаз $\frac{d}{dt} \widehat{\varphi}_{f_1, f_{p1}}(t)$ и $\frac{d}{dt} \widehat{\varphi}_{f_2, f_{p1}}(t)$ в первом слагаемом (39) представляют собой доплеровские сдвиги частот принимаемого сигнала и могут быть оценены непосредственно в процессе его обработки. Кроме того, они могут быть получены в соответствующих точках системы фазовой автоматической подстройки частоты каналов приемника ретранслированных сигналов. Второе слагаемое в (39) учитывает взаимное перемещение НС и СР. Его вычисление предполагает многократное измерение координат СР на всем интервале наблюдения, оценивание орбитальных параметров спутника методом наименьших квадратов [11], получение точных ультрабыстрых эфемерид НС, аппроксимацию дальностей $r^{(I)}(t)$, $r^{(II)}(t)$ и вычисление производной для заданной аппроксимации. Требуемая точность вычисления радиальной скорости СР относительно НС ввиду большого множителя во втором слагаемом (39) составляет доли или единицы см/с.

Законы фазовой модуляции ретранслированного сигнала (36) дают возможность записать временные представления принимаемых на ПП сигналов от n -го НС в виде

$$\dot{S}_{f_1, f_{p1(2)}}^{(n)}(t) = A_{f_1, f_{p1(2)}}^{(n)} D_{f_1}^{(n)}(t - \tau_{f_1, f_{p1(2)}}(t)) X_{f_1}^{(n)}(t - \tau_{f_1, f_{p1(2)}}(t)) e^{j\varphi_{f_1, f_{p1(2)}}(t)}, \quad (40)$$

$$\dot{S}_{f_2, f_{p1(2)}}^{(n)}(t) = A_{f_2, f_{p1(2)}}^{(n)} D_{f_2}^{(n)}(t - \tau_{f_2, f_{p1(2)}}(t)) X_{f_2}^{(n)}(t - \tau_{f_2, f_{p1(2)}}(t)) e^{j\varphi_{f_2, f_{p1(2)}}(t)},$$

где $D_{f_1}^{(n)}(t)$, $D_{f_2}^{(n)}(t)$, $X_{f_1}^{(n)}(t)$, $X_{f_2}^{(n)}(t)$ – навигационные сообщения и дальномерные коды n -го НС [2] на соответствующих частотах f_1 и f_2 ; $A_{f_1(2), f_{p1(2)}}^{(n)}$ – амплитуды сигналов на входе прием-

⁷URL: <https://www.glonass-iac.ru>

ника наземного ПП; $\tau_{f_{1(2)}, f_{p1(2)}}$ – задержки сигналов от n -го НС на трассе распространения НС – СР – ПП, определяемые по формуле

$$\tau_{f_{1(2)}, f_{p1(2)}}(t) = \frac{1}{c} \left(r^{(I)} \left(t - \frac{r^{(II)}(t)}{c} \right) + \frac{A \cdot TEC^{(I)}(t)}{f_{1(2)}^2} + r^{(II)}(t) + \frac{A \cdot TEC^{(II)}(t)}{f_{p1(2)}^2} \right). \quad (41)$$

Дальномерные коды $X_{f_1}^{(n)}(t)$, $X_{f_2}^{(n)}(t)$ известны из работы [2], навигационные сообщения $D_{f_1}^{(n)}(t)$, $D_{f_2}^{(n)}(t)$ могут быть оценены при приеме на ПП прямых навигационных сигналов от НС. Отметим, что навигационные сообщения и дальномерные коды принимают значения ± 1 . Это соответствует фазовой манипуляции $0, \pi$. Задержки $\tau_{f_{1(2)}, f_{p1(2)}}$ с неопределенностью, связанной с неизвестными значениями $TEC^{(I)}$, $TEC^{(II)}$, и законы фазовой модуляции $\varphi_{f_{1(2)}, f_{p1(2)}}(t)$ оцениваются по известным с высокой точностью координатам НС, ПП и прогнозируемым координатам СР. Представления (40) обуславливают возможность приема и когерентного накопления сигналов на ПП путем фазовой демодуляции принимаемых сигналов по навигационным сообщениям, дальномерным кодам и законам $\varphi_{f_{1(2)}, f_{p1(2)}}(t)$ и возможность их накопления в узкополосных фильтрах аналогично [14].

Структура аппаратуры ретрансляции и расчет ее требуемых характеристик. Структурно-функциональная схема аппаратуры ретрансляции изображена на рис. 3. С выхода микрополосковой приемной антенны A_{in} на частотный диапазон 1,2...1,6 ГГц принятые навигационные сигналы поступают на малошумящий усилитель (МШУ) и далее разветвляются на два полосно-пропускающих фильтра (ППФ) с центральными частотами 1575,42 и 1227,6 МГц и полосами пропускания около 2 МГц, что соответствует ширине спектра навигационного сигнала по первым нулям. Сигналы с выхода фильтров поступают на четыре балансных смесителя (БС). На вторые входы БС поступают сигналы гетеродинов, формируемые из сигналов высокостабильного опорного генератора (ОГ) с использованием синтезаторов частоты с фиксированными коэффициентами деления частот опорного сигнала и генератора, управляемого напряжением, с частотами 1575,42–150, 1575,42–400, 1227,6–150 и 1227,6–400 МГц. Сигналы выходов БС суммируются и усиливаются в цепочке усилителей на частотах 150 и 400 МГц и поступают на согласующие устройства с фазовращателями СУ+ФВ, имеющие два выхода со сдвигом сигналов на $\pi/2$. Квадратурные выходные сигналы СУ+ФВ на рабочих частотах 150 и 400 МГц суммируются и поступают на элементы турникетной передающей антенны.

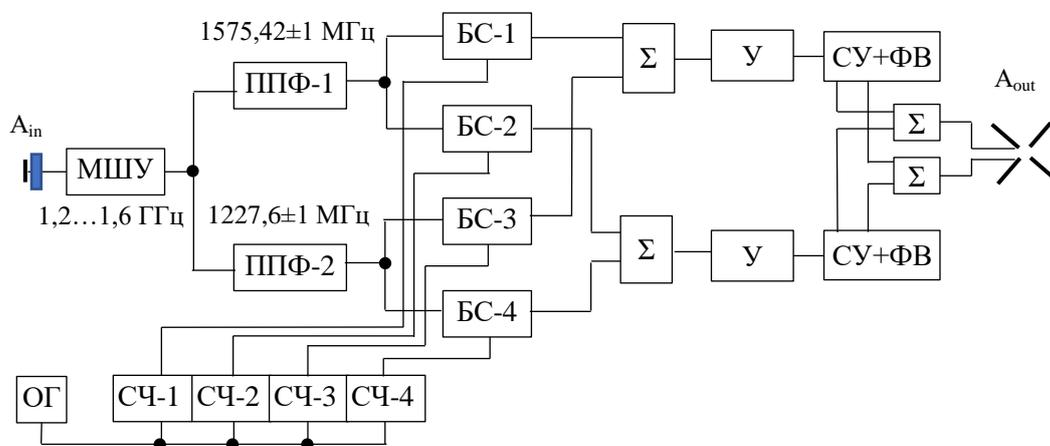


Рис. 3. Структурно-функциональная схема аппаратуры ретрансляции
Fig. 3. Structural and functional diagram of relay equipment

Основными параметрами аппаратуры являются эффективный (с учетом всех потерь в тракте) коэффициент ретрансляции и мощность выходного усилителя. Определим указанные параметры при следующих исходных данных: эквивалентная изотропная излучаемая мощность навигационного сигнала на каждой из частот по кодам C/A, L2OC $P_{\text{HC}} = 26,8$ дБВт; дальность между НС и СР $r_1 = 21\,000$ км; дальность между СР и ПП $r_2 = 1000$ км; спектральная плотность мощности внутренних шумов аппаратуры ретрансляции и приемного тракта ПП $N_0 = 5,8 \cdot 10^{-21}$ Вт/Гц ($K_{\text{ш}} = 1,5$ дБ, $T = 300$ К); средний по рабочему сектору коэффициент усиления микрополосковой приемной антенны аппаратуры ретрансляции с учетом потерь в тракте до МШУ $G_r = 1,5$; коэффициент усиления передающей антенны аппаратуры ретрансляции $G_{tr} = 1$; коэффициент усиления антенны ПП (трехэлементной антенны Яги – Уда с ориентацией на СР) $G_{\text{ПП}} = 7,3$ дБ; коэффициент потерь полезного сигнала в приемном тракте ПП $k_{\text{ПП}} = 0,5$; коэффициент поляризационных потерь $k_{\pi} = 0,5$.

Отношение сигнал/шум на выходе согласованного фильтра для одного периода навигационного сигнала определяется выражением

$$\gamma = \frac{P_{\text{HC}} G_r G_{tr} G_{\text{ПП}} \lambda_{\text{HC}}^2 \lambda_p^2 k_{\text{ПП}} k_{\pi} T_0}{(4\pi)^4 r_1^2 r_2^2 N_0} K_p, \quad (42)$$

где $\lambda_{\text{HC}} \approx 0,19$ – наименьшая длина волны для навигационного сигнала, $\lambda_p = 0,75$ или 2 м – длина волны ретранслированного сигнала, $T_0 = 1$ мс – длительность сигнала для частоты f_1 .

Основной особенностью рассматриваемого ретранслятора является существенное усиление собственных шумов, так как мощность принимаемого навигационного сигнала (как и для типового навигационного приемника) на $15...20$ дБ [2] меньше мощности шума МШУ. Коэффициент ретрансляции ограничен мощностью усиленного шума

$$P_{\text{ш}} = N_0 \Delta f_p K_p, \quad (43)$$

где $\Delta f_p = 4$ МГц – общая полоса пропускания ретранслятора. Отношение сигнал/шум при когерентном накоплении в течение времени $T_{\text{кн}}$ составит

$$\rho(K_p) = \gamma(K_p) \frac{T_{\text{кн}}}{T_0}. \quad (44)$$

Зависимости отношения сигнал/шум, мощности шума на выходе ретранслятора и СКО оценивания ПЭС на трассах НС – СР ($\sigma_{\text{ТЕС1}}$) и СР – ПП ($\sigma_{\text{ТЕС2}}$) от K_p показаны на рис. 4. Зависимости построены согласно выражениям (15), (19) для случая определения ПЭС на трассе СР – ПП при ретрансляции одной частоты f_1 и наличии одного и шести НС, а для случая определения ПЭС на трассе НС – СР зависимости построены при ретрансляции двух частот f_1, f_2 на частоту ретрансляции 150 МГц, что более энергетически выгодно, чем на частоту 400 МГц, при наличии одного и пяти ПП. Пунктирные и сплошные линии на рис. 4, *a* соответствуют частотам ретрансляции 150 и 400 МГц, зависимости на рис. 4, *b* построены для времени когерентного накопления 1 с.

На рис. 4 видно, что достижение СКО оценки ПЭС на трассе СР – ПП порядка $0,1$ ТЕСу при наличии в зоне видимости шести спутников GPS возможно при коэффициенте ретрансляции порядка 128 дБ и мощности шума на выходе ретранслятора $0,2$ Вт. Отношения сигнал/шум на выходе накопителя составят $32,3$ и $40,8$ дБ для частот ретрансляции 400 и 150 МГц соответственно. При мощности изучения СР $0,2$ Вт, высоте орбиты 500 км и ширине спектра шу-

ма 4 МГц спектральная плотность мощности сигнала на выходе расположенной на Земле изотропной приемной антенны на согласованной поляризации будет примерно на 3 дБ ниже стандартной спектральной плотности мощности шумов стандартного приемника 10^{-20} Вт/Гц, т. е. влияние излучения СР на другие радиоэлектронные средства будет пренебрежимо малым. Следует отметить, что при оценивании ПЭС на трассе СР – ПП при ретрансляции только одной частоты f_1 на частоты 150/400 МГц можно использовать и спутники ГЛОНАСС. Однако это приведет к необходимости расширения общей полосы ретрансляции до примерно 10 МГц либо последовательной работы по НС на разных частотах и ухудшению точности оценивания ПЭС из-за меньшего в два раза в сравнении с GPS коэффициента сжатия навигационного сигнала [2].

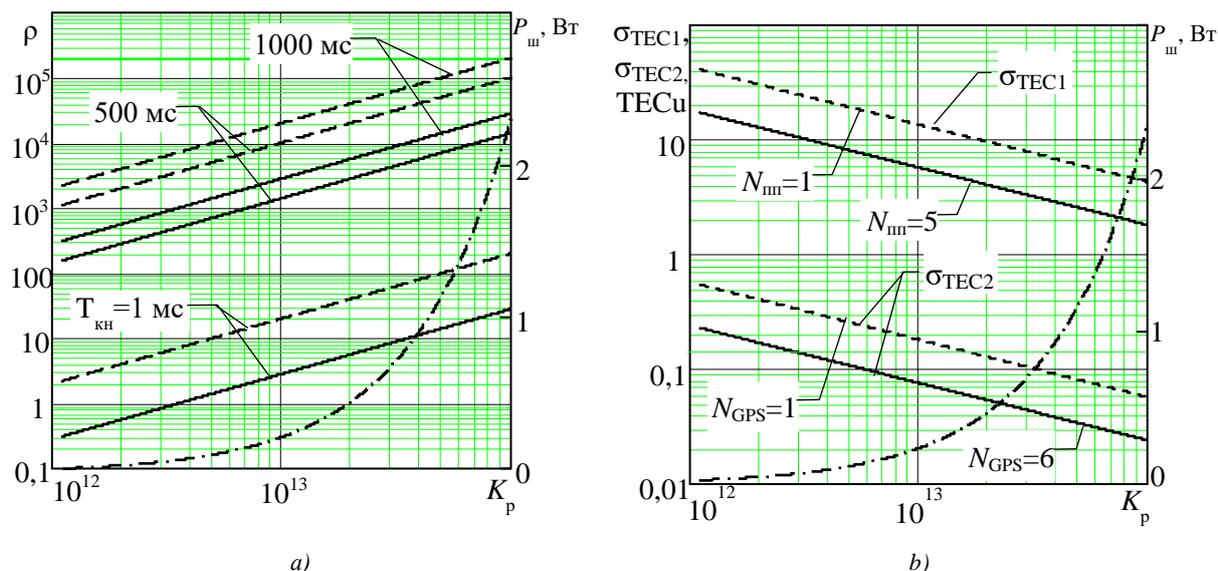


Рис. 4. Зависимость отношения сигнал/шум на выходе ретранслятора (а) и СКО оценивания ПЭС (b) от коэффициента ретрансляции
Fig. 4. Dependence of the signal-to-noise ratio at the output of the repeater (a) and RMS of TEC estimation (b) on the relay coefficient

Достижение приемлемой точности оценивания ПЭС на трассе НС – СР дальностным методом требует больших значений коэффициента ретрансляции и, соответственно, большей его мощности. Так, при ограничении выходной мощности ретранслятора величиной 1 Вт требуемый коэффициент ретрансляции составит 137 дБ при СКО оценки ПЭС для одного и пяти ПП порядка 8 и 3 TECu соответственно. Анализ точности фазоразностного оценивания ПЭС на трассе НС – СР выходит за рамки настоящей статьи.

Точность измерения ПЭС на трассах СР – ПП дальностным методом является вполне достаточной. Резервы повышения точности измерения ПЭС на трассах НС – СР наряду с усреднением по времени наблюдения заключаются в увеличении числа ПП или коэффициента усиления приемных антенн аппаратуры ПП.

Результаты моделирования. Моделирование предлагаемого способа оценивания ПЭС проводилось путем воспроизведения движения НС и СР по заданным орбитам и ПП на поверхности Земли, имитации неравномерной электронной концентрации в ионосфере, расчета истинных значений ПЭС на трассах распространения, воспроизведения принимаемых сигналов и их обработки, а также оценивания ПЭС на трассе СР – ПП.

Рассматриваемая ситуация изображена на рис. 5, где горизонтальные и вертикальные оси соответствуют географической долготе и широте, а значения электронной концентрации в виде вертикального ПЭС представлены цветовой шкалой (справа) в TECu. СР, показанный в виде квадрата со стрелочкой, отображающей его скорость, пролетает через зоны прямой видимости пяти ПП (ромбы) на временном интервале видимости около 500 с. Высота орбиты СР составля-

ет 500 км. Один из восьми видимых НС GPS, попадающих в заданный масштаб изображения, показан красным кружком. Высота максимума электронной концентрации совпадает с высотой орбиты СР. В этих условиях примерно половина из представленных на рис. 5 значений вертикального ПЭС будет относиться к трассам СР – ПП, а вторая половина – к трассам НС – СР. Значения ПЭС на указанных трассах могут в два-три раза превышать величину вертикального ПЭС ввиду большего пути в ионосфере.

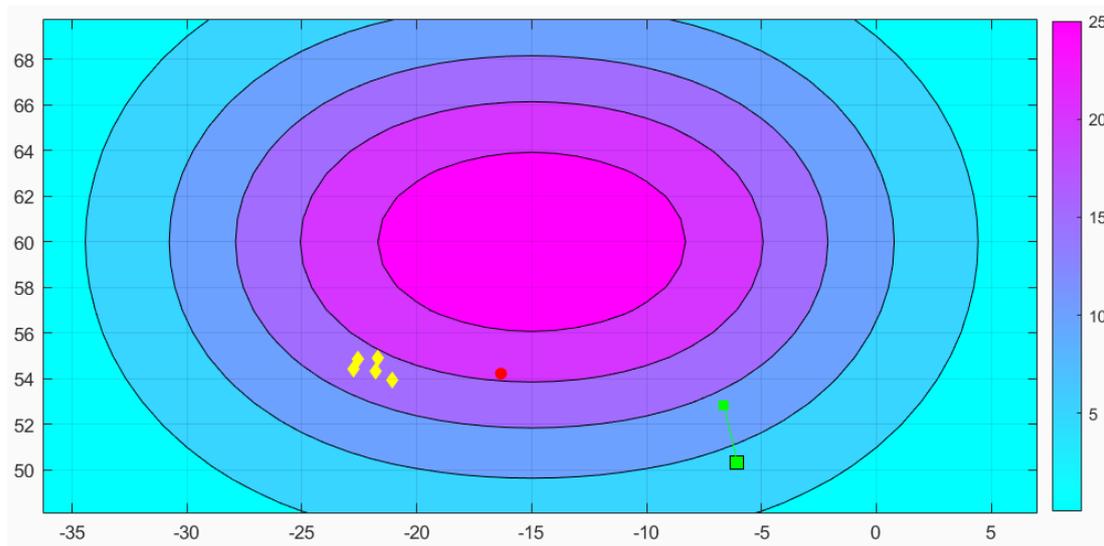


Рис. 5. Ситуация наблюдения

Fig. 5. Surveillance situation

На рис. 6 приведены значения истинных и измеренных ПЭС на трассе СР – первый ПП (квадратный маркер, на рис. 5 крайний слева) и на трассе СР – пятый ПП (круглый маркер, на рис. 5 крайний справа). Расчеты проводились при коэффициенте ретрансляции 130 дБ. Истинные значения ПЭС отражены соответствующими сплошными линиями.

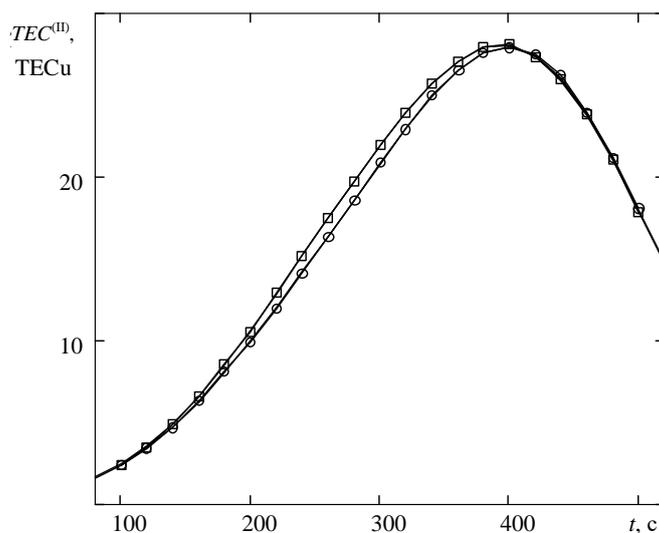


Рис. 6. Результаты оценивания ПЭС на трассах СР – ПП

Fig. 6. TEC evaluation results on the SR – RP routes

Значения ПЭС для двух ПП незначительно различаются в соответствии с геометрией наблюдения. Истинные значения ПЭС изменяются в пределах 1...27 TECu. Измеренные ПЭС

на трассе СР – ПП являются точными, а СКО не превышает 0,1 ТЕСц, что соответствует теоретическому значению при достигаемом отношении сигнал/шум по ретранслированному навигационному сигналу.

Заключение. В статье предложен новый способ определения ПЭС в ионосфере на основе ретрансляции навигационных сигналов ГНСС GPS наноспутником-ретранслятором. Аппаратура ретрансляции при оценивании ПЭС на трассе СР – ПП имеет малую, порядка 0,2 Вт, мощность излучения и максимально простую структурную схему, не использующую цифровые устройства с низкой радиационной стойкостью. Дальнейшие исследования связаны с моделированием обработки ретранслированного сигнала, оценкой эффективности фазового метода определения ПЭС, а также высокоточным определением координат и параметров движения СР по принимаемому ретранслированному сигналу.

Вклад авторов. *И. В. Белоконов* проводил консультации по техническим параметрам и применению наноспутников. *А. М. Крот* вывел аналитические соотношения определения ПЭС. *С. В. Козлов* сформулировал задачи исследования, обосновал временное представление сигналов. *Е. А. Каплярчук* разработал дальностный метод определения ПЭС на основе ретранслированного сигнала, определил требования к аппаратуре ретрансляции, провел расчеты и моделирование. *И. Э. Савиных* определила требования к данным и процедуры измерения ПЭС для выполнения радиотомографии ионосферы. *А. С. Шапкин* обосновал фазовый метод вычисления ПЭС на основе ретранслированного сигнала.

Список использованных источников

1. Куницын, В. Е. Радиотомография ионосферы / В. Е. Куницын, Е. Д. Терещенко, Е. С. Андреева. – М. : Физматлит, 2007. – 336 с.
2. ГЛОНАСС. Принципы построения и функционирования / под ред. А. И. Перова, В. Н. Харисова. – 4-е изд., перераб. и доп. – М. : ИПРЖР, 2010. – 800 с.
3. Афраймович, Э. Л. GPS-мониторинг верхней атмосферы Земли / Э. Л. Афраймович, Н. П. Первалова. – Иркутск : ГУ НЦ ВСНЦ СО РАН, 2006. – 480 с.
4. Семейство наноспутников изучения ионосферы на базе платформы SamSat разработки Самарского университета / И. В. Белоконов [и др.] // Восьмой Белорусский космический конгресс : материалы конгресса : в 2 т., Минск, 25–27 окт. 2022 г. – Минск : ОИПИ НАН Беларуси, 2022. – Т. 1. – С. 167–170.
5. Application of FORMOSAT-3/COSMIC mission to global Earth monitoring / C.-J. Fong [et al.] // Space 2005, Long Beach, California, 30 Aug. – 01 Sept. 2005. – Long Beach, 2005. – P. 6774. <https://doi.org/10.2514/6.2005-6774>
6. Романов, А. А. Измерение полного электронного содержания ионосферы Земли с помощью много-частотного когерентного зондирующего сигнала / А. А. Романов, А. В. Новиков // Вопросы электромеханики. Тр. НПП ВНИИЭМ. – М. : ФГУП «НПП ВНИИЭМ», 2009. – Т. 111. – С. 31–36.
7. Анализ современных возможностей создания малых космических аппаратов для дистанционного зондирования Земли / Н. Н. Севастьянов [и др.] // Тр. МФТИ. – 2009. – Т. 1, № 3. – С. 15–23.
8. Использование сигналов ГНСС для исследования состояния ионосферы / И. В. Белоконов [и др.] // Навигация и управление движением : тез. докл. Междунар. семинара, Самара, 28 сент. – 2 окт. 2020 г. / Самар. нац. исслед. ун-т им. С. П. Королёва. – Самара, 2020. – С. 85–86.
9. Николаев, П. Н. Алгоритм быстрой обратной проекции с фильтрацией в 2D-ионосферной радиотомографии с использованием межспутниковых измерений / П. Н. Николаев, О. В. Филонин, И. В. Белоконов // Advances in Space Research. – 2021. – Т. 68, № 10. – С. 4167–4188. <https://doi.org/10.1016/j.asr.2021.07.042>
10. Ширман, Я. Д. Теория и техника обработки радиолокационной информации на фоне помех / Я. Д. Ширман, В. Н. Манжос. – М. : Радио и связь, 1981. – 416 с.
11. Саврасов, Ю. С. Алгоритмы и программы в радиолокации / Ю. С. Саврасов. – М. : Радио и связь, 1985. – 216 с.
12. Купряшкин, И. Ф. Малогабаритные многофункциональные РЛС с непрерывным частотно-модулированным излучением. – М. : Радиотехника, 2020. – 280 с.
13. Ле, В. К. Алгоритмы длительного когерентного накопления отраженного сигнала при ненулевых высших производных дальности до радиолокационной цели в спектральной области / В. К. Ле, С. В. Козлов // Докл. БГУИР. – 2021. – № 5. – С. 35–44. <http://dx.doi.org/10.35596/1729-7648-2021-19-5-35-44>

14. Куан, Н. В. Пассивная радиолокационная система мониторинга движения судов в прибрежных районах с использованием спутниковых сигналов подсвета / Н. В. Куан // Изв. высших учебных заведений России. Радиоэлектроника. – 2020. – Т. 23, № 3. – С. 41–52. <https://doi.org/10.32603/1993-8985-2020-23-3-41-52>
15. Beacon satellite receiver for ionospheric tomography / J. Vierinen [et al.] // Radio Science. – 2014. – Vol. 49, iss. 12. – P. 1141–1152. <https://doi.org/10.1002/2014RS005434>
16. Дэвис, К. Радиоволны в ионосфере : пер. с англ. / К. Дэвис. – М. : Мир, 1973. – 502 с.
17. Тихонов, В. И. Оптимальный прием сигналов. – М. : Радио и связь, 1983. – 320 с.
18. Генике, А. А. Глобальные спутниковые системы определения местоположения и их применение в геодезии / А. А. Генике, Г. Г. Побединский. – М. : Картгеоцентр, 2004. – 355 с.

References

1. Kunitsyn V. E., Tereshchenko E. D., Andreeva E. S. Radiotomografija ionosfery. *Radio Tomography of the Ionosphere*. Moscow, Fizmatlit, 2007, 336 p. (In Russ.).
2. GLONASS. Principy postroenija i funkcionirovanija. *GLONASS. Principles of Construction and Functioning*. In A. I. Perov, V. N. Kharisov (eds.). 4th ed. Moscow, IPRZhR, 2010, 800 p. (In Russ.).
3. Afraimovich E. L., Perevalova N. P. GPS-monitoring verhnej atmosfery Zemli. *GPS Monitoring of the Earth's Upper Atmosphere*. Irkutsk, Gosudarstvennoe uchrezhdenie Nauchnyj centr Vostochno-Sibirskogo nauchnogo centra Sibirskogo otdelenija Rossijskoj akademii nauk, 2006, 480 p. (In Russ.).
4. Belokonov I. V., Boltov E. A., Elisov N. A., Lomaka I. A., Nikolaev P. N., Shafran S. V. *A family of nanosatellites for studying the ionosphere based on the SamSat platform developed by Samara University*. Vos'moj Belorusskij kosmicheskij kongress : materialy kongressa : v 2 t., Minsk, 25–27 oktjabrja 2022 g. [*Eighth Belarusian Space Congress : Materials of the Congress : in 2 Volumes, Minsk, 25–27 October 2022*]. Minsk, Ob"edinennyj institut problem informatiki Nacional'noj akademii nauk Belarusi, 2022, pp. 167–170 (In Russ.).
5. Fong C.-J., Wu B.-H., Yen N., Chen P. Application of FORMOSAT-3/COSMIC mission to global Earth monitoring. *Space 2005, Long Beach, California, 30 August – 01 September 2005*. Long Beach, 2005, p. 6774. <https://doi.org/10.2514/6.2005-6774>
6. Romanov A. A., Novikov A. V. *Measurement of the total electron content of the Earth's ionosphere using a multi-frequency coherent sounding signal*. Voprosy jelektromehaniki. Trudy Nauchno-proizvodstvennogo predprijatija Vserossijskogo nauchno-issledovatel'skogo instituta jelektromehaniki [*Questions of electromechanics. Proceedings of the Research and Production Enterprise of the All-Russian Research Institute of Electromechanics*], Moscow, Federal'noe gosudarstvennoe unitarnoe predpriatie "Nauchno-proizvodstvennoe predpriatie – Vserossijskij nauchno-issledovatel'skij institut jelektromehaniki s zavodom im. A. G. Iosif'jana", 2009, vol. 111, pp. 31–36 (In Russ.).
7. Sevastyanov N. N., Branets V. N., Panchenko V. A., Kazinsky N. V., Kondranin T. V., Negodyaev S. S. *Analysis of Modern Possibilities of Creating Small Spacecraft for Remote Sensing of the Earth*. Trudy Moskovskogo fiziko-tehnicheskogo instituta [*Proceedings of the Moscow Institute of Physics and Technology*], 2009, vol. 1, no. 3, pp. 15–23 (In Russ.).
8. Belokonov I. V., Krot A. M., Kumarin A. A., Nikolaev P. N., Filonin O. V. *Use of GNSS signals to study the state of the ionosphere*. Navigacija i upravlenie dvizheniem : tezisy dokladov Mezhdunarodnogo seminar, Samara, 28 sentjabrja – 2 oktjabrja 2020 g. Samarskij nacional'nyj issledovatel'skij universitet imeni akademika S. P. Koroleva. *Navigation and Traffic Control : Abstracts of Reports of the International Seminar, Samara, 28 September – 2 October 2020*. Samara, Samara National Research University named after academician S. P. Korolev, 2020, pp. 85–86 (In Russ.).
9. Nikolaev P. N., Filonin O. V., Belokonov I. V. Fast back projection algorithm with filtering in 2D ionospheric radio tomography using intersatellite measurements. *Advances in Space Research*, 2021, vol. 68, no. 10, pp. 4167–4188. <https://doi.org/10.1016/j.asr.2021.07.042>
10. Shirman Ya. D., Manzhos V. N. Teorija i tehnika obrabotki radiolokacionnoj informacii na fone pomeh. *Theory and Technique of Processing Radar Information Against the Background of Interference*. Moscow, Radio i svjaz', 1981, 416 p. (In Russ.).
11. Savrasov Yu. S. Algoritmy i programmy v radiolokacii. *Algorithms and Programs in Radar*. Moscow, Radio i svjaz', 1985, 216 p. (In Russ.).
12. Kupryashkin I. F. Malogabaritnye mnogofunktional'nye RLS s nepreryvnym chastotno-modulirovannym izlucheniem. *Small-sized Multifunctional Radars with Continuous Partial-modulated Radiation*. Moscow, Radiotekhnika, 2020, 280 p. (In Russ.).

13. Le V. K., Kozlov S. V. *Algorithms for long-term coherent accumulation of the reflected signal with non-zero higher derivatives of the distance to the radar target in the spectral region*. Doklady Belorusskogo gosudarstvennogo universiteta informatiki i radioelektroniki [Reports of the Belarusian State University of Informatics and Radioelectronics], 2021, no. 5, pp. 35–44 <http://dx.doi.org/10.35596/1729-7648-2021-19-5-35-44> (In Russ.).

14. Kuan N. V. *Passive radar system for monitoring the movement of ships in coastal areas using satellite illumination signals*. Izvestija vysshih uchebnyh zavedenij Rossii. Radioelektronika [News of higher educational institutions in Russia, Radioelectronics], 2020, vol. 23, no. 3, pp. 41–52. <https://doi.org/10.32603/1993-8985-2020-23-3-41-52> (In Russ.).

15. Vierinen J., Norberg J., Lehtinen M. S., Amm O., Roininen L., Väänänen A., Erickson P. J., McKay-Bukowski D. Beacon satellite receiver for ionospheric tomography. *Radio Science*, 2014, vol. 49, iss. 12, pp. 1141–1152. <https://doi.org/10.1002/2014RS005434>

16. Davies K. *Ionospheric Radio Waves*. Blaisdell Publishing Company, 1969, 460 p.

17. Tikhonov V. I. Optimal'nyj priem signalov. *Optimum Signal Reception*. Moscow, Radio i svjaz', 1983, 320 p. (In Russ.).

18. Genike A. A., Pobedinsky G. G. Global'nye sputnikovyje sistemy opredelenija mestopolozhenija i ih primenenie v geodezii. *Global Satellite Positioning Systems and Their Application in Geodesy*. Moscow, Kartgeotsentr, 2004, 355 p. (In Russ.).

Информация об авторах

Белоконов Игорь Витальевич, доктор технических наук, профессор, заведующий межвузовской кафедрой космических исследований, Самарский национальный исследовательский университет имени академика С. П. Королёва.
E-mail: priem@ssau.ru

Крот Александр Михайлович, доктор технических наук, профессор, заведующий лабораторией моделирования самоорганизующихся систем, Объединенный институт проблем информатики Национальной академии наук Беларуси.
E-mail: alxkrot@newman.bas-net.by
<https://orcid.org/0000-0003-0705-010X>

Козлов Сергей Вячеславович, доктор технических наук, профессор, профессор кафедры информационных радиотехнологий, Белорусский государственный университет информатики и радиоэлектроники.
E-mail: kozlov@bsuir.by

Каплярчук Евгений Александрович, аспирант кафедры информационных радиотехнологий, Белорусский государственный университет информатики и радиоэлектроники.
E-mail: ek.genia13@gmail.com

Савиных Ирина Эдуардовна, инженер-программист, лаборатория моделирования самоорганизующихся систем, Объединенный институт проблем информатики Национальной академии наук Беларуси.
E-mail: rct.savinykh@gmail.com

Шапкин Александр Сергеевич, аспирант, лаборатория моделирования самоорганизующихся систем, Объединенный институт проблем информатики Национальной академии наук Беларуси.
E-mail: shap1kin2@gmail.com

Information about the authors

Igor V. Belokonov, D. Sc. (Eng.), Prof., Head of the Interuniversity Department of Space Research, Samara National Research University named after academician S. P. Korolev.
E-mail: priem@ssau.ru

Alexander M. Krot, D. Sc. (Eng.), Prof., Head of the Laboratory of Self-organization System Modeling, The United Institute of Informatics Problems of the National Academy of Sciences of Belarus.
E-mail: alxkrot@newman.basnet.by
<https://orcid.org/0000-0003-0705-010X>

Sergey V. Kozlov, D. Sc. (Eng.), Prof., Prof. at the Information Radioengineering Department, Belarusian State University of Informatics and Radioelectronics.
E-mail: kozlov@bsuir.by

Yauheni A. Kapliarchuk, Postgraduate Student at the Information Radioengineering Department, Belarusian State University of Informatics and Radioelectronics.
E-mail: ek.genia13@gmail.com

Irina E. Savinykh, Software Engineer, Laboratory of Self-organization System Modeling, The United Institute of Informatics Problems of the National Academy of Sciences of Belarus.
E-mail: rct.savinykh@gmail.com

Aliaksandr S. Shapkin, Postgraduate Student, Laboratory of Self-organization System Modeling, The United Institute of Informatics Problems of the National Academy of Sciences of Belarus.
E-mail: shap1kin2@gmail.com

ОБРАБОТКА СИГНАЛОВ, ИЗОБРАЖЕНИЙ, РЕЧИ, ТЕКСТА И РАСПОЗНАВАНИЕ ОБРАЗОВ

SIGNAL, IMAGE, SPEECH, TEXT PROCESSING AND PATTERN RECOGNITION



УДК 004.93'11
<https://doi.org/10.37661/1816-0301-2023-20-2-28-38>

Оригинальная статья
Original Paper

Методика формирования базы гистопатологических изображений папиллярного рака щитовидной железы для глубокого обучения

М. В. Фридман^{1,2}, А. А. Косарева^{3✉}, Э. В. Снежко⁴, П. В. Камлач³, В. А. Ковалев⁴

¹Минский городской клинический
онкологический центр,
пр. Независимости, 64, 220013, Минск, Беларусь

²Республиканский центр опухолей щитовидной железы,
пр. Независимости, 64, 220013, Минск, Беларусь

³Белорусский государственный университет
информатики и радиоэлектроники,
ул. П. Бровки, 6, 220013, Минск, Беларусь

✉E-mail: kosareva@bsuir.by

⁴Объединенный институт проблем информатики
Национальной академии наук Беларуси,
ул. Сурганова, 6, Минск, 220012, Беларусь

Аннотация

Цели. Морфологический анализ образцов папиллярного рака щитовидной железы важен для дальнейшего планирования лечения. Автоматизация этого процесса связана с использованием традиционных и нейросетевых способов извлечения признаков изображений. Подготовительная работа предполагает подготовку набора данных, содержащих изображения, которые применяются для обучения нейронных сетей. В статье рассматривается вопрос выбора признаков для разметки гистологических изображений, особенности работы с полнослайдовыми изображениями, предлагается методика подготовки данных для разработки системы поиска похожих фрагментов гистопатологического изображения рака щитовидной железы. Исследуется влияние размера представительного фрагмента полнослайдового изображения папиллярного рака щитовидной железы на точность классификации обученной нейронной сети Efficient-NetV0, оцениваются слабые стороны использования фрагментов изображений разного представительного размера и причины неудовлетворительной точности классификации на больших увеличениях.

Материалы и методы. Использовались гистопатологические полнослайдовые изображения 129 пациентов. Гистологические микропрепараты, содержащие элементы опухоли и окружающей ткани, сканировались в аппарате Aperio AT2 (Leica Biosystems, Германия) с максимальным разрешением. Разметка про-

водилась в программном пакете ASAP. Для выбора оптимального представительного размера фрагмента решалась задача классификации с применением предобученной нейронной сети EfficientNetB0.

Результаты. Предложена методика подготовки базы данных гистопатологических изображений папиллярного рака щитовидной железы, проведены эксперименты по определению оптимального представительного размера фрагмента изображения. Наилучший результат точности определения класса тестовой выборки показал размер представительного фрагмента 394,32×394,32 мкм.

Заключение. Анализ влияния представительных размеров фрагментов гистопатологических изображений выявил проблемные места при решении задачи классификации, обусловленные спецификой нарезки и окрашивания изображений, морфологической сложностью и текстурным различием изображений одного класса. Поскольку проблема подготовки набора данных для обучения нейронной сети на решение задачи поиска инвазии сосудов на гистопатологическом изображении является нетривиальной, требуются дополнительные этапы подготовки данных.

Ключевые слова: медицинская визуализация, сверточная нейронная сеть, глубокое обучение, компьютерная диагностика, папиллярный рак щитовидной железы, архитектура рака

Для цитирования. Методика формирования базы гистопатологических изображений папиллярного рака щитовидной железы для глубокого обучения / М. В. Фридман [и др.] // Информатика. – 2023. – Т. 20, № 2. – С. 28–38. <https://doi.org/10.37661/1816-0301-2023-20-2-28-38>

Конфликт интересов. Авторы заявляют об отсутствии конфликта интересов.

Поступила в редакцию | Received 17.02.2023

Подписана в печать | Accepted 10.03.2023

Опубликована | Published 29.06.2023

Papillary thyroid carcinoma whole-slide images as a basis for deep learning

Michael V. Fridman^{1,2}, Aleksandra A. Kosareva^{3✉}, Eduard V. Snezhko⁴, Pavel V. Kamlach³, Vassili A. Kovalev⁴

¹*Minsk City Clinical Oncology Center,
av. Nezavisimosti, 64, Minsk, 220013, Belarus*

²*Republican Centre for Thyroid Tumours,
av. Nezavisimosti, 64, Minsk, 220013, Belarus*

³*Belarusian State University of Informatics and Radioelectronics,
st. P. Brovki, 6, Minsk, 220012, Belarus*

✉*E-mail: kosareva@bsuir.by*

⁴*The United Institute of Informatics Problems
of the National Academy of Sciences of Belarus,
st. Surganova, 6, Minsk, 220012, Belarus*

Abstract

Objectives. Morphological analysis of papillary thyroid cancer is a cornerstone for further treatment planning. Traditional and neural network methods of extracting parts of images are used to automate the analysis. It is necessary to prepare a set of data for teaching neural networks to develop a system of similar anatomical region in the histopathological image. Authors discuss the second selection of signs for the marking of histological images, methodological approaches to dissect whole-slide images, how to prepare raw data for a future analysis. The influence of the representative size of the fragment of the full-to-suction image of papillary thyroid cancer on the accuracy of the classification of trained neural network EfficientNetB0 is conducted. The analysis of the resulting results is carried out, the weaknesses of the use of fragments of images of different representative size and the cause of the unsatisfactory accuracy of the classification on large increase are evaluated.

Materials and methods. Histopathological whole-slide images of 129 patients were used. Histological micropreparations containing elements of a tumor and surrounding tissue were scanned in the Aperio AT2 (Leica Biosystems, Germany) apparatus with maximum resolution. The marking was carried out in the ASAP software package. To choose the optimal representative size of the fragment the problem of classification was solved using the pre-study neural network EfficientNetB0.

Results. A methodology for preparing a database of histopathological images of papillary thyroid cancer was proposed. Experiments were conducted to determine the optimal representative size of the image fragment. The best result of the accuracy of determining the class of test sample showed the size of a representative fragment as 394.32×394.32 microns.

Conclusion. The analysis of the influence of the representative sizes of fragments of histopathological images showed the problems in solving the classification tasks because of cutting and staining images specifics, morphological complex and textured differences in the images of the same class. At the same time, it was determined that the task of preparing a set of data for training neural network to solve the problem of finding invasion of vessels in a histopathological image is not trivial and it requires additional stages of data preparation.

Keywords: medical imaging, convolutional neural network, deep learning, computer-aided diagnosis, papillary thyroid cancer, cancer architectonics

For citation. Fridman M. V., Kosareva A. A., Snezhko E. V., Kamlach P. V., Kovalev V. A. *Papillary thyroid carcinoma whole-slide images as a basis for deep learning*. *Informatika [Informatics]*, 2023, vol. 20, no. 2, pp. 28–38 (In Russ.). <https://doi.org/10.37661/1816-0301-2023-20-2-28-38>

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

Введение. Подавляющее большинство злокачественных новообразований щитовидной железы диагностируются как папиллярный рак [1]. Как правило, заболевание не угрожает жизни пациента, однако нередко встречаются и варианты этой опухоли с неблагоприятным прогнозом [2, 3]. Поэтому морфологический анализ и сопоставления между клиническим поведением и морфологическими характеристиками новообразования важны для дальнейшего планирования лечения [4], установления степени распространения опухолевого процесса, оценки агрессивного потенциала карциномы. Процесс «чтения» гистологического микропрепарата может занимать много времени, а большое количество характеристик, требующих оценки, не позволяет проводить полноценную диагностику в рутинном режиме, что существенно затрудняет прогнозирование. В настоящее время развитие технологий визуализации позволяет автоматизировать определенные этапы интеллектуального анализа микропрепаратов, обнаруживать детали, известные экспертам, но ускользающие от внимания практических врачей, облегчать трудоемкие и затратные по времени процедуры. В связи с этим целесообразно проанализировать существующие методы автоматического определения морфологических деталей на полнослайдовых изображениях с гистотопографических срезов папиллярного рака щитовидной железы и разработать систему, которая позволит количественно оценивать размер опухоли, особенности архитектоники, инвазивные свойства и другие маркеры, необходимые для диагностики и определения прогностически значимых характеристик не только экспертом, но и практикующим врачом.

Основная проблема разметки – корректное выделение морфологических структур на гистопатологических изображениях, по которым выполняется окончательное заключение по диагнозу. Процедура разметки может быть качественно выполнена только высококвалифицированным специалистом-гистопатологом с соответствующим базовым образованием и большим практическим опытом. В этом отношении главной задачей компьютеризированного анализа гистопатологических изображений является автоматизация процесса выделения указанных объектов и морфологических структур изображения. В настоящее время основным способом решения данной задачи является использование методов глубокого обучения и сверточных нейронных сетей.

Цель исследования заключалась в разработке методики подготовки фрагментов полнослайдовых гистопатологических изображений папиллярного рака щитовидной железы для дальнейшего использования их при разработке интеллектуальных систем.

Методы исследования гистопатологических изображений. Существующие методы автоматического исследования гистопатологических изображений можно разделить на традиционные и нейросетевые.

К традиционным инструментам анализа изображений можно отнести метод локальных бинарных шаблонов (local binary patterns, LBP), масштабно-инвариантную трансформацию признаков (scale-invariant feature transform, SIFT), расчет матриц встречаемости (gray-level co-occurrence matrices, GLCM). Все они нашли применение в рамках решения задачи поиска схожей анатомической области изображений при работе с компьютерными томограммами легких. При выделении текстурных признаков изображения наилучший результат показали методы GLCM и LBP [5]. Метод LBP нашел применение и при анализе полнослайдовых изображений, полученных по результатам исследования биоптатов почечной ткани [6]. Также совмещение традиционного и нейросетевого подходов позволило авторам снизить число ложных срабатываний алгоритма до 3 % [6]. Анализ методом GLCM молекулярных изменений опухолевой ткани [7] обеспечил получение количественных характеристик.

Для улучшения качества поиска схожей области нередко совместно с традиционными используются методы машинного обучения. В исследованиях [5–7] применялись метод опорных векторов (support vector machines, SVM) и метод k -ближайших соседей (k -nearest neighbors algorithm, k -NN). Еще один подход базируется на совместном использовании SVM, k -NN, правила ближайшего соответствия (closest matching rule, CMR) и байесовского алгоритма (naive Bayes algorithm, NB) [8].

Нейросетевой анализ гистопатологического изображения щитовидной железы обусловлен применением классификаторов, основанных на использовании архитектур нейронных сетей VGG, Inceptionv3 и семейства ResNet [9], причем наилучший результат показала нейронная сеть ResNet50. Извлеченные нейросетевые признаки оценивались по четырем классификаторам и трем методам хеширования. Применялся полуавтоматический подход, характеризующийся сложностью обработки полнослайдовых изображений: поступающее на вход нейронной сети изображение предварительно размечалось специалистом, определяющим область интереса.

Существенная проблема работы с полнослайдовым изображением возникает из-за его размера. Одна картинка может достигать $100\,000 \times 300\,000$ пикселей, так как содержит в себе информацию о последовательно отсканированном гистопатологическом срезе на разных увеличениях.

Таким образом, подготовка набора данных крайне важна для дальнейшей разработки системы анализа гистопатологических изображений, причем результаты применения методов глубокого обучения чувствительны к размеру, сбалансированности и особенно к содержанию обучающей выборки. Выбор размера фрагментов полнослайдового изображения также оказывает сильное влияние на интерпретацию результатов анализа в связи с тем, что при разном увеличении распознаются разные признаки гистопатологического изображения.

Материалы для подготовки набора данных гистопатологических изображений. Набор данных был составлен на основе гистопатологических полнослайдовых изображений 129 пациентов. Все пациенты были разделены на две группы по времени рождения: до аварии на Чернобыльской АЭС (104 пациента, из них 35 мужчин и 69 женщин) и после (25 пациентов, из них 14 мужчин и 11 женщин).

Материалом для исследования послужили фрагменты папиллярной карциномы щитовидной железы. Был выполнен микроскопический анализ с использованием стандартной окраски гематоксилином и эозином. Гистологические микропрепараты, содержащие элементы опухоли и окружающей ткани, сканировались в аппарате Aperio AT2 (Leica Biosystems, Германия) с максимальным разрешением.

Виртуальные полнослайдовые изображения были описаны медицинским специалистом (М. В. Фридманом) с выделением таких характеристик, как размер опухоли, вариант гистологического строения, архитектура, одиночность или множественность фокусов роста опухоли, локализация, наличие лимфоидных инфильтратов, инвазия лимфатических и кровеносных сосудов, присутствие слоистых микрокальцинатов (псамомных телец), фиброза и фоновой пато-

логии. Были выделены семь паттернов (классов), соответствующих солидной, фолликулярной, папиллярной, папиллярно-фолликулярной, фолликулярно-солидной, папиллярно-солидной и смешанной архитектонике (рис. 1). В отдельный класс были определены фрагменты, содержащие псамомные тельца (рис. 1, *e*). Их наличие может свидетельствовать о давности патологического процесса, а расположение в межфолликулярном пространстве – о лимфогенной внутрижелезистой диссеминации по лимфатическим каналам.

Фрагменты изображений, содержащие фиброзную ткань (рис. 1, *f*) и области фоновой патологии (аутоиммунный тиреоидит, зоб и аденома, рис. 1, *g*), также были выделены в классы. Выделение фрагментов фоновой патологии необходимо для реализации автоматической классификации злокачественных и доброкачественных опухолей. В отдельный класс были выделены и участки неизменной ткани (рис. 1, *h*).

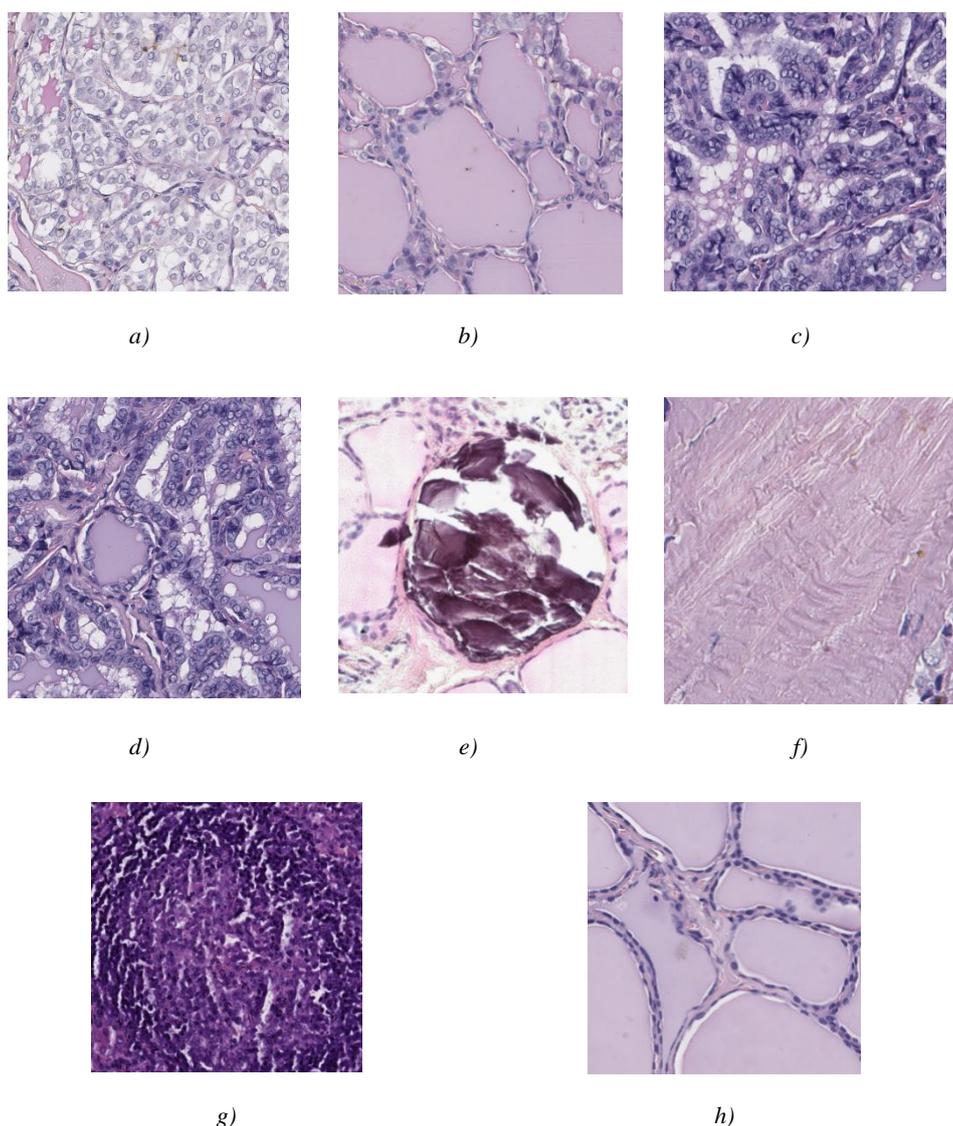


Рис. 1. Фрагменты микропрепаратов, содержащие: *a*) солидную; *b*) фолликулярную; *c*) папиллярную; *d*) папиллярно-фолликулярную архитектонику; *e*) псамомное тельце; *f*) фиброзную ткань; *g*) фоновую патологию; *h*) неизменную ткань

Fig. 1. Tiles of micropreparations containing: a) solid; b) follicular; c) papillary; d) papillary and follicular architectonics; e) psammoma body; f) fibrous tissue; g) background pathology; h) unchanged tissue

Первоначально в отдельный класс для разметки входили области, содержащие инвазию лимфатических и кровеносных сосудов. Однако в ходе разметки выбранного класса оказалось, что размеры (калибр и толщина стенки) сосудов могут значительно отличаться между собой. Это сильно влияло на вариативность размеров фрагментов изображений при подготовке набора данных и отражалось на размерах скользящего окна при поиске схожих фрагментов среди всего изучаемого поля. Поскольку решение задачи поиска инвазии сосудов на изображении требовало отдельного исследования, фрагменты изображения с инвазией сосудов были исключены из общего набора данных.

Методика подготовки данных для обучения нейронных сетей. Для разметки изображений на выбранные классы использовалось программное обеспечение ASAP [10]. Изображения всех классов размечались с помощью полигональных масок за исключением класса, содержащего изображения псамомных телец. Псамомные тельца размечались точно, и изображения нарезались так, чтобы псамомное тельце располагалось в центре фрагмента согласно координатам проставленных точек. В ходе разметки каждая маска подписывалась названием соответствующего класса, что позволило в дальнейшем автоматизировать процесс нарезки фрагментов. После разметки все изображения были разрезаны на фрагменты, содержащие признаки выделенных классов.

Была реализована программа, позволяющая выделять фрагменты из подготовленных изображений в соответствии с размеченными классами, выбирать размер фрагмента, а также долю выделенной маски в фрагменте. В результате нарезаемые по краям маски-полигона фрагменты содержали большую часть изображения-признака.

Также было проведено исследование по выбору представительного размера фрагмента для последующего анализа. Для этого были подготовлены четыре набора данных для обучения нейронных сетей с разным масштабом отображения (49,29; 98,58; 197,16 и 394,32 мкм по одной стороне фрагмента). Каждый из исследуемых масштабов выбирался в соответствии с уровнем увеличения гистопатологического изображения. Масштаб 788,64 мкм по одной стороне фрагмента не рассматривался, так как при таком способе выделения фрагментов их количество было недостаточным для обучения нейронной сети.

Подготовленные наборы включали в себя два класса данных: изображения нормальной ткани и изображения с признаками наличия раковой опухоли. Каждый класс содержал 2000 фрагментов гистопатологических изображений, полученных из 43 полнослайдовых изображений 27 пациентов. Тестовая выборка представляла собой фрагменты гистопатологических изображений других 10 пациентов (12 полнослайдовых изображений).

Для решения задачи использовалась предобученная нейронная сеть EfficientNetB0 [11].

Общая блок-схема предлагаемой методики подготовки базы гистопатологических изображений папиллярного рака щитовидной железы изображена на рис. 2.

Результаты экспериментальных исследований. Были проведены эксперименты по определению оптимального представительного размера фрагмента изображения. Результаты тестирования обученных на разных наборах данных нейросетевых моделей представлены в таблице. Точность определения класса оценивалась с помощью метрики F_1 -мера [12, 13]:

$$F_1 = \frac{2(P \cdot R)}{P + R},$$

где P – доля верно отнесенных к классу «фрагменты с признаками наличия раковой опухоли» изображений среди всех изображений, отнесенных к данному классу; R – доля верно отнесенных к классу «фрагменты с признаками наличия раковой опухоли» изображений среди всех изображений данного класса.

Наилучший результат точности определения класса тестовой выборки показал размер представительного фрагмента $394,32 \times 394,32$ мкм. При этом размер представительного фрагмента $197,16 \times 197,16$ мкм показал лучший результат при определении класса «фрагменты с признаками наличия раковой опухоли», но худший – при соотношении к верному классу «фрагменты изображений нормальной ткани».

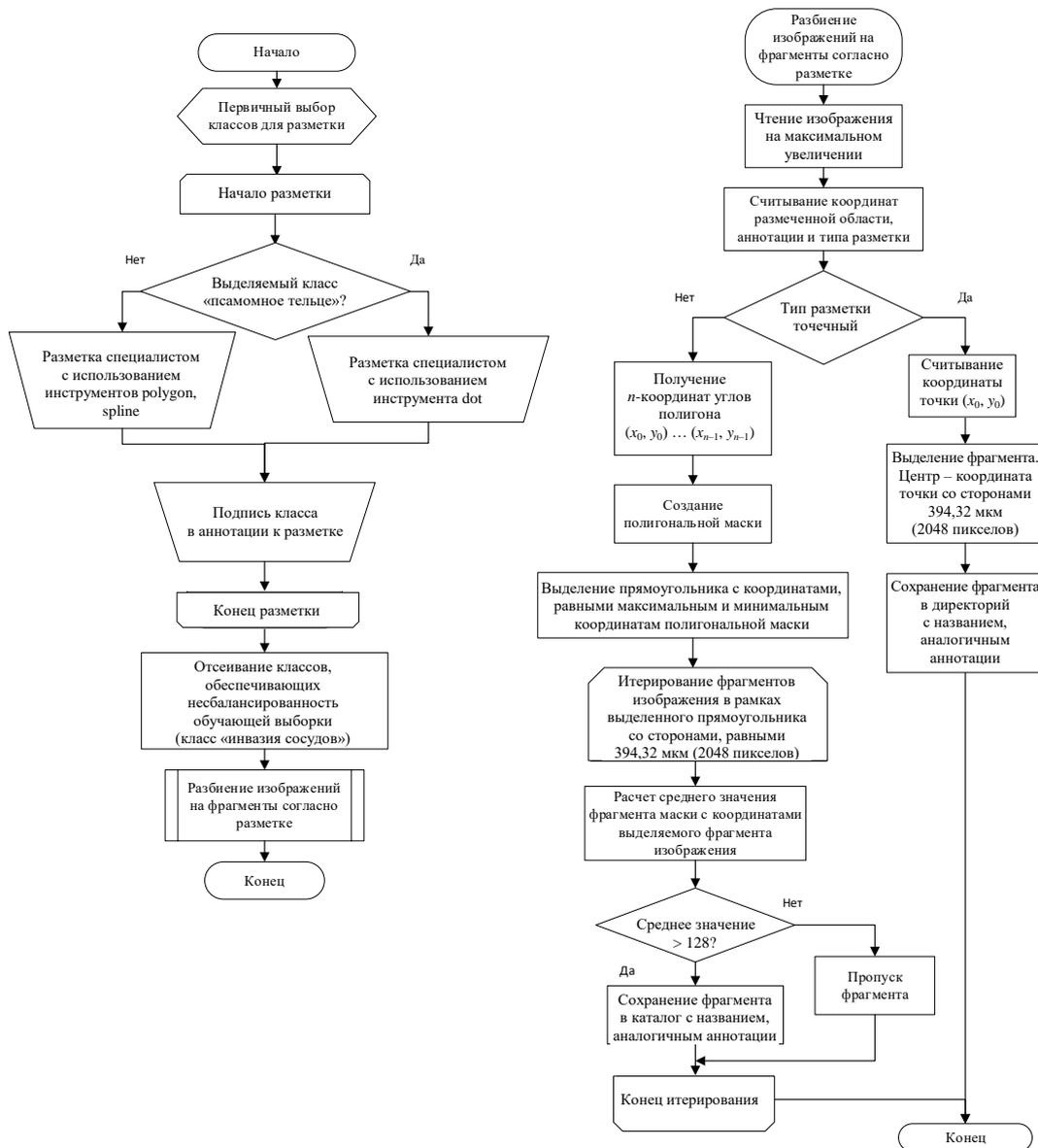


Рис. 2. Блок-схема предлагаемой методики формирования базы гистопатологических изображений
Fig. 2. The block diagram of the proposed methodology for the formation of the base of histopathological images

Результаты тестирования нейронной сети EfficientNetB0, обученной на фрагментах разного представительного размера

The results of testing the neural network EfficientNetB0, trained on tiles of different representative size

Размер представительного фрагмента при обучении EfficientNetB0, мкм×мкм <i>Size of a representative fragment when training EfficientNetB0, $\mu\text{m} \times \mu\text{m}$</i>	Мера F_1 , рассчитанная по результатам классификации изображений тестового набора данных <i>Score F_1 calculated according to the results of the classification of tiles of the test dataset</i>
49,29×49,29	0,66
98,58×98,58	0,68
197,16×197,16	0,70
394,32×394,32	0,83

Модели нейронных сетей, обученные на фрагментах с представительными размерами $49,29 \times 49,29$ и $98,58 \times 98,58$ мкм, показали худший результат. На возможном наименьшем по размеру представительном фрагменте нейронная сеть не обучилась, а на в два раза большем фрагменте показала неудовлетворительную точность как для нормальной ткани, так и для изображений, содержащих признаки раковой опухоли.

Есть несколько причин и проблем, влияющих на результат проведенного эксперимента. Результаты классификации при обучении нейронной сети на изображениях с малым размером представительного фрагмента ($49,29 \times 49,29$ и $98,58 \times 98,58$ мкм) объясняются тем, что при составлении обучающей выборки в оба класса попадали изображения с плохо отличимыми признаками. На рис. 3 представлены примеры схожих изображений разных классов в представительном размере фрагмента $98,58 \times 98,58$ мкм. Как видно из рисунка, большую часть фрагмента занимает коллоид, поэтому отнесение нейросетью данных изображений к одному из двух классов является затруднительным. Одинаковый размер фолликулов в (и вне) опухоли усложняет дифференциальную диагностику и при рутинном исследовании, и при применении нейросети.

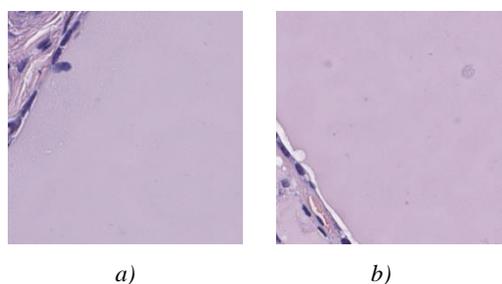


Рис. 3. Фрагменты гистопатологического изображения с представительным размером $98,58 \times 98,58$ мкм, содержащие примеры: a) нормальной ткани; b) папиллярного рака фолликулярной архитектоники

Fig. 3. Tiles of a histopathological image with a present size $98.58 \times 98.58 \mu\text{m}$ containing examples of: a) normal tissue; b) papillary cancer of follicular architectonics

Помимо информации, содержащейся в фрагментах изображений разных размеров, можно выделить еще несколько проблем, влияющих на неточность классификации. На рис. 4, a показан неверно отнесенный к классу нормальной ткани фрагмент внутриопухолевого фиброза, что объясняется отсутствием специфичности фиброзной ткани, на рис. 4, c – фрагмент нормальной ткани, включающий в себя изображение множества прозрачных пузырьков воздуха, которые покрывают поверхность микропрепарата.

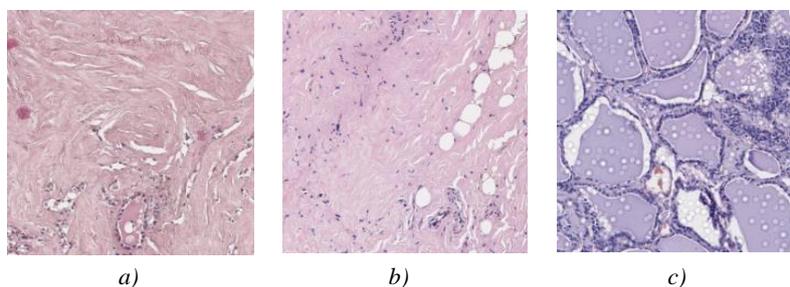


Рис. 4. Фрагменты гистопатологического изображения с представительным размером $394,32 \times 394,32$ мкм, содержащие примеры: a) неверно классифицированной фиброзной ткани; b) зоны из анатомической капсулы щитовидной железы; c) неизменной тиреоидной ткани с артефактами покрытия микропрепарата стеклом

Fig. 4. Tiles of the histopathological image with a representing size of $394.32 \times 394.32 \mu\text{m}$, containing examples of: a) an incorrectly classic fibrous tissue; b) zones from anatomical capsule of thyroid gland; c) unchanged thyroid tissue with artifacts of the coating of glass

Заключение. В результате исследований была разработана методика подготовки фрагментов гистопатологических изображений, включающая выбор классов и способа разметки класса, подписи масок, а также программа для нарезки фрагментов, позволяющая выбрать размер фрагмента и долю наличия изображения-признака на фрагменте после нарезки.

Предложен оптимальный размер представительного фрагмента – 394,32×394,32 мкм. Критерием выбора стала F_1 -мера, рассчитанная при классификации фрагментов гистопатологических изображений на тестовом наборе данных. Анализ влияния представительных размеров фрагментов гистопатологических изображений показал проблемные места при решении задачи классификации, обусловленные спецификой нарезки и окрашивания изображений, морфологической сложностью и текстурным различием изображений одного класса.

Также было определено, что проблема подготовки набора данных для обучения нейронной сети на решение задачи поиска инвазии сосудов на гистопатологическом изображении является нетривиальной и поэтому требует дополнительных этапов подготовки данных.

Подготовленный набор данных можно использовать для разработки системы второго мнения, направленной на поиск злокачественных и доброкачественных новообразований, на определение архитектоники папиллярного рака, на поиск псамомных телец и фиброзной ткани.

В дальнейшем планируется проведение работ по апробации традиционных и нейросетевых методов извлечения признаков изображений для реализации системы поиска похожих фрагментов гистопатологических изображений.

Вклад авторов. *М. В. Фридман* выполнил экспертную разметку гистопатологических изображений. *А. А. Косарева* осуществила подготовку фрагментов гистопатологических изображений для обучения нейронных сетей, исследовала зависимости точности классификации фрагмента от его представительного размера, проанализировала результаты. *Э. В. Снежко* оказывал консультативную помощь при проведении экспериментов. *П. В. Камлач* выполнил постановку задачи и научное редактирование статьи. *В. А. Ковалев* определил направление и цель исследования, общую структуру эксперимента.

Список использованных источников

1. A national cancer data base report on 53,856 cases of thyroid carcinoma treated in the U.S., 1985–1995 / S. A. Hundahl [et al.] // *Cancer*. – 1998. – Vol. 83, iss. 12. – P. 2638–22948. [https://doi.org/10.1002/\(SICI\)1097-0142\(19981215\)83:12<2638::AID-CNCR31>3.0.CO;2-1](https://doi.org/10.1002/(SICI)1097-0142(19981215)83:12<2638::AID-CNCR31>3.0.CO;2-1)
2. Гистологическая и молекулярно-генетическая характеристика клинически агрессивных вариантов папиллярного рака щитовидной железы / А. В. Боголюбова [и др.] // *Архив патологии*. – 2019. – Т. 81, № 1. – С. 46–51. <https://doi.org/10.17116/patol20198101146>
3. Morphological and clinical presentation of papillary thyroid carcinoma in children and adolescents of Belarus: the influence of radiation exposure and the source of irradiation / M. Fridman [et al.] // *Experimental and Molecular Pathology*. – 2015. – Vol. 98, iss. 3. – P. 527–531. <https://doi.org/10.1016/j.yexmp.2015.03.039>
4. Xu, B. Crucial parameters in thyroid carcinoma reporting – challenges, controversies and clinical implications / B. Xu, R. A. Ghossein // *Histopathology*. – 2018. – Vol. 72, iss. 1. – P. 32–39. <https://doi.org/10.1111/his.13335>
5. Косарева, А. А. Определение схожего анатомического участка на КТ-изображении грудной клетки с использованием традиционных методов выделения признаков изображения / А. А. Косарева, П. В. Камлач, В. А. Ковалев // *Доклады БГУИР*. – 2022. – Т. 20, № 5. – С. 48–56. <http://dx.doi.org/10.35596/1729-7648-2022-20-5-48-56>
6. Multi-radial LBP features as a tool for rapid glomerular detection and assessment in whole slide histopathology images / O. Simon [et al.] // *Scientific Reports*. – 2018. – Vol. 8. – P. 1–11. <https://doi.org/10.1038/s41598-018-20453-7>
7. A novel method for morphological pleomorphism and heterogeneity quantitative measurement: Named cell feature level co-occurrence matrix / A. Saito [et al.] // *J. of Pathology Informatics*. – 2016. – Vol. 7, iss. 1. – P. 1–9. <https://doi.org/10.4103/2153-3539.189699>

8. Jothi, J. A. A. Automatic classification of thyroid histopathology images using multi-classifier system / J. A. A. Jothi, V. M. A. Rajam // *Multimedia Tools and Applications*. – 2017. – Vol. 76, iss. 18. – P. 18711–18730. <https://doi.org/10.1007/s11042-017-4363-0>
9. Interactive thyroid whole slide image diagnostic system using deep representation / P. Chen [et al.] // *Computer Methods and Programs in Biomedicine*. – 2020. – Vol. 195. – P. 105630. <https://doi.org/10.1016/j.cmpb.2020.105630>
10. Litjens, G. Automate Slide Analysis Platform (ASAP) [Electronic resource] / G. Litjens. – 2017. – Mode of access: <https://github.com/geertlitjens/ASAP>. – Date of access: 10.09.2022
11. Tan, M. EfficientNet: Rethinking model scaling for convolutional neural networks / M. Tan, Q. V. Le // *Proc. of the 36th Intern. Conf. on Machine Learning, ICML 2019, Long Beach, 9–15 June 2019*. – Long Beach, 2019. – P. 6105–6114.
12. Rijsbergen, C. J. van. *Information Retrieval* / C. J. van Rijsbergen. – 2nd ed. – Butterworth-Heinemann, 1979. – 224 p.
13. Sasaki, Y. The Truth of the F-measure [Electronic resource] / Y. Sasaki. – 2007. – Mode of access: <https://www.cs.odu.edu/~mukka/cs795sum09dm/Lecturenotes/Day3/F-measure-YS-26Oct07.pdf>. – Date of access: 17.02.2023.

References

1. Hundahl S. A., Fleming I. D., Fremgen A. M., Menck H. R. A national cancer data base report on 53,856 cases of thyroid carcinoma treated in the U.S., 1985–1995. *Cancer*, 1998, vol. 83, iss. 12, pp. 2638–22948. [https://doi.org/10.1002/\(SICI\)1097-0142\(19981215\)83:12<2638::AID-CNCR31>3.0.CO;2-1](https://doi.org/10.1002/(SICI)1097-0142(19981215)83:12<2638::AID-CNCR31>3.0.CO;2-1)
2. Bogolyubova A. V., Abrosimov A. Yu., Selivanova L. S., Belousov P. V. *Histopathological and molecular genetic characteristics of clinically aggressive variants of papillary thyroid carcinoma*. *Arkhiv patologii [Pathology Archive]*, 2019, vol. 81, no. 1, pp. 46–51 (In Russ.). <https://doi.org/10.17116/patol20198101146>
3. Fridman M., Lam A. K., Krasko O., Schmid K. W., Branovan D. I., Demidchik Y. Morphological and clinical presentation of papillary thyroid carcinoma in children and adolescents of Belarus: the influence of radiation exposure and the source of irradiation. *Experimental and Molecular Pathology*, 2015, vol. 98, iss. 3, pp. 527–531. <https://doi.org/10.1016/j.yexmp.2015.03.039>
4. Xu B., Ghossein R. A. Crucial parameters in thyroid carcinoma reporting – challenges, controversies and clinical implications. *Histopathology*, 2018, vol. 72, iss. 1, pp. 32–39. <https://doi.org/10.1111/his.13335>
5. Kosareva A. A., Kamlach P. V., Kovalev V. A. Determination of a similar anatomical area on a chest CT image using traditional image feature extraction methods. *Doklady Belorusskogo gosudarstvennogo universiteta informatiki i radioelektroniki [Reports of the Belarusian State University of Informatics and Radioelectronics]*, 2022, vol. 20, no. 5, pp. 48–56 (In Russ.). <http://dx.doi.org/10.35596/1729-7648-2022-20-5-48-56>
6. Simon O., Yacoub R., Jain S., Tomaszewski J. E., Sarder P. Multi-radial LBP features as a tool for rapid glomerular detection and assessment in whole slide histopathology images. *Scientific Reports*, 2018, vol. 8, pp. 1–11. <https://doi.org/10.1038/s41598-018-20453-7>
7. Saito A., Numata Y., Hamada T., Horisawa T., Cosatto E., ..., Yamamoto Y. A novel method for morphological pleomorphism and heterogeneity quantitative measurement: Named cell feature level co-occurrence matrix. *Journal of Pathology Informatics*, 2016, vol. 7, iss. 1, pp. 1–9. <https://doi.org/10.4103/2153-3539.1896998>
8. Jothi J. A. A., Rajam V. M. A. Automatic classification of thyroid histopathology images using multi-classifier system. *Multimedia Tools and Applications*, 2017, vol. 76, iss. 18, pp. 18711–18730. <https://doi.org/10.1007/s11042-017-4363-09>
9. Chen P., Shi X., Liang Y., Li Y., Yang L., Gader P. D. Interactive thyroid whole slide image diagnostic system using deep representation. *Computer Methods and Programs in Biomedicine*, 2020, vol. 195, p. 105630. <https://doi.org/10.1016/j.cmpb.2020.105630>
10. Litjens G. *Automate Slide Analysis Platform (ASAP)*, 2017. Available at: <https://github.com/geertlitjens/ASAP> (accessed 10.09.2022).
11. Tan M., Le Q. V. EfficientNet: Rethinking model scaling for convolutional neural networks. *Proceedings of the 36th International Conference on Machine Learning, ICML 2019, Long Beach, 9–15 June 2019*. Long Beach, 2019, pp. 6105–6114.
12. Rijsbergen C. J. van. *Information Retrieval*, 2nd edition. Butterworth-Heinemann, 1979, 224 p.
13. Sasaki Y. *The Truth of the F-measure*, 2007. Available at: <https://www.cs.odu.edu/~mukka/cs795sum09dm/Lecturenotes/Day3/F-measure-YS-26Oct07.pdf> (accessed 17.02.2023).

Информация об авторах

Фридман Михаил Валерьевич, доктор медицинских наук, заведующий патологоанатомической лабораторией учреждения здравоохранения «Минский городской клинический онкологический центр»; сотрудник Республиканского центра опухолей щитовидной железы.

E-mail: kupriyan@rambler.ru

Косарева Александра Андреевна, ассистент кафедры электронной техники и технологии, аспирант Белорусского государственного университета информатики и радиоэлектроники.

E-mail: kosareva@bsuir.by

Снежко Эдуард Витальевич, кандидат технических наук, заведующий лабораторией анализа биомедицинских изображений Объединенного института проблем информатики Национальной академии наук Беларуси.

Камлач Павел Викторович, кандидат технических наук, доцент, заместитель декана факультета компьютерного проектирования Белорусского государственного университета информатики и радиоэлектроники.

Ковалев Василий Алексеевич, кандидат технических наук, доцент кафедры биомедицинской информатики Белорусского государственного университета, ведущий научный сотрудник лаборатории анализа биомедицинских изображений Объединенного института проблем информатики Национальной академии наук Беларуси.

Information about the authors

Michael V. Fridman, D. Sc. (Med.), Head of the Pathoanatomical Laboratory of the Health Care Institution "Minsk City Clinical Oncology Center"; Employee of the Republican Centre for Thyroid Tumours.

E-mail: kupriyan@rambler.ru

Aleksandra A. Kosareva, Assistant Lecturer of Electronic Engineering and Technology Department, Postgraduate Student of the Belarusian State University of Informatics and Radioelectronics.

E-mail: kosareva@bsuir.by

Eduard V. Snezhko, Ph. D. (Eng.), Head of the Biomedical Image Analysis Department of The United Institute of Informatics Problems of the National Academy of Sciences of Belarus.

Pavel V. Kamlach, Ph. D. (Eng.), Assoc. Prof., Deputy Dean of the Faculty of Computer Design of the Belarusian State University of Informatics and Radioelectronics.

Vassili A. Kovalev, Ph. D. (Eng.), Assoc. Prof. of the Biomedical Informatics Department of the Belarusian State University, Leading Researcher of the Biomedical Image Analysis Department of The United Institute of Informatics Problems of the National Academy of Sciences of Belarus.

ЛОГИЧЕСКОЕ ПРОЕКТИРОВАНИЕ

LOGICAL DESIGN



УДК 519.714
<https://doi.org/10.37661/1816-0301-2023-20-2-39-64>

Оригинальная статья
Original Paper

Применение диаграмм решений не полностью определенных функций k -значной логики при синтезе логических схем

П. Н. Бибило

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

Аннотация

Цели. Рассматривается проблема схемной реализации не полностью определенных функций k -значной логики, заданных табличными представлениями. Изучается этап технологически независимой оптимизации. Целью этого этапа является получение по табличным представлениям не полностью определенных функций k -значной логики минимизированных представлений систем полностью определенных булевых функций, по которым выполняется технологическое отображение (technology mapping) – второй этап синтеза логических схем.

Методы. При синтезе логических схем на этапе технологически независимой оптимизации предлагается использовать доопределения многозначных диаграмм решений (Reduced Ordered Multi-valued Decision Diagrams, ROMDD), которые далее называются MDD, и доопределения бинарных диаграмм решений (Binary Decision Diagram, BDD), задающих не полностью определенные системы булевых функций. Доопределение MDD ориентировано на уменьшение числа вершин графа MDD, соответствующих кофакторам разложения Шеннона многозначной функции.

Результаты. Задача минимизации MDD сведена к решению задач минимальной раскраски неориентированных графов несовместимости кофакторов. Кодирование многозначных значений аргументов и значений функций k -значной логики двоичными кодами приводит к системам не полностью определенных булевых функций, которые также доопределяются с целью минимизации их многоуровневых BDD-представлений.

Заключение. Предложенный подход позволяет в два этапа провести доопределение частичных многозначных функций до полностью определенных булевых функций. На втором этапе используются известные и эффективные методы доопределения BDD, задающих системы не полностью определенных булевых функций. В результате такого двухэтапного подхода получаются минимизированные BDD-представления систем полностью определенных функций. По полностью определенным булевым функциям выполняется технологическое отображение в заданную библиотеку логических элементов, т. е. покрытие оптимизированных описаний систем булевых функций описаниями логических элементов.

Ключевые слова: не полностью определенные функции, k -значная логика, Multi-valued Decision Diagram (MDD), булевы функции, Binary Decision Diagram (BDD), разложение Шеннона, синтез логической схемы, VHDL, СБИС

Для цитирования. Бибилло, П. Н. Применение диаграмм решений не полностью определенных функций k -значной логики при синтезе логических схем / П. Н. Бибилло // Информатика. – 2023. – Т. 20, № 2. – С. 39–64. <https://doi.org/10.37661/1816-0301-2023-20-2-39-64>

Конфликт интересов. Автор заявляет об отсутствии конфликта интересов.

Поступила в редакцию | Received 06.02.2023

Подписана в печать | Accepted 20.03.2023

Опубликована | Published 29.06.2023

Application of decision diagrams of incompletely specified of k -valued logic functions in the synthesis of logical circuits

Petr N. Bibilo

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

Abstract

Objectives. The problem of circuit implementation of incompletely specified (partial) k -valued logic functions given by tabular representations is considered. The stage of technologically independent optimization is studied to obtain minimized representations of systems of completely specified Boolean functions from tabular representations of partial functions of k -valued logic. According to these representations of Boolean functions, technological mapping is performed at the second stage of the synthesis of logic circuits.

Methods. Using additional definitions of Multi-valued Decision Diagrams (MDD) representing partial functions of k -valued logic, and Binary Decision Diagrams (BDD) representing partial systems of Boolean functions at the stage of technologically independent optimization is proposed. The task of additional definition of MDD is oriented to reducing the number of vertices of the MDD graph that correspond to the cofactors of the Shannon expansion of a multi-valued function.

Results. The MDD minimization problem is reduced to solving the problems of coloring undirected graphs of incompatibility of cofactors by minimum number of colors. Encoding of multi-valued values of arguments and values of functions of k -valued logic by binary codes leads to systems of partial Boolean functions, which are also further defined in order to minimize their multi-level BDD representations.

Conclusion. The proposed approach makes it possible to define partial multi-valued functions to fully defined Boolean functions in two stages. At the second stage, well-known and effective methods are used to redefine BDD representing systems of partial Boolean functions. As a result of this two-step approach, minimized BDD representations of systems of completely defined functions are obtained. According to completely defined Boolean functions, a technological mapping into a given library of logical elements is performed, i.e. the optimized descriptions of Boolean function systems are covered with descriptions of logical elements.

Keywords: partial functions, k -valued logic, Multi-valued Decision Diagram (MDD), Boolean functions, Binary Decision Diagram (BDD), Shannon expansion, digital logic synthesis, VHDL, VLSI

For citation. Bibilo P. N. *Application of decision diagrams of incompletely specified of k -valued logic functions in the synthesis of logical circuits.* Informatika [Informatics], 2023, vol. 20, no. 2, pp. 39–64 (In Russ.). <https://doi.org/10.37661/1816-0301-2023-20-2-39-64>

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

Введение. Наиболее распространенной формой задания исходных спецификаций блоков комбинационной логики сверхбольших интегральных схем (СБИС) при их схемной реализации (синтезе логических схем) являются полностью либо не полностью определенные (частичные) булевы функции. На практике исходные спецификации могут задаваться также в виде компактных табличных заданий k -значных, $k \geq 3$, частичных функций, аргументы и значения которых определяются из ограниченного диапазона неотрицательных целых чисел. Примерами служат модулярные устройства [1–3]. Выполняя кодирование значений аргументов и значений функций двоичными кодами, можно получать системы булевых функций и сводить задачи схемной реализации функций k -значной логики к задачам схемной реализации систем булевых функций. Однако при таком подходе размерности задач предварительной логической оптимизации возрастают, особенно в тех случаях, когда приходится реализовывать системы k -значных (многозначных) функций. Кроме того, усложняется процесс замены неопределенных значений функций определенными значениями, выполняемый с целью улучшения основных характеристик логических схем, таких как площадь (число транзисторов), энергопотребление и быстродействие.

В настоящей работе для предварительной технологически независимой оптимизации систем не полностью определенных многозначных функций предлагается использовать графовый аппарат сокращенных упорядоченных многозначных диаграмм решений (Reduced Ordered Multivalued Decision Diagrams, ROMDD) [4], которые далее будем называть MDD. Это позволит находить целесообразные доопределения исходных многозначных функций, после чего можно будет переходить к частичным булевым (двухзначным) функциям, получающимся в результате кодирования двоичными кодами аналитических представлений, соответствующих оптимизированным MDD. Такой подход приводит к многоуровневым представлениям систем булевых функций в виде логических уравнений, соответствующих BDD, как важного и хорошо изученного [5–11] частного случая MDD для $k = 2$. Методы доопределения BDD представлений систем частичных булевых функций известны в литературе [11], поэтому общий процесс доопределения исходных многозначных функций может быть разбит на два этапа: вначале выполняется доопределение частичных многозначных функций, а затем – доопределение частичных булевых функций, полученных кодированием многозначных значений булевыми кодами. Минимизация числа частичных подфункций на уровнях MDD сведена к комбинаторным задачам раскраски вершин неориентированных графов несовместимости кофакторов в минимальное число цветов. При таком двухэтапном подходе к технологически независимой оптимизации происходит постепенное доопределение частичных многозначных функций до полностью определенных булевых функций, для которых выполняется технологическое отображение (technology mapping) минимизированных описаний полностью определенных булевых функций в заданную библиотеку логических элементов.

Основные определения. Пусть $S = \{0, 1, 2, \dots, k - 1\}$ и $S^n = S \times S \times \dots \times S$. Полностью определенной k -значной функцией, зависящей от n k -значных аргументов x_1, \dots, x_n , называется отображение

$$f : S^n \rightarrow S.$$

Далее для краткости будем называть аргументы таких функций и сами функции *многозначными*, отличая их от булевых (двухзначных) аргументов и функций. Аргументы x_i булевых функций, как и сами булевы функции f , принимают значения 0, 1. Таким образом, рассматриваемые полностью определенные многозначные функции f и их аргументы принимают значения из множества S . Если многозначная функция на некоторых наборах значений аргументов принимает неопределенное значение «–», то будем называть ее частичной (не полностью определенной) многозначной функцией.

Рассмотрим две частичные многозначные функции f^i, f^j . Частичная функция f^j реализует частичную функцию f^i , либо частичная функция f^i реализуется частичной функцией f^j ($f^i < f^j$

либо $f^j \succ f^i$), если на одном и том же наборе значений аргументов для каждой пары значений функций f^i, f^j выполняется одно (любое) из следующих отношений:

$$-\prec 0, -\prec 1, -\prec 2, \dots, -\prec k-1, \quad (1)$$

$$0\prec 0, 1\prec 1, 2\prec 2, \dots, k-1\prec k-1. \quad (2)$$

Функцию f^j будем называть доопределением функции f^i . Видно, что при доопределении частичной функции любое ее неопределенное значение « \prec » может быть заменено любым значением $0, 1, 2, \dots, k-1$ из множества S . Если для каждой пары соответствующих значений функций f^i, f^j выполняется одно из отношений, входящее в соотношения (2), то функции f^i, f^j являются полностью определенными и равными.

Таким образом, отношение равенства частичных функций представляет собой частный случай отношения реализации. Если для каждой пары соответствующих значений частичных булевых функций f^i, f^j выполняется одно (любое) из пяти отношений

$$0\prec 0, 1\prec 1, -\prec -, -\prec 0, -\prec 1,$$

то выполняется отношение реализации $f^i \prec f^j$ булевых функций. Упорядоченную систему булевых функций будем называть также векторной булевой функцией.

Пример 1. В табл. 1 приведена система $F = \{f^1, f^2, f^3, f^4\}$ частичных трехзначных функций, для которых выполняются отношения реализации

$$\begin{aligned} f^1 \prec f_*^1, f^2 \prec f_*^2, f^3 \prec f_*^3, f^4 = f_*^4, \\ f^1 \prec f_{**}^1, f^2 \prec f_{**}^2, f^3 \prec f_{**}^3, f^4 \prec f_{**}^4, \\ f_*^1 \prec f_{**}^1, f_*^2 \prec f_{**}^2, f_*^3 \prec f_{**}^3, f_*^4 \prec f_{**}^4. \end{aligned}$$

Таблица 1
Полностью определенные и частичные многозначные функции

Table 1
Fully defined and partial multi-valued functions

Пример 1 Example 1		Система F частичных многозначных функций System F of partial multi-valued functions				Реализующие частичные многозначные функции Implementing partial multi-valued functions				Реализующие полностью определенные многозначные функции Implementing fully defined multi-valued functions			
Многозначные аргументы Multi-valued argument		f^1	f^2	f^3	f^4	f_*^1	f_*^2	f_*^3	f_*^4	f_{**}^1	f_{**}^2	f_{**}^3	f_{**}^4
x_1	x_2												
0	0	-	-	1	2	1	-	1	2	1	2	1	2
0	1	1	-	-	1	1	1	-	1	1	1	2	1
0	2	2	-	1	2	2	0	1	2	2	0	1	2
1	0	1	2	0	2	1	2	0	2	1	2	0	2
1	1	-	0	-	1	-	0	1	1	1	0	1	1
1	2	1	1	0	2	1	1	0	2	1	1	0	2
2	0	2	2	2	2	2	2	2	2	2	2	2	2
2	1	0	0	-	0	0	0	2	0	0	0	2	0
2	2	1	1	-	1	1	1	-	1	1	1	2	1

Постановка задачи. На основе табличного задания системы F частичных многозначных функций требуется получить логическую схему, реализующую табличное задание системы F .

Предлагается выполнить следующие этапы проектирования требуемой логической схемы:

1. Построение и минимизация MDD, которая представляет функции системы F .

2. Кодирование значений многозначных аргументов и многозначных функций системы F булевыми кодами и переход от каждой многозначной функции системы F к своей системе частичных булевых функций.

3. Доопределение частичных булевых функций и получение функционального описания системы полностью определенных булевых функций в виде логических уравнений.

4. Синтез логической схемы в заданном технологическом базисе по полученным на этапе 3 логическим уравнениям.

Далее при решении задач, возникающих на этапах 1–4, будем придерживаться следующих соглашений (ограничений).

Соглашение 1. На этапе 1 будем строить MDD, доопределяя многозначные функции системы F в процессе построения MDD.

Заметим, что могут быть и другие подходы. Например, когда MDD построена, а затем осуществляется ее доопределение с целью сокращения сложности графа MDD.

Соглашение 2. В качестве кодов значений аргументов и значений функций будем использовать двоичные представления неотрицательных чисел из множества $S = \{0, 1, 2, \dots, k - 1\}$. При кодировании значений « \rightarrow » многозначной функции булевыми (двоичными) кодами неопределенное значение « \rightarrow » многозначной функции заменяется вектором неопределенных значений соответствующей системы *частичных* булевых функций. При кодировании могут быть использованы и другие коды (например, унарный код). Задача выбора эффективного способа кодирования представляет отдельный интерес и в данной работе не рассматривается.

Соглашение 3. Для системы частичных булевых функций будем оптимизировать BDD-представление и получать в результате выполнения этапа 3 логические уравнения, соответствующие BDD-представлению системы *полностью определенных* булевых функций.

Заметим, что от BDD-представления можно переходить к другим формам задания систем булевых функций. Например, можно дополнительно упрощать логические уравнения, соответствующие BDD-представлениям булевых функций, либо переходить к минимизированным дизъюнктивным нормальным формам (ДНФ) [12, 13], булевым сетям [14] и т. д. От формы задания реализуемой системы булевых функций зависят результаты синтеза в промышленных синтезаторах логических схем.

Соглашение 4. На этапе 4 будем использовать библиотеку логических элементов заказных КМОП СБИС, описанную в статье [15], и использовать в качестве системы синтеза логических схем синтезатор LeonardoSpectrum [16]. Синтез будет выполняться с целью получения однотактных (не конвейерных) реализаций схем. Синтезатор LeonardoSpectrum после синтеза схемы подсчитывает сложность (площадь) схем из библиотечных элементов как сумму площадей всех логических элементов схемы и выдает значение данного параметра под названием *Area* (площадь). Заметим, что задержка синтезированной схемы вычисляется в виде значения параметра *Delay*.

Построение и минимизация MDD, реализующей функции системы F (этап 1). Разложение Шеннона и MDD. Введем понятие литерала многозначной переменной x_i , $i = 1, \dots, n$. Если переменная x_i принимает значение $j \in S$, то пишем $x_i^{\{j\}}$. Выражение $x_i^{\{j\}}$ будем называть *одноэлементным литералом* переменной x_i . Пусть $T = \{j_1, \dots, j_t\}$, $T \subseteq S$. Под *многоэлементным литералом* (или просто литералом) будем понимать запись $x_i^{\{T\}}$ и интерпретировать литерал $x_i^{\{T\}}$ как булеву функцию: $x_i^{\{T\}} = 1$, если значение α переменной x_i принадлежит множеству T ; $x_i^{\{T\}} = 0$, если значение α переменной x_i не принадлежит множеству T . Если $T = S$, то литерал $x_i^{\{T\}} = 1$ (т. е. равен константе 1).

Результат подстановки значения $j \in S$ переменной x_i в частичную многозначную функцию $f(x_1, \dots, x_i, \dots, x_n)$ будем называть кофактором [4] многозначной функции $f(x_1, \dots, x_i, \dots, x_n)$ по литералу $x_i^{\{j\}}$ и запишем в виде

$$f_{x_i^{\{j\}}} = f(x_1, \dots, x_{i-1}, j, x_{i+1}, \dots, x_n).$$

Разложением Шеннона частичной многозначной функции $f(x_1, \dots, x_i, \dots, x_n)$ по переменной x_i называется представление

$$f = (x_i^{(0)} \& f_{x_i^{(0)}}) \vee (x_i^{(1)} \& f_{x_i^{(1)}}) \vee \dots \vee (x_i^{(k-1)} \& f_{x_i^{(k-1)}}), \quad (3)$$

где $\&$ – логическая конъюнкция, \vee – логическая дизъюнкция.

Путем нахождения одинаковых частичных кофакторов $f(x_1, \dots, x_{i-1}, j, x_{i+1}, \dots, x_n)$ можно сократить число дизъюнктивных членов в выражении (3) и получить множество попарно различающихся кофакторов для последующего разложения всех их по одной и той же переменной x_j . Выполнив такие процедуры разложения кофакторов по одной и той же переменной и найдя одинаковые частичные кофакторы, получим многоуровневое представление частичной многозначной функции. На последнем шаге разложения (или ранее) кофакторы вырождаются до констант 0, 1, 2, ..., $k - 1$. Многоуровневое представление можно задать бесконтурным ориентированным графом (орграфом), т. е. MDD. Кофакторы соответствуют вершинам орграфа MDD; ориентированные дуги, всегда идущие сверху вниз, задают факт порождения кофакторов нижнего уровня из кофакторов верхнего уровня. Пометки дуг соответствуют литералам, по которым порождается кофактор. Листовые вершины MDD соответствуют значениям многозначной функции.

В рассматриваемом случае наблюдается аналогия с построением сокращенных упорядоченных BDD для булевых функций [10, 11]. Построение MDD для системы F многозначных функций аналогично: одинаковые кофакторы находятся среди кофакторов всех функций системы F , разложение Шеннона для каждой функции системы F (либо ее кофакторов) ведется по одной и той же последовательности переменных разложения (для одной и той же перестановки переменных). В работе [4] изучались MDD для полностью определенных многозначных функций.

Рассмотрим пример 1 (см. табл. 1) построения MDD для системы F многозначных частичных функций $F = \{f^1, f^2, f^3, f^4\}$ путем получения кофакторов сначала по переменной x_1 (табл. 2 и 3), затем по переменной x_2 , т. е. перестановка переменных, по которой строится MDD, имеет вид $\langle x_1, x_2 \rangle$.

Таблица 2

Кофакторы p_1, \dots, p_5 разложения Шеннона функций системы F по переменной x_1

Table 2

Cofactors p_1, \dots, p_5 of the Shannon expansion of the functions of the system F by the variable x_1

x_2	$p_1 = f_{x_1^{(0)}}^1$	$p_2 = f_{x_1^{(2)}}^1 = f_{x_1^{(0)}}^3$	$p_3 = f_{x_1^{(2)}}^1$	$p_4 = f_{x_1^{(0)}}^2$	$p_5 = f_{x_1^{(2)}}^2 = f_{x_1^{(1)}}^2 = f_{x_1^{(2)}}^4$
0	–	1	2	–	2
1	1	–	0	–	0
2	2	1	1	–	1

Таблица 3

Кофакторы p_6, p_7, p_8 разложения Шеннона функций системы F по переменной x_1

Table 3

Cofactors p_6, p_7, p_8 of the Shannon expansion of the functions of the system F by the variable x_1

x_2	$p_6 = f_{x_1^{(1)}}^3$	$p_7 = f_{x_1^{(2)}}^3$	$p_8 = f_{x_1^{(0)}}^4 = f_{x_1^{(1)}}^4$
0	0	2	2
1	–	–	1
2	0	–	2

Используя обозначения кофакторов из табл. 2 и 3, допишем выражения разложения Шеннона многозначных функций f^1, f^2, f^3, f^4 (см. табл. 1) по переменной x_1 :

$$\begin{aligned} f^1 &= x_1^{(0)} \& p^1 \vee x_1^{(1)} \& p^2 \vee x_1^{(2)} \& p^3 = x_1^{(0)} p^1 \vee x_1^{(1)} p^2 \vee x_1^{(2)} p^3, \\ f^2 &= x_1^{(0)} \& p^4 \vee x_1^{(1)} \& p^3 \vee x_1^{(2)} \& p^5 = x_1^{(0)} p^4 \vee x_1^{(1)} p^3 \vee x_1^{(2)} p^5, \\ f^3 &= x_1^{(0)} \& p^2 \vee x_1^{(1)} \& p^6 \vee x_1^{(2)} \& p^7 = x_1^{(0)} p^2 \vee x_1^{(1)} p^6 \vee x_1^{(2)} p^7, \\ f^4 &= x_1^{(0,1)} \& p^8 \vee x_1^{(2)} \& p^5 = x_1^{(0,1)} p^8 \vee x_1^{(2)} p^5. \end{aligned}$$

В дальнейшем будем опускать знак $\&$ конъюнкции там, где это уместно. Разложение по переменной x_2 полученных кофакторов p_1, \dots, p_8 тривиально.

Граф MDD, построенный по перестановке $\langle x_1, x_2 \rangle$ переменных, изображен на рис. 1. Ориентация дуг не показана, так как все дуги ориентированы сверху вниз.

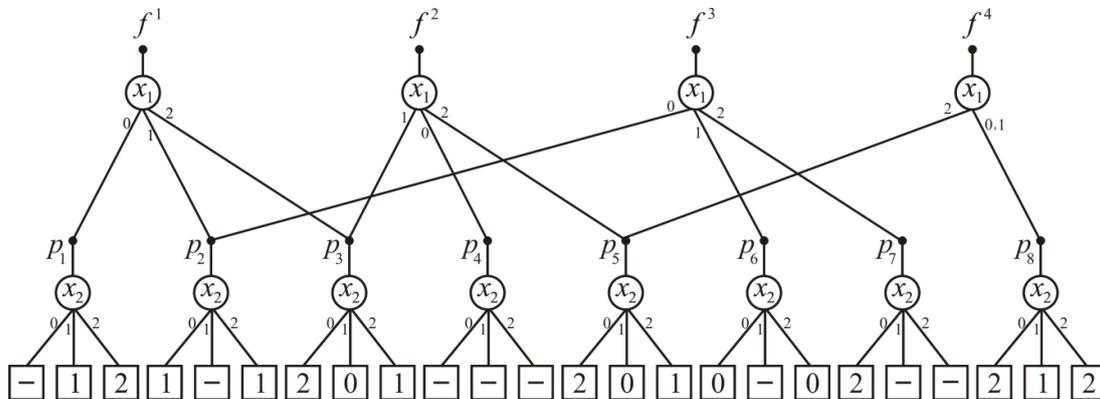


Рис. 1. MDD для системы функций F

Fig. 1. MDD for a system of functions F

Для упрощения изображения MDD листовые вершины для кофакторов, зависящих от переменной x_2 , продублированы. Неопределенные листовые вершины отличаются от определенных: при доопределении MDD каждая неопределенная листовая вершина рассматривается отдельно и доопределяется независимо от других неопределенных вершин. По MDD можно сформировать многозначные функции, рассматривая пути из корневых вершин, соответствующих функциям системы F , к листовым вершинам, задающим значения компонентной функции системы, и записывая конъюнкцию литералов, помечающих проходимые дуги, аналогично тому, как это делается для BDD [11]. Например, функция f^4 имеет представление

$$\begin{aligned} f^4 &= x_1^{(2)} x_2^{(0)} f^4(2,0) \vee x_1^{(2)} x_2^{(1)} f^4(2,1) \vee x_1^{(2)} x_2^{(2)} f^4(2,2) \vee \\ &\vee x_1^{(0,1)} x_2^{(0)} f^4(0,0) \vee x_1^{(0,1)} x_2^{(1)} f^4(0,1) \vee x_1^{(0,1)} x_2^{(2)} f^4(0,2). \end{aligned} \quad (4)$$

Так как $f^4(0,0) = f^4(1,0) = 2, f^4(0,1) = f^4(1,1) = 1, f^4(0,2) = f^4(1,2) = 2, f^4(2,1) = 0$, в выражении (4) вместо $f^4(0,0), f^4(0,1), f^4(0,2)$ могут быть записаны $f^4(1,0), f^4(1,1), f^4(1,2)$ соответственно.

Если понимать запись $f^{4(\alpha)}$ как «многозначная функция f^4 принимает значение $\alpha \in \{-\} \cup S$ », то формула (4) запишется в виде

$$f^4 = x_1^{(2)} x_2^{(0)} f^{4(2)} \vee x_1^{(2)} x_2^{(1)} f^{4(-)} \vee x_1^{(2)} x_2^{(2)} f^{4(1)} \vee x_1^{(0,1)} x_2^{(0)} f^{4(2)} \vee x_1^{(0,1)} x_2^{(1)} f^{4(1)} \vee x_1^{(0,1)} x_2^{(2)} f^{4(2)}. \quad (5)$$

В выражении (5) символы & (логическая конъюнкция) опущены, и (5) можно записать по табличному заданию f^4 (см. табл. 1). Аналогичным образом могут быть записаны представления других многозначных функций из табл. 1.

Доопределение кофакторов MDD. Введем понятие несовместимости многозначных функций либо их кофакторов. Кофакторы $f(x_1, \dots, x_r)$, $g(x_1, \dots, x_r)$ назовем *несовместимыми*, если найдется хотя бы один набор значений $(x_1^{(j_1)}, x_2^{(j_2)}, \dots, x_r^{(j_r)})$ переменных x_1, \dots, x_r , для которого оба значения $f(x_1^{(j_1)}, \dots, x_r^{(j_r)})$, $g(x_1^{(j_1)}, \dots, x_r^{(j_r)})$ определены и не равны. Если же такой набор $(x_1^{(j_1)}, x_2^{(j_2)}, \dots, x_r^{(j_r)})$ не найдется, то кофакторы $f(x_1, \dots, x_r)$, $g(x_1, \dots, x_r)$ назовем *совместимыми*.

Например, кофакторы p_1, p_2 (см. табл. 2), зависящие от многозначной переменной x_2 , несовместимы, поскольку найдется такое значение $x_2^{(2)}$, что $p_1(2) = 2, p_2(2) = 1$ (оба значения определены и не равны). Кофакторы p_4, p_5 являются совместимыми. Очевидно, что совместимые частичные кофакторы могут быть доопределены до одного и того же кофактора. Например, совместимые кофакторы p_4, p_5 могут быть доопределены до кофактора g , такого, что $g(0) = 2, g(1) = 1, g(2) = 1$, и для которого выполняются отношения реализации $p_4 \prec g, p_5 \prec g$. Кофакторы p_3, p_5 являются полностью определенными и равными.

Задача А. Для заданной системы F частичных многозначных функций требуется найти минимальную по мощности систему H многозначных функций, таких, что каждая из функций исходной системы F реализуется хотя бы одной из функций системы H .

Решение задачи А сводится к построению неориентированного графа G отношения несовместимости функций из системы F , раскраске графа G в минимальное число цветов (красок) и формированию по каждому множеству одноцветно раскрашенных вершин (функций) одной из функций системы H .

Минимальное число p_{\min} частичных функций в системе H определяется из соотношения

$$p_{\min} = \chi(G), \quad (6)$$

где $\chi(G)$ – хроматическое число графа G [17].

Рассмотрим решение задачи А на примере кофакторов p_1, \dots, p_8 из табл. 2 и 3. Граф G отношения несовместимости и раскраска его вершин показаны на рис. 2. Формирование функций h^1, h^2, h^3, h^4 , являющихся решением задачи А, представлено в табл. 4.

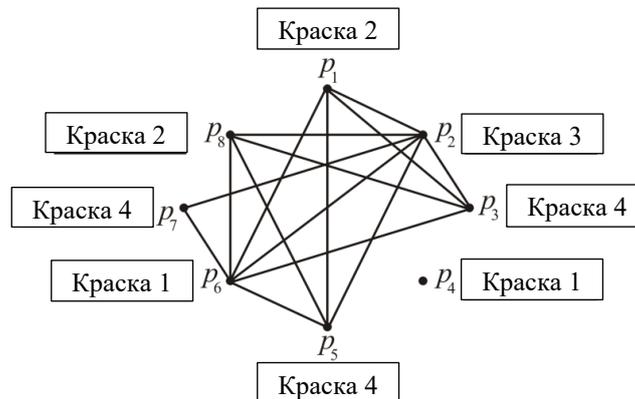


Рис. 2. Граф G отношения несовместимости кофакторов
Fig. 2. Graph G of the incompatibility relationship of cofactors

Таблица 4
Сокращение числа кофакторов на основе решения задачи о раскраске графа

Table 4
Reduction of the number of cofactors based on solving the graph coloring problem

Кофактор Cofactor $p_i(x_2)$	x_2			Краска Color	Доопределение Additional definition			Реализация Implementation	
	$x_2^{\{0\}}$	$x_2^{\{1\}}$	$x_2^{\{2\}}$		Кофактор Cofactor $h^i(x_2)$	$x_2^{\{0\}}$	$x_2^{\{1\}}$		$x_2^{\{2\}}$
p_4	-	1	-	1	h^1	0	1	0	$p_4 \prec h^1, p_6 \prec h^1$
p_6	0	-	0						
p_1	-	1	2	2	h^2	2	1	2	$p_1 \prec h^2, p_8 \prec h^2$
p_8	2	1	2						
p_2	1	-	1	3	h^3	1	-	1	$p_2 \prec h^3$
p_3	2	0	1	4	h^4	2	0	1	$p_3 \prec h^4, p_5 \prec h^4, p_7 \prec h^4$
p_5	2	0	1						
p_7	2	-	-						

Заменяя кофакторы в MDD (см. рис. 1) реализующими их кофакторами, можно уменьшить число функциональных вершин MDD за счет доопределения кофакторов (рис. 3). Примеры правил сокращения MDD показаны на рис. 4, они аналогичны известным правилам редуцирования BDD. В результате выполнения этапа 1 была построена MDD, у которой только две листовые вершины (рис. 3) имеют неопределенное значение.

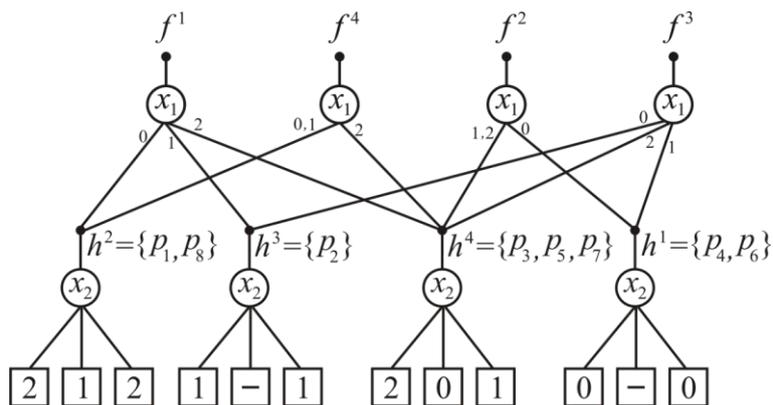


Рис. 3. Доопределение MDD
Fig. 3. MDD additional definition

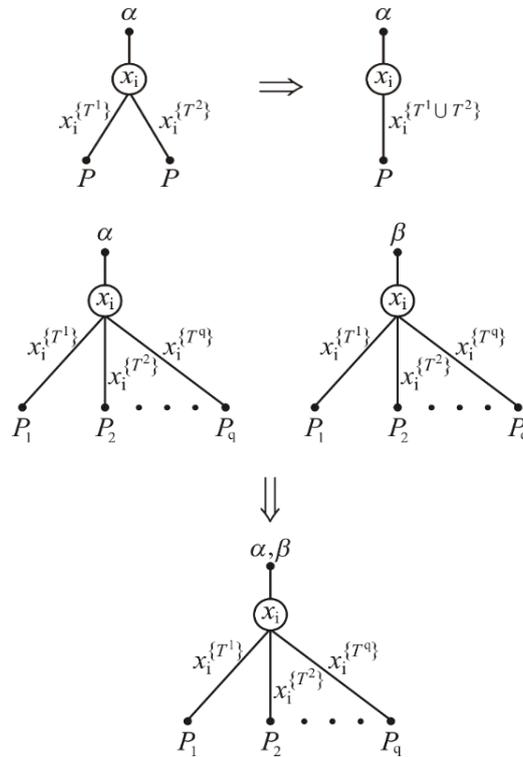


Рис. 4. Локальные правила редукции MDD

Fig. 4. Local MDD reduction rules

Если для некоторого кофактора все листовые значения являются неопределенными, то будем обозначать такой кофактор p_- . Он представляет собой частичную многозначную функцию, принимающую неопределенные значения для всех наборов значений многозначных аргументов. Кофактор p_- совместим с любым другим частичным кофактором.

Эвристический алгоритм поиска лучшей перестановки переменных разложения Шеннона для построения MDD состоит из следующих шагов:

Шаг 1. Поочередно рассматриваются все входные переменные. Для каждой переменной строится начальный уровень MDD, затем для полученных частичных кофакторов решается задача А. Выбирается та переменная x_i , для которой в результате решения задачи А в системе H будет наименьшее число кофакторов (эвристика 1). Переменная x_i определяет первую переменную искомой перестановки переменных, по которой будет строиться MDD.

Шаг 2 (итеративный). К полученной переменной (перестановке) поочередно добавляется одна из оставшихся переменных и решается задача А. Выбирается та из переменных x_{i_2} , которая обеспечивает получение наилучшего по числу функций системы H решения задачи А (эвристика 2). Такой выбор переменных продолжается до тех пор, пока не будут определены все переменные в искомой перестановке $\langle x_{i_1}, x_{i_2}, \dots, x_{i_n} \rangle$.

Заметим, что аналогичный алгоритм оказался эффективным для минимизации числа вершин BDD, представляющей систему полностью определенных булевых функций [18].

Кодирование значений многозначных аргументов и многозначных функций (этап 2). Длина m булева кода для многозначной переменной x_i и многозначной функции f^j системы F определяется из соотношения $m = \lceil \log_2 k \rceil$, где k – мощность множества S , а $\lceil A \rceil$ – ближайшее целое, большее либо равное A .

Выполним кодирование многозначных переменных и литералов согласно табл. 5 и 6. В рассматриваемом примере $k = 3$, поэтому каждая многозначная переменная x_1, x_2 и каждая многозначная функция f^1, f^2, f^3, f^4 системы F кодируются своим булевым кодом длиной $m = 2$ (табл. 5).

Таблица 5
Кодирование многозначных переменных, функций и кофакторов булевыми переменными

Table 5
Encoding of multi-valued variables, functions and cofactors with Boolean variables

Многозначная переменная, функция, кофактор <i>Multi-valued variable, function, cofactor</i>	Булев код <i>Boolean code</i>
x_1	(v^1, v^2)
x_2	(w^1, w^2)
f^1	(f_1^1, f_2^1)
f^2	(f_1^2, f_2^2)
f^3	(f_1^3, f_2^3)
f^4	(f_1^4, f_2^4)
h^1	(h_1^1, h_2^1)
h^2	(h_1^2, h_2^2)
h^3	(h_1^3, h_2^3)
h^4	(h_1^4, h_2^4)
p_-	(c_-, c_-)

Таблица 6
Кодирование литералов

Table 6
Encoding literals

Многозначная переменная <i>Multi-valued variable</i>	Булев код <i>Boolean code</i>	Литерал многозначной переменной <i>Literal of a multi-valued variable</i>	Кодирующее выражение <i>Encoding expression</i>
x_1	(v^1, v^2)	$x_1^{(0)}$	$\bar{v}^1 \bar{v}^2$
		$x_1^{(1)}$	$\bar{v}^1 v^2$
		$x_1^{(2)}$	$v^1 \bar{v}^2$
x_2	(w^1, w^2)	$x_2^{(0)}$	$\bar{w}^1 \bar{w}^2$
		$x_2^{(1)}$	$\bar{w}^1 w^2$
		$x_2^{(2)}$	$w^1 \bar{w}^2$

Одноэлементные литералы многозначных переменных кодируются булевыми выражениями в виде произведений литералов булевых переменных (табл. 6). Многоэлементным литералам соответствуют некоторые ДНФ – дизъюнкции полных произведений одноэлементных литера-

лов, которые могут быть упрощены по законам булевой алгебры. Например, двухэлементный литерал $x_1^{(0,2)}$ кодируется ДНФ $\bar{v}^1\bar{v}^2 \vee v^1\bar{v}^2 = \bar{v}^2$. Для некоторых комбинаций булевых кодирующих переменных могут отсутствовать соответствующие им многозначные значения, в рассматриваемом примере такими комбинациями кодирующих булевых переменных являются (v^1v^2) , (w^1w^2) . Эти комбинации использовались бы, если бы элемент 3 входил в множество S . В результате кодирования MDD для каждой из многозначных функций исходной системы превращается в BDD векторной булевой функции, который можно задать множеством BDD, реализующих компонентные функции векторной булевой функции (пример дан ниже).

Согласно MDD (см. рис. 3) запишем следующие представления доопределенных частичных многозначных функций системы F :

$$\begin{aligned} f^1 &= x_1^{(0)}h^2 \vee x_1^{(1)}h^3 \vee x_1^{(2)}h^4, \\ f^2 &= x_1^{(1,2)}h^4 \vee x_1^{(0)}h^1, \\ f^3 &= x_1^{(0)}h^3 \vee x_1^{(1)}h^1 \vee x_1^{(2)}h^4, \\ f^4 &= x_1^{(0,1)}h^2 \vee x_1^{(2)}h^4. \end{aligned}$$

Тогда уравнения для разложений Шеннона соответствующих двухкомпонентных частичных векторных булевых функций примут вид

$$\begin{aligned} (f_1^1, f_2^1) &= \bar{v}^1\bar{v}^2(h_1^2, h_2^2) \vee \bar{v}^1v^2(h_1^3, h_2^3) \vee v^1\bar{v}^2(h_1^4, h_2^4) \vee v^1v^2(c_-, c_-), \\ (f_1^2, f_2^2) &= (\bar{v}^1v^2 \vee v^1\bar{v}^2)(h_1^4, h_2^4) \vee \bar{v}^1\bar{v}^2(h_1^1, h_2^1) \vee v^1v^2(c_-, c_-), \\ (f_1^3, f_2^3) &= \bar{v}^1\bar{v}^2(h_1^3, h_2^3) \vee \bar{v}^1v^2(h_1^1, h_2^1) \vee v^1\bar{v}^2(h_1^4, h_2^4) \vee v^1v^2(c_-, c_-), \\ (f_1^4, f_2^4) &= (\bar{v}^1\bar{v}^2 \vee \bar{v}^1v^2)(h_1^2, h_2^2) \vee v^1\bar{v}^2(h_1^4, h_2^4) \vee v^1v^2(c_-, c_-). \end{aligned} \quad (7)$$

Здесь кофакторы являются частичными булевыми функциями h_i^j , зависящими от булевых переменных w^1, w^2 . Булевы функции h_i^j заданы в табл. 7.

В уравнения (7) для каждой из векторных булевых функций (f_1^1, f_2^1) , (f_1^2, f_2^2) , (f_1^3, f_2^3) , (f_1^4, f_2^4) добавлен дизъюнктивный член $v^1v^2(c_-, c_-)$. Из этого следует, что на наборах переменных, где $v^1 = 1$, $v^2 = 1$, все значения компонентных булевых функций являются неопределенными:

v^1	v^2	w^1	w^2	f_1^1	f_2^1	f_1^2	f_2^2	f_1^3	f_2^3	f_1^4	f_2^4
1	1	0	0	–	–	–	–	–	–	–	–
1	1	0	1	–	–	–	–	–	–	–	–
1	1	1	0	–	–	–	–	–	–	–	–
1	1	1	1	–	–	–	–	–	–	–	–

Вместе с тем эти значения могут заменяться определенными при логической оптимизации. В двухкомпонентном векторе (c_-, c_-) через c_- обозначена частичная булева функция, принимающая неопределенные значения для всех наборов значений булевых аргументов.

Таким образом, если в результате выполнения этапа 1 MDD для многозначной функции не содержит неопределенных листовых вершин (исходная частичная многозначная функция доопределена до полностью определенной), то в результате кодирования на этапе 2 полученная система булевых функций может оказаться частичной.

Доопределение частичных булевых функций (этап 3). Можно провести доопределение частичных булевых кофакторов из табл. 7 до полностью определенных, заменяя булевы неопределенные значения «-» нулями и единицами так, чтобы многоуровневое BDD-представление реализующих полностью определенных булевых функций стало как можно менее сложным. Проведение таких доопределений описано в статье [19], где эта задача сведена к нахождению кратчайшей имплицитующей формы троичной матрицы, задающей значения частичных булевых функций.

В рассматриваемом примере булевых функций (табл. 7) в последней строке указаны алгебраические формы полностью определенных булевых функций, полученные путем доопределения частичных функций h_i^j . Например, заменяя неопределенные значения функции h_2^3 единицами, получаем, что $h_2^3 = 1$. Это аналогично и для других функций. Приведем отношения реализации частичных булевых кофакторов полностью определенными булевыми кофакторами, заданными в последней строке табл. 7:

$$h_1^1 < 0, h_2^1 < 0, h_1^2 < \bar{w}_2, h_2^2 < w_2, h_1^3 < 0, h_2^3 < 1, h_1^4 < \bar{w}^1\bar{w}^2, h_2^4 < w^1w^2.$$

Таблица 7
Замена многозначных кофакторов системами частичных булевых функций

Table 7
Replacement of multi-valued cofactors by systems of partial Boolean functions

Булевы переменные <i>Boolean variables</i>		Многозначный кофактор <i>Multi-valued cofactor</i>							
		h^1		h^2		h^3		h^4	
		Система частичных булевых функций <i>System of partial Boolean functions</i>							
		1		2		3		4	
w^1	w^2	h_1^1	h_2^1	h_1^2	h_2^2	h_1^3	h_2^3	h_1^4	h_2^4
0	0	0	0	1	0	0	1	1	0
0	1	-	-	0	1	-	-	0	0
1	0	0	0	1	0	0	1	0	1
1	1	-	-	-	-	-	-	-	-
Реализующие булевы функции <i>Implementing Boolean functions</i>		0	0	\bar{w}^2	w^2	0	1	$\bar{w}^1\bar{w}^2$	w^1w^2

От векторных уравнений (7) перейдем к алгебраическому заданию частичных компонентных булевых функций:

$$\begin{aligned} f_1^1 &= \bar{v}^1\bar{v}^2h_1^2 \vee \bar{v}^1v^2h_1^3 \vee v^1\bar{v}^2h_1^4 \vee v^1v^2c_-; & f_2^1 &= \bar{v}^1\bar{v}^2h_2^2 \vee \bar{v}^1v^2h_2^3 \vee v^1\bar{v}^2h_2^4 \vee v^1v^2c_-; \\ f_1^2 &= (\bar{v}^1v^2 \vee v^1\bar{v}^2)h_1^4 \vee \bar{v}^1\bar{v}^2h_1^1 \vee v^1v^2c_-; & f_2^2 &= (\bar{v}^1v^2 \vee v^1\bar{v}^2)h_2^4 \vee \bar{v}^1\bar{v}^2h_2^1 \vee v^1v^2c_-; \\ f_1^3 &= \bar{v}^1\bar{v}^2h_1^3 \vee \bar{v}^1v^2h_1^1 \vee v^1\bar{v}^2h_1^4 \vee v^1v^2c_-; & f_2^3 &= \bar{v}^1\bar{v}^2h_2^3 \vee \bar{v}^1v^2h_2^1 \vee v^1\bar{v}^2h_2^4 \vee v^1v^2c_-; \\ f_1^4 &= (\bar{v}^1\bar{v}^2 \vee \bar{v}^1v^2)h_1^2 \vee v^1\bar{v}^2h_1^4 \vee v^1v^2c_-; & f_2^4 &= (\bar{v}^1\bar{v}^2 \vee \bar{v}^1v^2)h_2^2 \vee v^1\bar{v}^2h_2^4 \vee v^1v^2c_- \end{aligned} \quad (8)$$

Константа c_- в задании (8) может быть доопределена до любой частичной либо полностью определенной функции и при упрощении логических выражений заменена этой функцией. Другими словами, функция c_- совместима с любой частичной функцией. Например,

$$v^1\bar{v}^2h_1^4 \vee v^1v^2c_- = v^1\bar{v}^2h_1^4 \vee v^1v^2h_1^4 = v^1(\bar{v}^2 \vee v^2)h_1^4 = v^1h_1^4.$$

В этом случае была использована замена c_- на h_1^4 . Упрощая уравнения (8) и записывая реализующие булевы кофакторы, получим более компактные уравнения компонентных полностью определенных булевых функций:

$$\begin{aligned} f_1^1 &= \bar{v}^1 \bar{v}^2 h_1^2 \vee \bar{v}^1 v^2 h_1^3 \vee v^1 h_1^4; & f_2^1 &= \bar{v}^1 \bar{v}^2 h_2^2 \vee \bar{v}^1 v^2 h_2^3 \vee v^1 h_2^4; \\ f_1^2 &= (\bar{v}^1 v^2 \vee v^1) h_1^4 \vee \bar{v}^1 \bar{v}^2 h_1^1; & f_2^2 &= (\bar{v}^1 v^2 \vee v^1) h_2^4 \vee \bar{v}^1 \bar{v}^2 h_2^1; \\ f_1^3 &= \bar{v}^1 \bar{v}^2 h_1^3 \vee \bar{v}^1 v^2 h_1^1 \vee v^1 h_1^4; & f_2^3 &= \bar{v}^1 \bar{v}^2 h_2^3 \vee \bar{v}^1 v^2 h_2^1 \vee v^1 h_2^4; \\ f_1^4 &= \bar{v}^1 h_1^2 \vee v^1 h_1^4; & f_2^4 &= \bar{v}^1 h_2^2 \vee v^1 h_2^4; \\ h_1^1 &= 0; & h_2^1 &= 0; & h_1^2 &= \bar{w}^2; & h_2^2 &= w^2; & h_1^3 &= 0; & h_2^3 &= 1; & h_1^4 &= \bar{w}^1 \bar{w}^2; & h_2^4 &= w^1 \bar{w}^2. \end{aligned} \quad (9)$$

Проверим, реализует ли с учетом выбранного кодирования значений многозначных переменных и функций логическая модель (9) исходное табличное задание (см. табл. 1). По уравнениям (9) построим таблицу истинности булевых функций (табл. 8), элиминируя (устраняя) промежуточные переменные.

Таблица 8
Таблица истинности системы булевых функций

Table 8
Truth table of a system of Boolean functions

x_1	x_2	f^1	f^2	f^3	f^4
$v^1 \ v^2$	$w^1 \ w^2$	$f_1^1 \ f_2^1$	$f_1^2 \ f_2^2$	$f_1^3 \ f_2^3$	$f_1^4 \ f_2^4$
00	00	10	00	01	10
00	01	01	00	01	01
00	10	10	00	01	10
00	11	01	00	01	01
01	00	01	10	00	10
01	01	01	00	00	01
01	10	01	01	00	10
01	11	01	00	00	01
10	00	10	10	10	10
10	01	00	00	00	00
10	10	01	01	01	01
10	11	00	00	00	00
11	00	10	10	10	10
11	01	00	00	00	00
11	10	01	01	01	01
11	11	00	00	00	00

Полужирным шрифтом в табл. 8 выделены коды значений аргументов и значений исходных многозначных функций, заданных в табл. 1. Курсивом отмечены булевы коды, которыми были заменены неопределенные значения многозначных функций. Выполняя замену неопределенных многозначных значений в табл. 1 определенными значениями, получим систему полностью определенных многозначных функций (табл. 9). Заметим, что различные неопределенные многозначные значения « \rightarrow » (см. табл. 1) заменены в табл. 9 определенными значениями из множества $\{0, 1, 2\}$.

Таблица 9
Доопределение многозначных функций системы F

Table 9
Additional definition of multi-valued functions of the system F

Многозначный аргумент Multi-valued argument		Полностью определенная многозначная функция Fully defined multi-valued function			
x_1	x_2	f^1	f^2	f^3	f^4
0	0	2	0	1	2
0	1	1	1	1	1
0	2	2	0	1	2
1	0	1	2	0	2
1	1	1	0	0	1
1	2	1	1	0	2
2	0	2	2	2	2
2	1	0	0	0	0
2	2	1	1	1	1

От первых шести уравнений (9) легко перейти к уравнениям, соответствующим BDD, проведя разложения Шеннона ДНФ булевых функций, и получить другой вид представления полностью определенных функций, реализующих исходную систему многозначных функций:

$$\begin{aligned}
 f_1^1 &= \bar{v}^1 z_1 \vee v^1 h_1^4; & z_1 &= \bar{v}^2 h_1^2 \vee v^2 h_1^3; & f_2^1 &= \bar{v}^1 z_3 \vee v^1 h_2^4; & z_3 &= \bar{v}^2 h_2^2 \vee v^2 h_2^3; \\
 f_1^2 &= \bar{v}^1 z_4 \vee v^1 h_1^4; & z_4 &= \bar{v}^2 h_1^1 \vee v^2 h_1^4; & f_2^2 &= \bar{v}^1 z_6 \vee v^1 h_2^4; & z_6 &= \bar{v}^2 h_2^1 \vee v^2 h_2^4; \\
 f_1^3 &= \bar{v}^1 z_9 \vee v^1 h_1^4; & z_9 &= \bar{v}^2 h_1^3 \vee v^2 h_1^1; & f_2^3 &= \bar{v}^1 z_{10} \vee v^1 h_2^4; & z_{10} &= \bar{v}^2 h_2^3 \vee v^2 h_2^1; \\
 f_1^4 &= \bar{v}^1 h_1^2 \vee v^1 h_1^4; & f_2^4 &= \bar{v}^1 h_2^2 \vee v^1 h_2^4; \\
 h_1^1 &= 0; & h_2^1 &= 0; & h_1^2 &= \bar{w}^2; & h_2^2 &= w^2; & h_1^3 &= 0; & h_2^3 &= 1; & h_1^4 &= \bar{w}^1 \bar{w}^2; & h_2^4 &= w^1 \bar{w}^2.
 \end{aligned}
 \tag{10}$$

Процесс перехода от разложения Шеннона для многозначной функции f^1 (см. рис. 3) к BDD показан на рис. 5. Для других многозначных функций – аналогично.

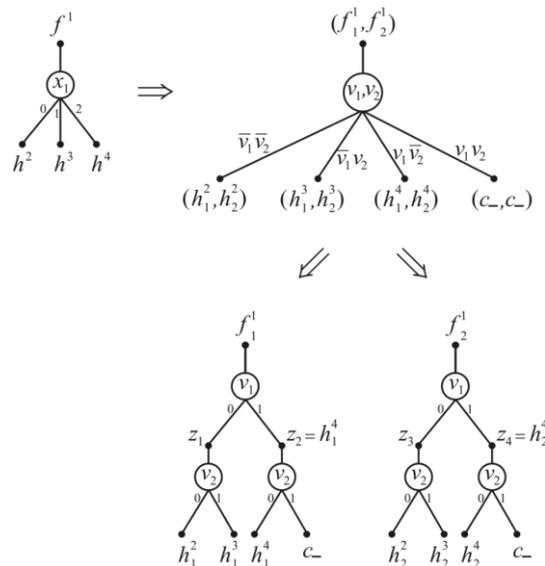


Рис. 5. Переход от разложения Шеннона многозначной функции (подграфа MDD) к BDD частных булевых функций
Fig. 5. Transition from the Shannon expansion of a multi-valued function (subgraph of MDD) to BDD partial Boolean functions

Для полученной системы полностью определенных булевых функций можно применять различные методы логической оптимизации, например минимизировать представления функций в классе ДНФ. Также можно использовать программу Espresso [12] для минимизации булевых функций из табл. 8 и получить минимизированные ДНФ, матричные формы которых заданы в табл. 10.

Таблица 10
Минимизированные ДНФ булевых функций

Table 10
Minimized SOPs of Boolean functions

x_1	x_2	f^1	f^2	f^3	f^4
$v^1 v^2$	$w^1 w^2$	$f_1^1 f_2^1$	$f_1^2 f_2^2$	$f_1^3 f_2^3$	$f_1^4 f_2^4$
0 0	- -	0 0	0 0	0 1	0 0
0 1	0 0	0 1	1 0	0 0	1 0
0 1	1 0	0 1	0 1	0 0	1 0
0 0	- 0	1 0	0 0	0 0	1 0
1 -	0 0	1 0	1 0	1 0	1 0
0 -	- 1	0 1	0 0	0 0	0 1
1 -	1 0	0 1	0 1	0 1	0 1

Табличное задание компонентных булевых функций из табл. 10 соответствует следующим формулам:

$$\begin{aligned}
 f_1^1 &= \bar{v}^1 \bar{v}^2 \bar{w}^2 \vee v^1 \bar{w}^1 \bar{w}^2; & f_2^1 &= \bar{v}^1 v^2 \bar{w}^1 \bar{w}^2 \vee \bar{v}^1 v^2 w^1 \bar{w}^2 \vee \bar{v}^1 w^2 \vee v^1 w^1; \\
 f_1^2 &= \bar{v}^1 v^2 \bar{w}^1 \bar{w}^2 \vee v^1 \bar{w}^1 \bar{w}^2; & f_2^2 &= \bar{v}^1 v^2 w^1 \bar{w}^2 \vee v^1 w^1; \\
 f_1^3 &= v^1 \bar{w}^1 \bar{w}^2; & f_2^3 &= \bar{v}^1 \bar{v}^2 \vee v^1 w^1; \\
 f_1^4 &= \bar{v}^1 v^2 \bar{w}^1 \bar{w}^2 \vee \bar{v}^1 v^2 w^1 \bar{w}^2 \vee \bar{v}^1 \bar{v}^2 \bar{w}^2 \vee v^1 \bar{w}^1 \bar{w}^2; & f_2^4 &= \bar{v}^1 w^2 \vee v^1 w^1.
 \end{aligned}
 \tag{11}$$

Синтез логической схемы (этап 4). По логическим уравнениям (9)–(11) были разработаны соответствующие VHDL-описания [20] путем замены логических операций операциями языка VHDL. Схемная реализация полученных VHDL-описаний в библиотеке проектирования отечественных КМОП СБИС выполнялась с помощью синтезатора LeonardoSpectrum [16]. Функции логических КМОП-элементов используемой библиотеки синтеза приведены в работе [15]. Для каждого VHDL-описания синтез осуществлялся с одними и теми же опциями управления синтезом. Для полученной схемы подсчитывались ее площадь (сумма площадей элементов, у. е.) и временная задержка. Результаты синтеза даны в табл. 11.

Таблица 11
Параметры логических схем, реализующих систему
многозначных функций

Table 11
Parameters of logic circuits implementing a system
of multi-valued functions

Функциональное описание схемы <i>Functional description of the circuit</i>	Площадь схемы, у. е. <i>Area of the circuit, c. u.</i>	Задержка схемы, нс <i>Circuit delay, ns</i>
Описание MDD (9)	5 519	1,61
Описание BDD (10)	5 284	2,07
Система ДНФ (11)	6 344	1,69
Исходное описание (см. табл. 1)	6 333	2,30

В строке «Исходное описание» табл. 11 представлены результаты синтеза минимизированной (программой Espresso) системы из восьми частичных функций, полученных из функций системы F кодированием многозначных значений как аргументов, так и функций системы F . Площадь схемы (5 284), полученной по BDD-представлению, оказалась наименьшей, а резуль-

таты синтеза схем после минимизации в классе ДНФ (две последние строки) показали худшие результаты. Можно заметить также, что изменения формы задания функций влияют на результаты синтеза даже для описаний схем небольшой размерности.

Пример 2. Проведем схемную реализацию функционального описания умножителя по модулю 5, заданного частичной многозначной функцией (табл. 12). Произведение по модулю p (основание модулярной вычислительной системы) для двух целых неотрицательных чисел (операндов) a, b , находящихся в диапазоне $\{0, 1, \dots, p - 1\}$, вычисляется по формуле

$$|a \times b|_p = (a \times b) - \left\lfloor \frac{a \times b}{p} \right\rfloor \times p,$$

где через $\lfloor k \rfloor$ обозначена целая часть числа, т. е. ближайшее целое, меньшее либо равное k . В случае если $(a \times b) < p$, то $|a \times b|_p = a \times b$. При схемной реализации умножителя по модулю p , подача на вход схемы значения, большего $p - 1$, хотя бы для одного из операндов a, b , соответствует неопределенному значению частичной многозначной функции y .

Таблица 12
Умножитель по модулю 5 (частичная многозначная функция)

Table 12
The multiplier modulo 5 is a partial multi-valued function

Область определенных значений многозначной функции <i>The domain of the defined values of a multi-valued function</i>			Область неопределенных значений многозначной функции <i>The domain of undefined values of a multi-valued function</i>		
<i>a</i>	<i>b</i>	<i>y</i>	<i>a</i>	<i>b</i>	<i>y</i>
0	0	0	0	5	-
0	1	0	0	6	-
0	2	0	0	7	-
0	3	0	1	5	-
0	4	0	1	6	-
1	0	0	1	7	-
1	1	1	2	5	-
1	2	2	2	6	-
1	3	3	2	7	-
1	4	4	3	5	-
2	0	0	3	6	-
2	1	2	3	7	-
2	2	4	4	5	-
2	3	1	4	6	-
2	4	3	4	7	-
3	0	0	5	0	-
3	1	3	5	1	-
3	2	1	5	2	-
3	3	4	5	3	-
3	4	2	5	4	-
4	0	0	5	5	-
4	1	4	5	6	-
4	2	3	5	7	-
4	3	2	6	0	-
4	4	1	6	1	-
			6	2	-
			6	3	-
			6	4	-
			6	5	-
			6	6	-
			6	7	-
			7	0	-
			7	1	-
			7	2	-
			7	3	-
			7	4	-
			7	5	-
			7	6	-
			7	7	-

Для умножителя по модулю 5 множество $S = \{0, 1, 2, 3, 4\}$ составляют пять чисел $|S| = 5$; для их представления нужен трехбитный двоичный код, так как $m = \lceil \log_2 5 \rceil = 3$.

Проведем схемную реализацию умножителя по модулю 5, не описывая подробно выполненные этапы 1–4. Будем строить MDD многозначной функции y по перестановке переменных $\langle a, b \rangle$. Разложение по переменной a показано на рис. 6, кофактор p_0 доопределяется до 0 (многозначного), кофактор p_1 – до многозначной переменной b . Разложение кофакторов p_2, p_3, p_4 по переменной b изображено на рис. 7.

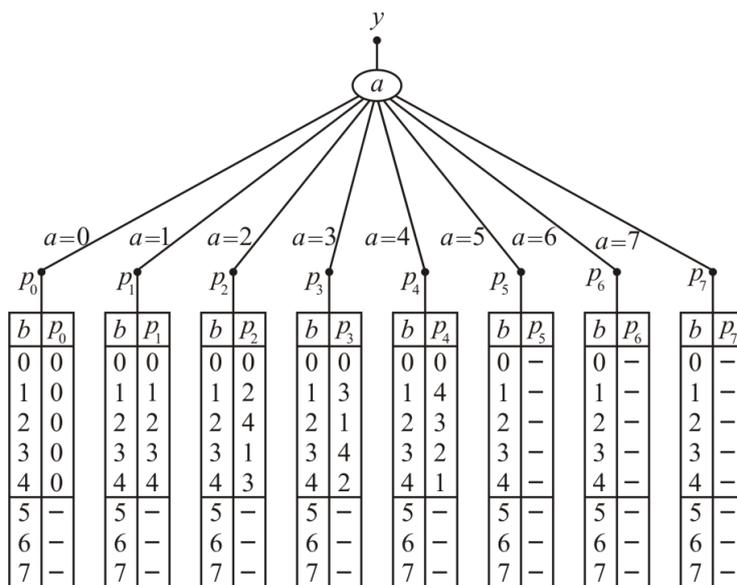


Рис. 6. Первый шаг построения MDD – разложение Шеннона по переменной a
Fig. 6. The first step of MDD construction is the Shannon expansion by variable a

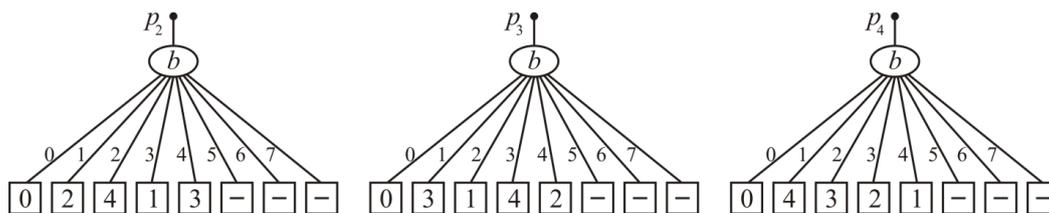


Рис. 7. Второй шаг построения MDD – разложение Шеннона нетривиальных кофакторов p_2, p_3, p_4 по переменной b
Fig. 7. The second step of MDD construction is the Shannon expansion of nontrivial cofactors p_2, p_3, p_4 by variable b

Разложение Шеннона многозначной функции y по переменной a имеет вид

$$y = a^{\{0\}} p_0 \vee a^{\{1\}} p_1 \vee a^{\{2\}} p_2 \vee a^{\{3\}} p_3 \vee a^{\{4\}} p_4 \vee a^{\{5\}} p_- \vee a^{\{6\}} p_- \vee a^{\{7\}} p_- \tag{12}$$

Многозначные кофакторы p_0, p_1, p_2, p_3, p_4 попарно несовместимы. Кофактор p_- совместим с любым из перечисленных ранее. Позже кофактор p_- будет заменен кофактором p_4 , эту замену можно будет сделать и после выполнения этапа 2 булева кодирования значений многозначных переменных. Кодирование переменных и функции булевыми кодами представлено в табл. 13.

Таблица 13
Кодирование переменных и функции умножителя по модулю 5

Table 13
Coding of variables and multiplier functions modulo 5

Многозначная переменная, функция <i>Multi-valued variable, function</i>	Булев код <i>Boolean code</i>
<i>a</i>	(a_1, a_2, a_3)
<i>b</i>	(b_1, b_2, b_3)
<i>y</i>	(y_0, y_1, y_2)

Кодирование каждого многозначного кофактора своим булевым кодом приводит к пяти системам булевых функций (табл. 14).

Таблица 14
Замена многозначных кофакторов системами частичных булевых функций

Table 14
Replacement of multi-valued cofactors by systems of partial Boolean functions

Булева переменная <i>Boolean variable</i>	Кофактор <i>Cofactor</i>				
	<i>p</i> ₀	<i>p</i> ₁	<i>p</i> ₂	<i>p</i> ₃	<i>p</i> ₄
	Система частичных булевых функций <i>System of partial Boolean functions</i>				
<i>b</i> ₁ <i>b</i> ₂ <i>b</i> ₃	<i>c</i> ₁ <i>c</i> ₂ <i>c</i> ₃	<i>c</i> ₄ <i>c</i> ₅ <i>c</i> ₆	<i>c</i> ₇ <i>c</i> ₈ <i>c</i> ₉	<i>c</i> ₁₀ <i>c</i> ₁₁ <i>c</i> ₁₂	<i>c</i> ₁₃ <i>c</i> ₁₄ <i>c</i> ₁₅
0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0
0 0 1	0 0 0	0 0 1	0 1 0	0 1 1	1 0 0
0 1 0	0 0 0	0 1 0	1 0 0	0 0 1	0 1 1
0 1 1	0 0 0	0 1 1	0 0 1	1 0 0	0 1 0
1 0 0	0 0 0	1 0 0	0 1 1	0 1 0	0 1 0
1 0 1	- - -	- - -	- - -	- - -	- - -
1 1 0	- - -	- - -	- - -	- - -	- - -
1 1 1	- - -	- - -	- - -	- - -	- - -

Анализируя булевы кофакторы из табл. 14, можно установить, что каждый из кофакторов *c*₁, *c*₂, *c*₃ доопределяется до булева нуля (*c*₁ = *c*₂ = *c*₃ = 0), а булевы кофакторы *c*₄, *c*₅, *c*₆ доопределяются до булевых переменных (*c*₄ = *b*₁, *c*₅ = *b*₂, *c*₆ = *b*₃). Кроме того, *c*₈ = *c*₁₁ (в табл. 14 *c*₈, *c*₁₁ выделены полужирным курсивом).

Используя коды многозначных переменных и кофакторов, перейдем от разложения Шеннона (12) частичной многозначной функции к представлению частичной векторной булевой функции:

$$(y_0, y_1, y_2) = \bar{a}_1 \bar{a}_2 \bar{a}_3 (c_1, c_2, c_3) \vee \bar{a}_1 \bar{a}_2 a_3 (c_4, c_5, c_6) \vee \bar{a}_1 a_2 \bar{a}_3 (c_7, c_8, c_9) \vee \bar{a}_1 a_2 a_3 (c_{10}, c_{11}, c_{12}) \vee a_1 \bar{a}_2 \bar{a}_3 (c_{13}, c_{14}, c_{15}) \vee a_1 \bar{a}_2 a_3 (c_-, c_-, c_-) \vee a_1 a_2 \bar{a}_3 (c_-, c_-, c_-) \vee a_1 a_2 a_3 (c_-, c_-, c_-). \quad (13)$$

Выражение (13) соответствует представлению частичной векторной функции, компонентами которой являются частичные булевы функции *y*₀, *y*₁, *y*₂, зависящие от булевых переменных *a*₁, *a*₂, *a*₃, *b*₁, *b*₂, *b*₃. По формуле (13) построим BDD частичной векторной булевой функции (рис. 8) и упростим его, используя соответствующее доопределение трехкомпонентных листовых вершин и правила сокращения BDD (рис. 9).

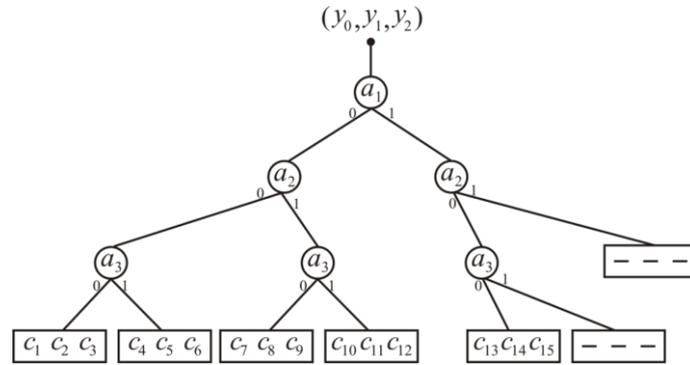


Рис. 8. BDD векторной булевой функции
Fig. 8. BDD of a vector Boolean function

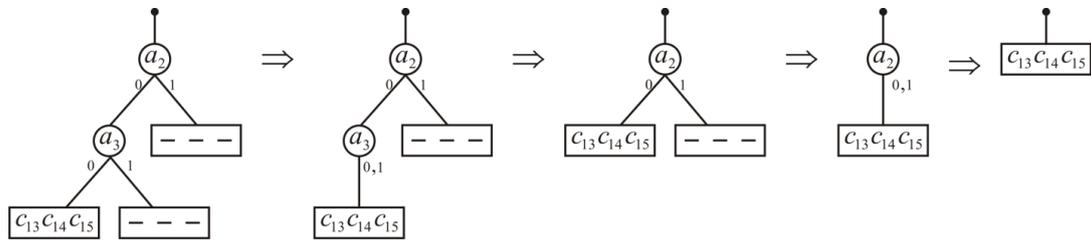


Рис. 9. Редукция подграфа BDD
Fig. 9. Reduction of the BDD subgraph

По формуле (13) запишем и упростим представления компонентных булевых функций y_0, y_1, y_2 :

$$\begin{aligned} y_0 &= \bar{a}_1 \bar{a}_2 \bar{a}_3 c_1 \vee \bar{a}_1 \bar{a}_2 a_3 c_4 \vee \bar{a}_1 a_2 \bar{a}_3 c_7 \vee \bar{a}_1 a_2 a_3 c_{10} \vee a_1 \bar{a}_2 \bar{a}_3 c_{13} \vee a_1 \bar{a}_2 a_3 c_- \vee a_1 a_2 \bar{a}_3 c_- \vee a_1 a_2 a_3 c_- = \\ &= \bar{a}_1 \bar{a}_2 \bar{a}_3 c_1 \vee \bar{a}_1 \bar{a}_2 a_3 c_4 \vee \bar{a}_1 a_2 \bar{a}_3 c_7 \vee \bar{a}_1 a_2 a_3 c_{10} \vee (a_1 \bar{a}_2 \bar{a}_3 c_{13} \vee a_1 \bar{a}_2 a_3 c_- \vee a_1 a_2 \bar{a}_3 c_- \vee a_1 a_2 a_3 c_-) = \\ &= \bar{a}_1 \bar{a}_2 \bar{a}_3 c_1 \vee \bar{a}_1 \bar{a}_2 a_3 c_4 \vee \bar{a}_1 a_2 \bar{a}_3 c_7 \vee \bar{a}_1 a_2 a_3 c_{10} \vee a_1 c_{13}. \end{aligned}$$

Здесь использована замена c_- на c_{13} :

$$\begin{aligned} (a_1 \bar{a}_2 \bar{a}_3 c_{13} \vee a_1 \bar{a}_2 a_3 c_- \vee a_1 a_2 \bar{a}_3 c_- \vee a_1 a_2 a_3 c_-) &= a_1 \bar{a}_2 \bar{a}_3 c_{13} \vee a_1 \bar{a}_2 a_3 c_{13} \vee a_1 a_2 \bar{a}_3 c_{13} \vee a_1 a_2 a_3 c_{13} = \\ &= a_1 c_{13} (\bar{a}_2 \bar{a}_3 \vee \bar{a}_2 a_3 \vee a_2 \bar{a}_3 \vee a_2 a_3) = a_1 c_{13} 1 = a_1 c_{13}. \end{aligned}$$

Так как $c_1 = 0, c_4 = b_1$, то

$$y_0 = \bar{a}_1 \bar{a}_2 a_3 b_1 \vee \bar{a}_1 a_2 \bar{a}_3 c_7 \vee \bar{a}_1 a_2 a_3 c_{10} \vee a_1 c_{13}. \tag{14}$$

Аналогично получим

$$\begin{aligned} y_1 &= \bar{a}_1 \bar{a}_2 \bar{a}_3 c_2 \vee \bar{a}_1 \bar{a}_2 a_3 c_5 \vee \bar{a}_1 a_2 \bar{a}_3 c_8 \vee \bar{a}_1 a_2 a_3 c_{11} \vee a_1 \bar{a}_2 \bar{a}_3 c_{14} \vee a_1 \bar{a}_2 a_3 c_- \vee a_1 a_2 \bar{a}_3 c_- \vee a_1 a_2 a_3 c_- = \\ &= \bar{a}_1 \bar{a}_2 \bar{a}_3 c_2 \vee \bar{a}_1 \bar{a}_2 a_3 c_5 \vee \bar{a}_1 a_2 \bar{a}_3 c_8 \vee \bar{a}_1 a_2 a_3 c_{11} \vee (a_1 \bar{a}_2 \bar{a}_3 c_{14} \vee a_1 \bar{a}_2 a_3 c_- \vee a_1 a_2 \bar{a}_3 c_- \vee a_1 a_2 a_3 c_-) = \\ &= \bar{a}_1 \bar{a}_2 \bar{a}_3 c_2 \vee \bar{a}_1 \bar{a}_2 a_3 c_5 \vee \bar{a}_1 a_2 \bar{a}_3 c_8 \vee \bar{a}_1 a_2 a_3 c_{11} \vee a_1 c_{14}. \end{aligned}$$

Поскольку $c_2 = 0, c_5 = b_2, c_8 = c_{11}$, то

$$y_1 = \bar{a}_1 \bar{a}_2 a_3 b_2 \vee \bar{a}_1 a_2 c_8 \vee a_1 c_{14}. \tag{15}$$

Функция y_2 запишется в виде

$$\begin{aligned} y_2 &= \bar{a}_1\bar{a}_2\bar{a}_3c_3 \vee \bar{a}_1\bar{a}_2a_3c_6 \vee \bar{a}_1a_2\bar{a}_3c_9 \vee \bar{a}_1a_2a_3c_{12} \vee a_1\bar{a}_2\bar{a}_3c_{14} \vee a_1\bar{a}_2a_3c_{15} \vee a_1a_2\bar{a}_3c_{12} \vee a_1a_2a_3c_{15} = \\ &= \bar{a}_1\bar{a}_2\bar{a}_3c_3 \vee \bar{a}_1\bar{a}_2a_3c_6 \vee \bar{a}_1a_2\bar{a}_3c_9 \vee \bar{a}_1a_2a_3c_{12} \vee (a_1\bar{a}_2\bar{a}_3c_{15} \vee a_1\bar{a}_2a_3c_{15} \vee a_1a_2\bar{a}_3c_{15} \vee a_1a_2a_3c_{15}) = \\ &= \bar{a}_1\bar{a}_2\bar{a}_3c_3 \vee \bar{a}_1\bar{a}_2a_3c_6 \vee \bar{a}_1a_2\bar{a}_3c_9 \vee \bar{a}_1a_2a_3c_{12} \vee a_1c_{15}. \end{aligned}$$

Так как $c_3 = 0$, $c_6 = b_3$, то

$$y_2 = \bar{a}_1\bar{a}_2a_3b_3 \vee \bar{a}_1a_2\bar{a}_3c_9 \vee \bar{a}_1a_2a_3c_{12} \vee a_1c_{15}. \quad (16)$$

Частичные булевы функции из табл. 14 (кофакторы) реализуются следующими логическими уравнениями:

$$\begin{aligned} c_7 &= b_2\bar{b}_3; \quad c_8 = \bar{b}_2s_3; \quad c_9 = \bar{b}_2s_5 \vee b_2s_3; \quad c_{10} = b_2s_3; \quad c_{12} = \bar{b}_2b_3 \vee b_2\bar{b}_3; \\ c_{13} &= \bar{b}_2b_3; \quad c_{14} = b_2; \quad c_{15} = \bar{b}_2s_5 \vee b_2\bar{b}_3; \quad s_3 = \bar{b}_3b_1 \vee b_3; \quad s_5 = \bar{b}_3b_1. \end{aligned} \quad (17)$$

Для получения формул (17) использовалась программа [11, с. 201] минимизации BDD-представления частичных функций с учетом возможности их доопределения. В целом система полностью определенных булевых функций, реализующая умножитель по модулю 5, имеет вид

$$\begin{aligned} y_0 &= \bar{a}_1\bar{a}_2a_3b_1 \vee \bar{a}_1a_2\bar{a}_3c_7 \vee \bar{a}_1a_2a_3c_{10} \vee a_1c_{13}; \\ y_1 &= \bar{a}_1\bar{a}_2a_3b_2 \vee \bar{a}_1a_2c_8 \vee a_1c_{14}; \\ y_2 &= \bar{a}_1\bar{a}_2a_3b_3 \vee \bar{a}_1a_2\bar{a}_3c_9 \vee \bar{a}_1a_2a_3c_{12} \vee a_1c_{15}; \\ c_7 &= b_2\bar{b}_3; \quad c_8 = \bar{b}_2s_3; \quad c_9 = \bar{b}_2s_5 \vee b_2s_3; \quad c_{10} = b_2s_3; \quad c_{12} = \bar{b}_2b_3 \vee b_2\bar{b}_3; \\ c_{13} &= \bar{b}_2b_3; \quad c_{14} = b_2; \quad c_{15} = \bar{b}_2s_5 \vee b_2\bar{b}_3; \quad s_3 = \bar{b}_3b_1 \vee b_3; \quad s_5 = \bar{b}_3b_1. \end{aligned} \quad (18)$$

Заменяя ДНФ булевых функций y_0, y_1, y_2 в формулах (18) разложениями Шеннона, получим многоуровневое BDD-представление (рис. 10), которому соответствуют уравнения

$$\begin{aligned} y_0 &= \bar{a}_1z_1 \vee a_1c_{13}; \quad z_1 = \bar{a}_2z_2 \vee a_2z_3; \quad z_2 = a_3b_1; \quad z_3 = \bar{a}_3c_7 \vee a_3c_{10}; \quad y_1 = \bar{a}_1z_4 \vee a_1c_{14}; \\ z_4 &= \bar{a}_2z_5 \vee a_2c_8; \quad z_5 = a_3b_2; \quad y_2 = \bar{a}_1z_6 \vee a_1c_{15}; \quad z_6 = \bar{a}_2z_7 \vee a_2z_8; \quad z_7 = a_3b_3; \quad z_8 = \bar{a}_3c_9 \vee a_3c_{12}; \\ c_7 &= b_2\bar{b}_3; \quad c_8 = \bar{b}_2s_3; \quad c_9 = \bar{b}_2s_5 \vee b_2s_3; \quad c_{10} = b_2s_3; \quad c_{12} = \bar{b}_2b_3 \vee b_2\bar{b}_3; \quad c_{13} = \bar{b}_2b_3; \\ c_{14} &= b_2; \quad c_{15} = \bar{b}_2s_5 \vee b_2\bar{b}_3; \quad s_3 = \bar{b}_3b_1 \vee b_3; \quad s_5 = \bar{b}_3b_1. \end{aligned} \quad (19)$$

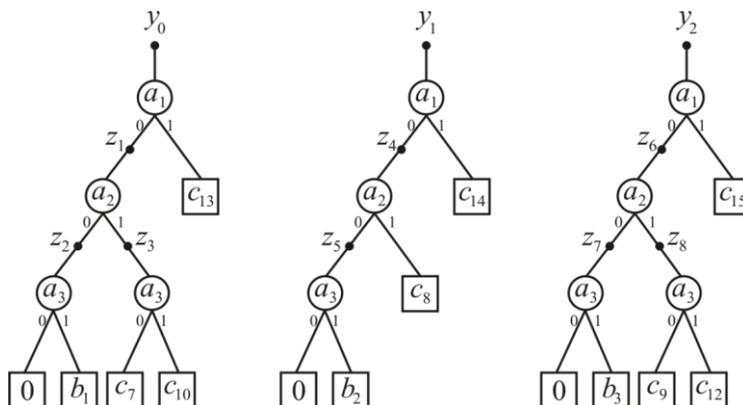


Рис. 10. BDD, заменяющие разложение Шеннона многозначной функции y по многозначной переменной a

Fig. 10. BDD replacing the Shannon expansion of a multi-valued function y by a multi-valued variable a

Синтез схем умножителя по модулю 5. Рассмотрим четыре варианта (способа) VHDL-описаний функций умножителя по модулю 5 и результаты синтеза логических схем по этим описаниям:

1. VHDL-описание полностью определенной многозначной функции, задающей область ее определенных значений.

Листинг 1. VHDL-описание полностью определенной многозначной функции умножителя по модулю 5

```
Library IEEE;
use IEEE.STD_LOGIC_1164.all;
entity mult_mod_5 is
  port ( a, b : in integer range 0 to 4;
        y : out integer range 0 to 4);
end mult_mod_5;
architecture beh of mult_mod_5 is
  begin
  process (a,b)
  begin
  if ((a = 0) or (b = 0)) then y <= 0; end if;
  if ((a = 1) and (b = 1)) then y <= 1; end if;
  if ((a = 1) and (b = 2)) then y <= 2; end if;
  if ((a = 1) and (b = 3)) then y <= 3; end if;
  if ((a = 1) and (b = 4)) then y <= 4; end if;
  if ((a = 2) and (b = 1)) then y <= 2; end if;
  if ((a = 2) and (b = 2)) then y <= 4; end if;
  if ((a = 2) and (b = 3)) then y <= 1; end if;
  if ((a = 2) and (b = 4)) then y <= 3; end if;
  if ((a = 3) and (b = 1)) then y <= 3; end if;
  if ((a = 3) and (b = 2)) then y <= 1; end if;
  if ((a = 3) and (b = 3)) then y <= 4; end if;
  if ((a = 3) and (b = 4)) then y <= 2; end if;
  if ((a = 4) and (b = 1)) then y <= 4; end if;
  if ((a = 4) and (b = 2)) then y <= 3; end if;
  if ((a = 4) and (b = 3)) then y <= 2; end if;
  if ((a = 4) and (b = 4)) then y <= 1; end if;
  end process;
end architecture beh;
```

2. VHDL-описание системы частичных функций, кодирование чисел стандартное.

Табл. 15 и листинг 2 (VHDL-описание) задают одну и ту же систему частичных булевых функций, реализующих умножитель по модулю 5.

Таблица 15

Умножитель по модулю 5 (система частичных булевых функций)

Table 15

The multiplier modulo 5 is a system of partial Boolean functions

Область определенных значений системы частичных булевых функций <i>The domain of the defined values of a system of partial Boolean functions</i>		Область неопределенных значений системы частичных булевых функций <i>The domain of undefined values of a system of partial Boolean functions</i>	
$a_1 a_2 a_3 b_1 b_2 b_3$	$y_0 y_1 y_2$	$a_1 a_2 a_3 b_1 b_2 b_3$	$y_0 y_1 y_2$
000 000	000	000 101	- - -
000 001	000	000 110	- - -
000 010	000	000 111	- - -
000 011	000	001 101	- - -
000 100	000	001 110	- - -
001 000	000	001 111	- - -
001 001	001	010 101	- - -
001 010	010	010 110	- - -
001 011	011	010 111	- - -
001 100	100	011 101	- - -
010 000	000	011 110	- - -

Окончание табл. 15
End of table 15

Область определенных значений системы частичных булевых функций <i>The domain of the defined values of a system of partial Boolean functions</i>					Область неопределенных значений системы частичных булевых функций <i>The domain of undefined values of a system of partial Boolean functions</i>			
a_1	a_2	a_3	b_1	b_2 b_3	y_0	y_1	y_2	
0	1	0	0	0	0	1	0	-
0	1	0	0	1	1	0	0	-
0	1	0	0	1	0	1	0	-
0	1	0	1	0	0	1	1	-
0	1	1	0	0	1	0	0	-
0	1	1	0	0	1	1	0	-
0	1	1	0	1	0	0	1	-
0	1	1	0	1	0	1	1	-
0	1	1	1	0	0	0	0	-
0	1	1	1	0	0	1	0	-
0	1	1	1	0	1	0	0	-
0	1	1	1	1	0	0	1	-
0	1	1	1	1	0	1	0	-
0	1	1	1	1	1	0	0	-
0	1	1	1	1	1	0	1	-
0	1	1	1	1	1	1	0	-
1	0	0	0	0	0	0	0	-
1	0	0	0	1	0	0	0	-
1	0	0	1	0	0	0	0	-
1	0	0	1	0	1	0	0	-
1	0	0	1	1	0	0	0	-
1	0	0	1	1	0	1	0	-
1	0	0	1	1	1	0	0	-
1	0	0	1	1	1	0	1	-
1	0	0	1	1	1	1	0	-
1	0	1	0	0	0	0	0	-
1	0	1	0	0	0	1	0	-
1	0	1	0	0	1	1	0	-
1	0	1	0	1	0	0	0	-
1	0	1	0	1	0	1	0	-
1	0	1	0	1	1	0	0	-
1	0	1	0	1	1	0	1	-
1	0	1	0	1	1	1	0	-
1	0	1	1	0	0	0	0	-
1	0	1	1	0	0	1	0	-
1	0	1	1	0	1	0	0	-
1	0	1	1	0	1	1	0	-
1	0	1	1	1	0	0	0	-
1	0	1	1	1	0	1	0	-
1	0	1	1	1	1	0	0	-
1	0	1	1	1	1	0	1	-
1	0	1	1	1	1	1	0	-
1	1	0	0	0	0	0	0	-
1	1	0	0	0	0	1	0	-
1	1	0	0	0	1	1	0	-
1	1	0	0	1	0	0	0	-
1	1	0	0	1	0	1	0	-
1	1	0	0	1	1	0	0	-
1	1	0	0	1	1	0	1	-
1	1	0	0	1	1	1	0	-
1	1	0	1	0	0	0	0	-
1	1	0	1	0	0	1	0	-
1	1	0	1	0	1	0	0	-
1	1	0	1	0	1	1	0	-
1	1	0	1	1	0	0	0	-
1	1	0	1	1	0	1	0	-
1	1	0	1	1	1	0	0	-
1	1	0	1	1	1	0	1	-
1	1	0	1	1	1	1	0	-
1	1	1	0	0	0	0	0	-
1	1	1	0	0	0	1	0	-
1	1	1	0	0	1	0	0	-
1	1	1	0	0	1	1	0	-
1	1	1	0	1	0	0	0	-
1	1	1	0	1	0	1	0	-
1	1	1	0	1	1	0	0	-
1	1	1	0	1	1	0	1	-
1	1	1	0	1	1	1	0	-
1	1	1	1	0	0	0	0	-
1	1	1	1	0	0	1	0	-
1	1	1	1	0	1	0	0	-
1	1	1	1	0	1	1	0	-
1	1	1	1	1	0	0	0	-
1	1	1	1	1	0	1	0	-
1	1	1	1	1	1	0	0	-
1	1	1	1	1	1	0	1	-
1	1	1	1	1	1	1	0	-

В представленном ниже VHDL-описании функций умножителя по модулю 5 вектор $y = (y_0, y_1, y_2)$ задает компонентные частичные булевы функции, а вектор $x = (a_1, a_2, a_3, b_1, b_2, b_3)$ – это вектор входных переменных.

Листинг 2. VHDL-описание системы частичных булевых функций, задающей функционирование умножителя по модулю 5

```

Library IEEE;
use IEEE.STD_LOGIC_1164.all;
entity mult_5_mod is
  port ( x : in std_logic_vector (1 to 6);
        y : out std_logic_vector (1 to 3));
end;
architecture BEHAVIOR of mult_5_mod is
begin
y <= "000" when x = "000000" else
     "000" when x = "000001" else
     "000" when x = "000010" else
     "000" when x = "000011" else
     "000" when x = "000100" else
     "000" when x = "001000" else
     "001" when x = "001001" else
     "010" when x = "001010" else
     "011" when x = "001011" else
     "100" when x = "001100" else
     "000" when x = "010000" else
     "010" when x = "010001" else
     "100" when x = "010010" else

```

```

"001" when x = "010011" else
"011" when x = "010100" else
"000" when x = "011000" else
"011" when x = "011001" else
"001" when x = "011010" else
"100" when x = "011011" else
"010" when x = "011100" else
"000" when x = "100000" else
"100" when x = "100001" else
"011" when x = "100010" else
"010" when x = "100011" else
"001" when x = "100100" else
"---";
end BEHAVIOR;

```

3. BDDI-оптимизация системы полностью определенных булевых функций.

Под BDDI (от англ. Binary Decision Diagram with Inverse cofactors) понимается ориентированный бесконтурный граф, задающий последовательные разложения Шеннона системы булевых функций $f(x_1, \dots, x_n)$ по всем ее переменным x_1, \dots, x_n при заданном порядке (перестановке) переменных, по которым проводятся разложения, при условии нахождения пар взаимно инверсных кофакторов [18]. Построение BDDI осуществлялось программой BDD_Builder [18] при «нулевом доопределении» исходной системы частичных функций, т. е. оптимизировалась система полностью определенных функций, получающаяся заменой в табл. 15 всех неопределенных значений частичных булевых функций нулями.

4. Оптимизация MDD и переход к графам BDD, которым соответствуют уравнения (19).

Результаты синтеза умножителя по модулю 5 приведены в табл. 16.

Таблица 16
Параметры логических схем, реализующих умножитель по модулю 5

Table 16
Parameters of logic circuits implementing a multiplier modulo 5

Вариант <i>Option</i>	Способ оптимизации, функциональное описание схемы <i>Optimization method, functional description of the circuit</i>	Площадь схемы, у. е. <i>Area of the circuit, c. u.</i>	Задержка схемы, нс <i>Circuit delay, ns</i>
1	Оптимизации нет, задание исходной многозначной полностью определенной функции в листинге 1	22 281	5,87
2	Оптимизации нет, неоптимизированное представление системы частичных функций в листинге 2	9 012	3,91
3	BDDI-оптимизация [18], исходное задание для BDDI-оптимизации в табл. 15	10 602	3,20
4	Предлагаемый способ: оптимизация MDD и переход к BDD, задание функций схемы уравнениями (19)	*8 554	*2,03

Четвертый вариант описания умножителя по модулю 5 (предлагаемый способ) позволил получить схему наименьшей площади и с наименьшей задержкой. Эти решения помечены в табл. 16 символом *.

Анализируя результаты синтеза, можно установить, что модели описания функционирования умножителя, использующие область неопределенных значений (второй и четвертый варианты), позволяют получать логические схемы меньшей площади и меньшей задержки по сравнению со способами описаний, базирующимися на описании только области определенных значений многозначной функции умножителя (первым вариантом) либо булевых полностью определенных функций (третьим вариантом).

Заключение. Предложенный способ рационального доопределения частичных многозначных диаграмм решений в процессе их построения позволяет находить совместимые кофакторы, которым после булева кодирования соответствуют подсистемы булевых функций. Это дает возможность сокращать число вершин MDD (кофакторов), что приводит к более экономичным по площади логическим схемам. Наряду с булевой минимизацией в классе ДНФ функциональных описаний двухоперандных модулярных устройств [2, 3] предложенный в статье способ доопределения MDD может применяться для оптимизации логических представлений последовательностей арифметических операций, операнды которых имеют ограниченные диапазоны значений, и для оптимизации многооперандных модулярных устройств, особенно в тех случаях, когда соответствующие им многозначные функции имеют большие области неопределенных значений.

Список использованных источников

1. Модулярные параллельные вычислительные структуры нейропроцессорных систем / Н. И. Червяков [и др.]. – М. : Физматлит, 2003. – 288 с.
2. Особенности проектирования модулярных умножителей с помощью современных САПР / Р. А. Соловьев [и др.]. // Проблемы разработки перспективных микро- и нанoeлектронных систем (МЭС). – 2016. – № 1. – С. 249–254.
3. Америкбаев, В. М. Реализация библиотеки модульных арифметических операций на основе алгоритмов минимизации логических функций / В. М. Америкбаев, Р. А. Соловьев, Д. В. Тельпухов // Изв. ЮФУ. Техн. науки. – 2013. – № 7. – С. 221–225.
4. Multi-Valued Decision Diagrams for Logic Synthesis and Verification / T. Kam [et al.]. – Berkeley : College of Engineering University of California, 1996. – 39 p.
5. Bryant, R. E. Graph-based algorithms for Boolean function manipulation / R. E. Bryant // IEEE Transactions on Computers. – 1986. – Vol. 35, no. 8. – P. 677–691.
6. Drechsler, R. Binary Decision Diagrams: Theory and Implementation / R. Drechsler, B. Becker. – Springer, 1998. – 210 p.
7. Ebendt, R. Advanced BDD Optimization / R. Ebendt, G. Fey, R. Drechsler. – Springer, 2005. – 222 p.
8. Bryant, R. E. Ordered binary decision diagrams / R. E. Bryant, C. Meinel // Logic Synthesis and Verification / eds.: S. Hassoun, T. Sasao, R. K. Brayton. – Kluwer Academic Publishers, 2002. – P. 285–307.
9. Meinel, C. Algorithms and Data Structures in VLSI Design: OBDD – Foundations and Applications / C. Meinel, T. Theobald. – Berlin, Heidelberg : Springer-Verlag, 1998. – 267 p.
10. Кнут, Д. Э. Искусство программирования. Т. 4, А. Комбинаторные алгоритмы. Ч. 1 : пер. с англ. / Д. Э. Кнут. – М. : Вильямс, 2013. – 960 с.
11. Бибило, П. Н. Применение диаграмм двоичного выбора при синтезе логических схем / П. Н. Бибило. – Минск : Беларус. навука, 2014. – 231 с.
12. Logic Minimization Algorithm for VLSI Synthesis / K. R. Brayton [et al.]. – Boston, Kluwer Academic Publishers, 1984. – 193 p.
13. Торопов, Н. Р. Минимизация систем булевых функций в классе ДНФ / Н. Р. Торопов // Логическое проектирование : сб. науч. тр. – Минск : Ин-т техн. кибернетики НАН Беларуси, 1999. – Вып. 4. – С. 4–19.
14. Брейтон, Р. К. Синтез многоуровневых комбинационных логических схем / Р. К. Брейтон, Г. Д. Хэчтел, А. Л. Санджованни-Винченелли // ТИИЭР. – 1990. – Т. 78, № 2. – С. 38–83.
15. Бибило, П. Н. Оценка энергопотребления логических КМОП-схем по их переключательной активности / П. Н. Бибило, Н. А. Кириенко // Микроэлектроника. – 2012. – № 1. – С. 65–77.
16. Бибило, П. Н. Системы проектирования интегральных схем на основе языка VHDL. StateCAD, ModelSim, LeonardoSpectrum / П. Н. Бибило. – М. : СОЛОН-Пресс, 2005. – 384 с.
17. Закревский, А. Д. Логический синтез каскадных схем / А. Д. Закревский. – М. : Наука, 1981. – 416 с.
18. Бибило, П. Н. Использование полиномов Жегалкина при минимизации многоуровневых представлений систем булевых функций на основе разложения Шеннона / П. Н. Бибило, Ю. Ю. Ланкевич // Программная инженерия. – 2017. – № 8. – С. 369–384.
19. Бибило П. Н. Минимизация BDDI-представлений систем не полностью определенных булевых функций / П. Н. Бибило // Программная инженерия. – 2020. – Т. 11, № 3. – С. 152–168.
20. Ashenden, P. J. VHDL-2008. Just the New Stuff / P. J. Ashenden, J. Lewis. – Burlington : Morgan Kaufman Publishers, 2008. – 909 p.

References

1. Chervyakov N. I., Sakhnyuk P. A., Shaposhnikov A. V., Ryadnov S. A. *Modulyarnye parallel'nye vychislitel'nye struktury neyroprotsessornykh system. Modular Parallel Computing Structures of Neuroprocessor Systems*. Moscow, Fizmatlit, 2003, 288 p. (In Russ.).
2. Solov'ev R. A., Tel'puhov D. V., Balaka E. S., Ruhlov V. S., Mihmel' A. S. *Design features of modular multipliers using modern CAD. Problemy razrabotki perspektivnykh mikro- i nanoelektronnykh sistem (MES) [Problems of Developing Promising Micro- and Nanoelectronic Systems (MES)]*, 2016, no. 1, pp. 249–254 (In Russ.).
3. Amerbaev V. M., Solov'ev R. A., Tel'puhov D. V. *Implementation of a library of modular arithmetic operations based on algorithms for minimizing logical functions. Izvestija Juzhnogo federal'nogo universiteta. Tehnicheskie nauki [Proceedings of the Southern Federal University. Technical Science]*, 2013, no. 7, pp. 221–225 (In Russ.).
4. Kam T., Villa T., Brayton R. K., Sangiovanni-Vincentelli A. L. *Multi-Valued Decision Diagrams for Logic Synthesis and Verification*. Berkeley, College of Engineering University of California, 1996, 39 p.
5. Bryant R. E. Graph-based algorithms for Boolean function manipulation. *IEEE Transactions on Computers*, 1986, vol. 35, no. 8, pp. 677–691.
6. Drechsler R. *Binary Decision Diagrams: Theory and Implementation*. Springer, 1998, 210 p.
7. Ebendt R., Fey G., Drechsler R. *Advanced BDD Optimization*. Springer, 2005, 222 p.
8. Bryant R. E., Meinel C. Ordered binary decision diagrams. *Logic Synthesis and Verification*. In S. Hassoun, T. Sasao, R. K. Brayton (eds.). Kluwer Academic Publishers, 2002, pp. 285–307.
9. Meinel C., Theobald T. *Algorithms and Data Structures in VLSI Design: OBDD – Foundations and Applications*. Berlin, Heidelberg, Springer-Verlag, 1998, 267 p.
10. Knuth D. E. *The Art of Computer Programming, Volume 4A: Combinatorial Algorithms, Part 1*. Addison-Wesley Professional, 2011, 912 p.
11. Bibilo P. N. *Primenenie diagram dvoichnogo vybora pri sinteze logicheskikh shem. Application of Binary Decision Diagrams in the Synthesis of Logic Circuits*. Minsk, Belaruskaja navuka, 2014, 231 p. (In Russ.).
12. Brayton R. K., Hachtel G. D., McMullen C., Sangiovanni-Vincentelli A. L. *Logic Minimization Algorithm for VLSI Synthesis*. Boston, Kluwer Academic Publishers, 1984, 193 p.
13. Toropov N. R. *Minimization of systems of Boolean functions in the class DNF. Logicheskoe proektirovanie : sbornik nauchnykh trudov [Logical Design: Collection of Scientific Papers]*. Minsk, Institut tehnicheckoj kibernetiki Nacional'noj akademii nauk Belarusi, 1999, iss. 4, pp. 4–19 (In Russ.).
14. Brayton R. K., Hachtel G. D., Sangiovanni-Vincentelli A. L. *Synthesis of multi-level combinational logic circuits. Trudy Institute inzhenerov po jelektronike i radiotekhnike [Proceedings of the Institute of Electronics and Radio Engineering]*, 1990, vol. 78, no. 2, pp. 38–83 (In Russ.).
15. Bibilo P. N., Kirienko N. A. *Estimating energy consumption in logical CMOS circuits based on their switching activity. Mikroelektronika [Microelectronics]*, 2012, no. 1, pp. 65–77 (In Russ.).
16. Bibilo P. N. *Cistemy proektirovaniya integral'nykh skhem na osnove yazyka VHDL. StateCAD, ModelSim, LeonardoSpectrum. Integrated Circuit Design Systems Based on the VHDL Language. StateCAD, ModelSim, LeonardoSpectrum*. Moscow, SOLON-Press, 2005, 384 p. (In Russ.).
17. Zakrevskij A. D. *Logicheskij sintez kaskadnykh skhem. Logical Synthesis of Cascading Circuit*. Moscow, Nauka, 1981, 416 p. (In Russ.).
18. Bibilo P. N., Lankevich Yu. Yu. *The use of Zhegalkin polynomials in minimizing multi-level representations of systems of Boolean functions based on the Shannon expansion. Programmnyaya inzheneriya [Software Engineering]*, 2017, no. 8, pp. 369–384 (In Russ.).
19. Bibilo P. N. *Minimization of Binary Decision Diagrams for Systems of Incompletely Defined Boolean Functions using inverse cofactors. Programmnyaya inzheneriya [Software Engineering]*, 2020, vol. 11, no. 3, pp. 152–168 (In Russ.).
20. Ashenden P. J., Lewis J. *VHDL-2008. Just the New Stuff*. Burlington, Morgan Kaufman Publishers, 2008, 909 p.

Информация об авторе

Бибилу Петр Николаевич, доктор технических наук, профессор, Объединенный институт проблем информатики Национальной академии наук Беларуси.
E-mail: bibilo@newman.bas-net.by

Information about the author

Petr N. Bibilo, D. Sc. (Eng.), Prof., The United Institute of Informatics Problems of the National Academy of Sciences of Belarus.
E-mail: bibilo@newman.bas-net.by



UDC 004.272.2
<https://doi.org/10.37661/1816-0301-2023-20-2-65-84>

Original Paper
Оригинальная статья

Generation of shortest path search dataflow networks of actors for parallel multi-core implementation

Anatoly A. Prihozhy

*Belarusian National Technical University,
av. Nezavisimosty, 65, Minsk, 220013, Belarus
E-mail: prihozhy@yahoo.com*

Abstract

Objectives. The problem of parallelizing computations on multicore systems is considered. On the Floyd – Warshall blocked algorithm of shortest paths search in dense graphs of large size, two types of parallelism are compared: fork-join and network dataflow. Using the CAL programming language, a method of developing actors and an algorithm of generating parallel dataflow networks are proposed. The objective is to improve performance of parallel implementations of algorithms which have the property of partial order of computations on multicore processors.

Methods. Methods of graph theory, algorithm theory, parallelization theory and formal language theory are used.

Results. Claims about the possibility of reordering calculations in the blocked Floyd – Warshall algorithm are proved, which make it possible to achieve a greater load of cores during algorithm execution. Based on the claims, a method of constructing actors in the CAL language is developed and an algorithm for automatic generation of dataflow CAL networks for various configurations of block matrices describing the lengths of the shortest paths is proposed. It is proved that the networks have the properties of rate consistency, boundedness, and liveness. In actors running in parallel, the order of execution of actions with asynchronous behavior can change dynamically, resulting in efficient use of caches and increased core load. To implement the new features of actors, networks and the method of their generation, a tunable multi-threaded CAL engine has been developed that implements a static dataflow model of computation with bounded sizes of buffers. From the experimental results obtained on four types of multi-core processors it follows that there is an optimal size of the network matrix of actors for which the performance is maximum, and the size depends on the number of cores and the size of graph.

Conclusion. It has been shown that dataflow networks of actors are an effective means to parallelize computationally intensive algorithms that describe a partial order of computations over decomposed data. The results obtained on the blocked algorithm of shortest paths search prove that the parallelism of dataflow networks gives higher performance of software implementations on multicore processors in comparison with the fork-join parallelism of OpenMP.

Keywords: dataflow, network of actors, CAL language, shortest paths, blocked algorithm, multi-core system, speedup

For citation. Prihozhy A. A. *Generation of shortest path search dataflow networks of actors for parallel multi-core implementation*. *Informatika [Informatics]*, 2023, vol. 20, no. 2, pp. 65–84.

<https://doi.org/10.37661/1816-0301-2023-20-2-65-84>

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

Received | Поступила в редакцию 20.02.2023
Accepted | Подписана в печать 21.03.2023
Published | Опубликована 29.06.2023

Генерация потоковых сетей акторов поиска кратчайших путей для параллельной многоядерной реализации

А. А. Прихожий

*Белорусский национальный технический университет,
пр. Независимости, 65, Минск, 220013, Беларусь
E-mail: prihozhy@yahoo.com*

Аннотация

Цели. Рассматривается задача распараллеливания вычислений на многоядерных системах. Посредством блочного алгоритма Флойда – Уоршалла поиска кратчайших путей на плотных графах большого размера сравниваются два вида параллелизма: разветвление/слияние и сетевой потоковый. С использованием языка программирования CAL разрабатываются методы построения акторов потока данных и алгоритм генерации параллельных сетей акторов. Целью работы является повышение производительности параллельных сетевых реализаций алгоритмов, обладающих свойством частичного порядка вычислений, на многоядерных процессорах.

Методы. Используются методы теории графов, теории алгоритмов, теории распараллеливания, теории формальных языков.

Результаты. Доказаны утверждения о возможности переупорядочивания вычислений в блочном алгоритме Флойда – Уоршалла, способствующие повышению загрузки ядер при реализации алгоритма. На основе утверждений разработан метод построения акторов на языке CAL и предложен алгоритм автоматической генерации CAL-сетей потока данных для различных конфигураций матриц блоков, описывающих длины кратчайших путей. Доказано, что сети обладают свойствами согласованности, ограниченности и живучести. В акторах, работающих параллельно, порядок выполнения действий с асинхронным поведением может динамически меняться, что приводит к эффективному использованию кэш и увеличению загрузки ядер. Для реализации новых возможностей акторов, сетей и метода их генерации разработан настраиваемый многопоточный CAL-движок, реализующий статическую модель потоковых вычислений с ограниченными размерами буферов. Из экспериментальных результатов, полученных на четырех типах многоядерных процессоров, следует, что существует оптимальный размер сетевой матрицы акторов, для которого производительность максимальна, и этот размер зависит от размера графа и количества ядер.

Заключение. Показано, что сети акторов потока данных являются эффективным средством распараллеливания алгоритмов с высокой вычислительной нагрузкой, описывающих частичный порядок вычислений над данными, декомпозированными на части. Результаты, полученные на блочном алгоритме поиска кратчайших путей, показали, что параллелизм сетей потока данных дает более высокую производительность программных реализаций на многоядерных процессорах по сравнению с параллелизмом разветвления/слияния стандарта OpenMP.

Ключевые слова: поток данных, сеть акторов, язык CAL, кратчайшие пути, блочный алгоритм, многоядерная система, ускорение

Для цитирования. Прихожий, А. А. Генерация потоковых сетей акторов поиска кратчайших путей для параллельной многоядерной реализации / А. А. Прихожий // Информатика. – 2023. – Т. 20, № 2. – С. 65–84. <https://doi.org/10.37661/1816-0301-2023-20-2-65-84>

Конфликт интересов. Автор заявляет об отсутствии конфликта интересов.

Introduction. The problem of finding the shortest and longest paths in weighted graphs [1–5] has many practical applications: computer games, signal processing, city and network traffic, video compression, microelectronics, optimization of computer systems and networks, task scheduling, bioinformatics, and many others. It is formulated in different settings and, therefore, is solved by algorithms of different computational complexity, from polynomial to exponential. In this paper we consider the all-pairs shortest path problem and the blocked Floyd – Warshall algorithm (*BFW*) [6–10] which decomposes the dense graph into subgraphs, has cubic computational complexity and is a basic one for the problem. *BFW* helps to 1) localize the data accesses within blocks and thereby reduce the

data miss count in the processor hierarchical memory; 2) organize the parallel computation of blocks on a multi-processor system. At the same time, *BFW* has drawbacks of recalculating all blocks in every iteration of the loop along graph vertices and of parallelizing the block calculations in the fork-join style, thus providing insufficient load of processors. Usually, *BFW* is implemented with OpenMP. Although the *BFW*'s complexity is polynomial, to handle large graphs *BFW* requires huge computational resources and much runtime even on multiprocessor systems. Scientific research was done, and works were published which improve the properties of *BFW*. Thus, [11] extended the homogeneous blocked Floyd – Warshall algorithm to a heterogeneous one recognizing four types of blocks and speeding up their computation. In [12], a threaded block-parallel algorithm is proposed which uses a cooperative scheduler of threads incorporated in the operating system. Work [13] aims for selecting the optimal size of block. Methods of efficient utilization of hierarchical caches are proposed in [14, 15]. A generalization of blocked Floyd – Warshall algorithm is proposed in [16] aiming at reducing the usage of slow global memory in implementations on GPU.

Parallel dataflow networks [17–20] have not been used yet for the realization of *BFW*; this topic is the subject of the paper. The dataflow actor concept aims at modelling of distributed knowledge-based algorithms. Actors match to the heterogeneous and concurrent dataflow nature of various kinds of embedded systems. The CAL dataflow language is suitable to model applications from cryptography, multimedia processing, network processing, control systems, reconfigurable systems, power optimization, monitoring of hardware and software, and others. Both hardware- and software-oriented CAL-compilers were developed. The concept of actors and principles of concurrency and asynchrony lie in the basis of CAL. Although CAL is a general-purpose actor-programming language, it was most successfully used in the MPEG standard known as reconfigurable video coding. The property of reconfigurability was introduced in CAL due to works [21–24]. In [21], the authors developed the multidimensional synchronous dataflow. In [22], the authors proposed the parameterized dataflow and used it for the reconfiguration of digital signal processing systems. Work [23] defined OpenDF as a dataflow toolset for reconfigurable hardware and multicore systems. The authors of [24] proposed the Boolean parametric data flow as a means for run-time reconfiguration of CAL programs. Since CAL aims for the creation of streaming applications, the authors of works [25–27] developed methods and tools for the synthesis and optimization of dataflow pipelines.

Although this paper develops and implements CAL-networks for parallel solving the all-pairs shortest path problem on multi-core systems, it shows the way of how actors and dataflow networks of actors can be created, generated, and implemented targeting other computationally heavy problems of large sizes with partial order of computations. The main contributions of the paper are:

1. It proves that in parallel *BFW* the block calculations can be moved across iterations of the loop along graph vertices and reordered, thus balancing the computational load among processors.
2. By means of simulation it is shown that the reordering of block calculations can increase the *BFW* speedup up to 25 %.
3. The approach has been developed which extends the reconfigurability principal and yields a method of automatic generation of CAL-actors and dataflow CAL-networks for various block-matrix configurations of shortest paths lengths.
4. Based on the C/C++ language, the generation tool and tunable multi-threaded CAL-engine are developed which create and implement the dataflow networks of actors on multi-core systems.
5. The computational experiments have shown that the optimal size of the block-matrix and CAL-network can be found which depends on the core count and graph size; the CAL-networks give the speedup that is up to 28 % higher than the number of cores and is higher than the speedup OpenMP gives.

CAL dataflow actor language. Many modern forms of computation are very well suited for data flow description and implementation. CAL is the high-level dataflow actor programming language [20–24] in which a program is defined as a network of actors that interact and communicate by sending and receiving data (tokens) along data lossless and order preserving communication channels. An actor is a computational entity that consists of input and output ports, state variables, actions, and a scheduler. Actors run in parallel. When an actor is fired, it consumes tokens from input ports, changes the internal state and produces tokens on output ports. The action is a piece of computation that an

actor performs during firing. An actor may contain any number of actions. When an actor is fired, it selects one of them based on the availability of input tokens and optionally based on conditions relating to the values of tokens and state variables. An action guard enables conditional action firing. A finite state machine (FSM) allows actions to be scheduled according to the current state of the actor and considering action priorities. CAL enables the description of different, but still actor-like, contexts, which have different kinds of objects (and types), different libraries, different primitive data objects and operators. The CAL model has the properties of strong encapsulation, explicit concurrency, and asynchrony (untimedness).

CAL as a domain-specific language provides useful abstractions for parallelizing computations and dataflow programming. It has been shown that the CAL dataflow networks offer a representation that can effectively support the tasks of parallelization and vectorization – thus providing a practical means of supporting multiprocessor systems and utilizing vector instructions. CAL has been used in a wide variety of applications and has been compiled to hardware and software implementations. It has been chosen by the ISO/IEC standardization organization in the MPEG standard called Reconfigurable Video Coding (RVC) (ISO/IEC 23001-4 and 23002-4).

The model of computation [17, 18] defines the semantics of the communication between the actors. It also defines which scheduling policies can be used to fire actors. There exists a variety of models of computation for CAL, which make trade-offs between expressiveness and analysability. The set of recognized dataflow models which are scheduled statically by compiler include Kahn process networks [17], synchronous dataflow networks (SDF) [18], parameterized synchronous dataflow (PSDF) [23], Boolean parametric dataflow (BPDF) [24], multidimensional synchronous dataflow (MDSDF) [22]. Other dataflow networks require dynamical scheduling, which induces a run-time overhead. The Kahn network is a group of deterministic sequential processes that communicate through unbounded FIFO channels. In SDF, the number of tokens read and written by each process is known ahead of time, and the channels have bounded FIFOs. SDF is divided into synchronous sub-networks connected by asynchronous links. PSDF supports dynamic reconfigurability and design reuse, but it does not allow the topology of the dataflow graph to change at runtime. BPDF allows restricted dynamic changes of the graph topology by disabling edges annotated with Boolean expressions.

To be scheduled statically, the dataflow network must have a basic iteration and have the properties of rate consistency, boundedness, and liveness [17–24]. The number of tokens consumed or produced at a given port at each firing is called the rate. The rate consistency of a dataflow network is checked by generating a system of balance equations, which must have a non-null solution for all possible values of parameters. The boundedness is guaranteed if the network returns to its initial state after each iteration. The network liveness is checked by finding a schedule for a basic iteration.

The CAL was first used on the Ptolemy II platform [20]. The complete OpenDF framework has been developed for simulating CAL networks and for generating hardware and software code [21]. The portable CAL interpreter used in the Moses project aimed for simulating a hierarchical network of actors. OpenDF is a compilation framework using a source-to-source compiler. Backends that generate VHDL/Verilog and C for integration with SystemC were developed.

A problem of automatic generation of dataflow networks. Nowadays, dynamic reconfigurable embedded systems [21, 22] are widely used, since they have the capability to modify their functionalities by adding or removing components, and by modifying interconnections among them. The basic idea behind these systems is to autonomously modify its functionalities according to the application's changes. Dynamic reconfiguration is the process of adding, deleting, or moving resources within the network configuration without deactivating the affected node. The models, architectures, and design methodologies of the reconfigurable systems have been developed. The PSDF approach [23] can dynamically reconfigure the behaviour of dataflow actors, edges, graphs, and subsystems by run-time modification of parameter values. It permits the parameter reconfiguration that does not change the subsystem interface behavior. BPDF [24] combines both the token-rate and topology reconfigurations, although, it does not reconfigure the topology significantly. The dataflow programming models are well-suited to program many-core streaming applications.

There are a variety of application problems where it is difficult or impossible to create a reconfigurable dataflow network; therefore, different dataflow networks must be generated depending on the problem parameters, problem size, and problem formulation. The networks can differ by the actors, input and output ports, actions, etc, and their quantity. In the paper, we consider such a problem, i.e., the all-pairs shortest path search in large graphs to be solved on a multi-processor system. By modifying the blocked Floyd – Warshall algorithm [6, 7], we create and generate dataflow parallel CAL-networks automatically and implement them efficiently on multi-core systems by means of creating a CAL-language-based multithreaded engine.

Block-parallel all-pairs shortest path algorithm. Let $G = (V, E)$ be a simple directed graph with real edge-weights consisting of a set V , $|V| = N$, of vertices and a set E of edges. Let W be the cost adjacency matrix for G . So $w_{i,i} = 0$, $1 \leq i \leq N$; $w_{i,j}$ is the cost (weight) of edge (i, j) if $(i, j) \in E$ and $w_{i,j} = \infty$ if $i \neq j$ and $(i, j) \notin E$. When G has no cycle with negative sum of weights, the dynamic programming Floyd – Warshall (FW) Algorithm 1 [1, 2] computes a series of distance matrices $D^0 \dots D^k \dots D^N$ such that $D^0 = W$ and each element $d_{i,j}^k$ of matrix D^k , $k = 1 \dots N$, is the length of the shortest path from i to j composed of the subset of vertices labelled 1 to k .

The authors of [6, 7, 11, 13] proposed a blocked version *BFW* of the Floyd – Warshall Algorithm 2. *BFW* divides set V of vertices into subsets $V_0 \dots V_{M-1}$ of size S and splits matrix D into blocks of size $S \times S$ each, creating a block-matrix $B[M \times M]$, where equality $M \cdot S = N$ holds. Algorithm 2 performs M iterations, each consisting of three phases: calculation of diagonal $D0$ block $B_{m,m}$ (accounts for paths inside the subgraph on subset V_m of vertices); calculation of $(M - 1)$ cross blocks $B_{v,m}$ of type C1 through block $B_{m,m}$ (accounts for paths from vertices of V_v to vertices of V_m); calculation of $(M - 1)$ cross blocks $B_{m,v}$ of type C2 through block $B_{m,m}$ (accounts for paths from vertices of V_m to vertices of V_v); calculation of $(M - 1)^2$ peripheral $P3$ blocks $B_{v,u}$ through blocks $B_{v,m}$ and $B_{m,u}$ (accounts for paths from vertices of V_v to vertices of V_u passing through vertices of V_m). In $B_{v,u}^m$, index m describes the block calculation level. Algorithm 3 (*BCA*) calculates all three types of blocks. In [7], the authors shown that *BFW* can be parallelized to *PBFW* due to all cross blocks can be calculated mutually in parallel as well as all peripheral blocks. Algorithm 2 describes the parallelism by means of OpenMP directives. In *BFW*, the blocks can be also calculated recursively [7].

Algorithm1: Floyd – Warshall *FW*

Input: A number N of graph vertices
Input: An edge cost matrix $W[N \times N]$
Output: Matrix D^N of distances
 $D^0 \leftarrow W$
for $k \leftarrow 0$ **to** $N - 1$ **do**
 for $i \leftarrow 0$ **to** $N - 1$ **do**
 for $j \leftarrow 0$ **to** $N - 1$ **do**
 $d_{i,j}^{k+1} \leftarrow \min(d_{i,j}^k, d_{i,k}^k + d_{k,j}^k)$
return D^N

Algorithm 3: Block calculation *BCA*

Input: A size S of block
Input: Blocks E, F and H
Output: Block B
For $k \leftarrow 0$ **to** $S - 1$ **do**
 For $i \leftarrow 0$ **to** $S - 1$ **do**
 for $j \leftarrow 0$ **to** $S - 1$ **do**
 $b_{i,j} \leftarrow \min(e_{i,j}, f_{i,k} + h_{k,j})$
return B

Algorithm 2: Block-parallel Floyd – Warshall *PBFW*

Input: A number N of graph vertices
Input: A matrix $W[N \times N]$ of graph edge weights
Input: A size S of block
Output: A blocked matrix $B^M[M \times M]$ of path distances
 $M \leftarrow N/S$ $B^0 \leftarrow W$
#pragma omp parallel
for $m \leftarrow 0$ **to** $M - 1$ **do**
 #pragma omp single
 $B_{m,m}^{m+1} \leftarrow \text{BCA}(B_{m,m}^m, B_{m,m}^m, B_{m,m}^m)$ // Diagonal D0
 for $v \leftarrow 0$ **to** $M - 1$ **do**
 if $v \neq m$ **then**
 #pragma omp task united
 $B_{v,m}^{m+1} \leftarrow \text{BCA}(B_{v,m}^m, B_{v,m}^m, B_{m,m}^{m+1})$ // Cross C1
 #pragma omp task united
 $B_{m,v}^{m+1} \leftarrow \text{BCA}(B_{m,v}^m, B_{m,v}^{m+1}, B_{m,v}^m)$ // Cross C2
 #pragma omp task wait
 for $v \leftarrow 0$ **to** $M - 1$ **do**
 if $v \neq m$ **then**
 for $u \leftarrow 0$ **to** $M - 1$ **do**
 if $u \neq m$ **then**
 #pragma omp task united
 $B_{v,u}^{m+1} \leftarrow \text{BCA}(B_{v,u}^m, B_{v,m}^{m+1}, B_{m,u}^{m+1})$ // Peripheral P3
 #pragma omp task wait
 return B^M

Fork-join parallelization potential. OpenMP parallelizes *PBFW* in the fork-join style. Since diagonal block *D0* is computed in series to all parallel cross blocks *C1* and *C2*, and all cross blocks are computed in series to all parallel blocks *P3*, (1) estimates the speedup the *PBFW* provides over *BFW* on *P* processors

$$speedup = M^2 / \left(1 + \lceil 2(M-1)/P \rceil + \lceil (M-1)^2/P \rceil \right) \quad (1)$$

The parallelization potential is quite non-uniform when considering the diagonal, cross and peripheral blocks. Let $P = 8$ and $M = 4$. Then, one step is needed for executing the diagonal block, where 1 processor is loaded, and 7 processors stand idle. One step is needed for executing 6 cross blocks, where 6 processors are loaded, and 2 processors stand idle. Two steps are needed for executing 9 peripheral blocks: in first step all 8 processors are loaded; in second step only 1 processor is loaded, and 7 others stand idle. As a result, according to (1) the speedup of *PBFW* is 4 instead of expected ideal 8.

The promising alternative to the fork-join is dataflow parallelism. In the paper we develop dataflow networks which have such a property that the calculations of peripheral blocks can be moved over iterations along m in Algorithm 2.

Reordering of block calculations. The authors of [7] proved the following assertion related to the *BFW* algorithm:

Claim 1. Suppose $d_{i,j}^{k+1}$, $k = 0 \dots N-1$, is computed as

$$d_{i,j}^{k+1} = \min(d_{i,j}^k, d_{i,k}^{k'} + d_{k,j}^{k''}) \quad (2)$$

for $k \leq k'$, $k'' \leq N$, then upon termination, the Floyd – Warshall algorithm correctly computes the all-pairs shortest paths.

In *BFW* and *PBFW*, the diagonal block $B_{m,m}^{m+1}$ of type *D0* is calculated through $B_{m,m}^m$ for which all $B_{m,m}^1 \dots B_{m,m}^{m-1}$ have been already calculated. The cross block $B_{v,m}^{m+1}$ of type *C1* is calculated through $B_{m,m}^{m+1}$ and $B_{v,m}^m$ for which all $B_{v,m}^1 \dots B_{v,m}^{m-1}$ have been already calculated. The same holds for the cross block $B_{m,v}^{m+1}$ of type *C2*. The peripheral block $B_{v,u}^{m+1}$ of type *P3* is calculated through $B_{v,u}^m$, $B_{v,m}^{m+1}$ and $B_{m,u}^{m+1}$. We formulate and prove Claim 2 which relaxes the requirement to blocks $B_{v,m}^{m+1}$ and $B_{m,u}^{m+1}$.

Claim 2. Suppose *P3*-type block $C_{v,u}^{m+1}$ (that is block $B_{v,u}^{m+1}$ calculated by a different algorithm *BFW'*), $m = 0 \dots M-1$, is computed as

$$C_{v,u}^{m+1} = BCA(C_{v,u}^m, C_{v,m}^{m'}, C_{m,u}^{m''}) \quad (3)$$

for $m+1 \leq m'$, $m'' \leq M$, then upon termination, *BFW'* correctly computes the all-pairs shortest paths.

Proof. In *BFW* and *BFW'*, blocks *D0*, *C1* and *C2* are calculated in the same manner. Let prove by induction that for *P3*-type blocks $C_{v,u}^{m+1}$ and $B_{v,u}^{m+1}$, $0 \leq m \leq M-1$ the following inequality holds:

$$C_{v,u}^{m+1} \leq B_{v,u}^{m+1}, \quad (4)$$

which means inequality $c_{i,j} \leq b_{i,j}$ for all pairs of matching elements of the corresponding blocks.

Base case. By definition we have $C_{v,u}^0 = B_{v,u}^0 = W_{v,u}$, therefore $C_{v,u}^0 \leq B_{v,u}^0$, $v, u = 0 \dots M-1$.

Induction step. Suppose the inequality as follows holds:

$$C_{v,u}^m \leq B_{v,u}^m, \quad v, u = 0 \dots M-1. \quad (5)$$

Then applying Algorithm 3 to the *BCA* call (3) we can conclude:

$$C_{v,u}^{m+1} \leq BCA(B_{v,u}^m, C_{v,m}^{m'}, C_{m,u}^{m''}) \quad (6)$$

$$\leq BCA(B_{v,u}^m, C_{v,m}^{m+1}, C_{m,u}^{m+1}) \quad (7)$$

$$\leq BCA(B_{v,u}^m, B_{v,m}^{m+1}, B_{m,u}^{m+1}) \leq B_{v,u}^{m+1} . \quad (8)$$

Inequality (6) is inferred from (3) and (5). Inequality (7) is inferred from Claim 1 applied S times to each element of block B calculated over blocks E, F and H in Algorithm 3 executed through the *BCA* call of (6):

$$b_{i,j}^{S \cdot (m+1)} = \min(e_{i,j}^{S \cdot m}, f_{i,k}^{S \cdot m'} + h_{k,j}^{S \cdot m''}) \leq \min(e_{i,j}^{S \cdot m}, f_{i,k}^{S \cdot (m+1)} + h_{k,j}^{S \cdot (m+1)}), \quad (9)$$

where $S \cdot m$ is the calculation level of elements of the block that is at level m of calculation. Inequality (8) that proves (4) is inferred from (7) considering inequality (5) that is used to prove the inequality $C_{v,m}^{m+1} \leq B_{v,m}^{m+1}$ of the type $C1$ blocks and to prove the inequality $C_{m,u}^{m+1} \leq B_{m,u}^{m+1}$ of the type $C2$ blocks of the *BFW'* and *BFW* algorithms. All these blocks are calculated over the diagonal blocks which meet (5).

On the other hand, since the traditional blocked Floyd – Warshall algorithm computes the shortest paths at termination and (3) computes the length of some paths, we have:

$$B_{v,u}^{m+1} \leq C_{v,u}^{m+1} . \quad (10)$$

It is derived from (4) and (10) that $B_{v,u}^M = C_{v,u}^M$, which completes the proof.

Claim 2 allows the delaying and reordering of peripheral blocks calculations. Such a reordering was used in the multi-threaded all-pairs shortest path algorithm [12] realized using a cooperative scheduler of threads.

Fig. 1, *a* depicts the estimated parallelization potential of the fork-join *PBFW* algorithm and the parallelization potential of an algorithm *PBFW''* obtained from the former one by reordering of block calculations and balancing the load of processors. To estimate the parallelization potential, a program in C++ was developed which implements *PBFW* and its modification *PBFW''*. Fig. 1, *b* shows that the reordering of block calculations can speed up the shortest paths parallel search up to 25 %. Moreover, the speedup can increase in case when the block execution time is variable. It should be noted that the gain of *PBFW''* against *PBFW* is being reduced with the growth of block count.

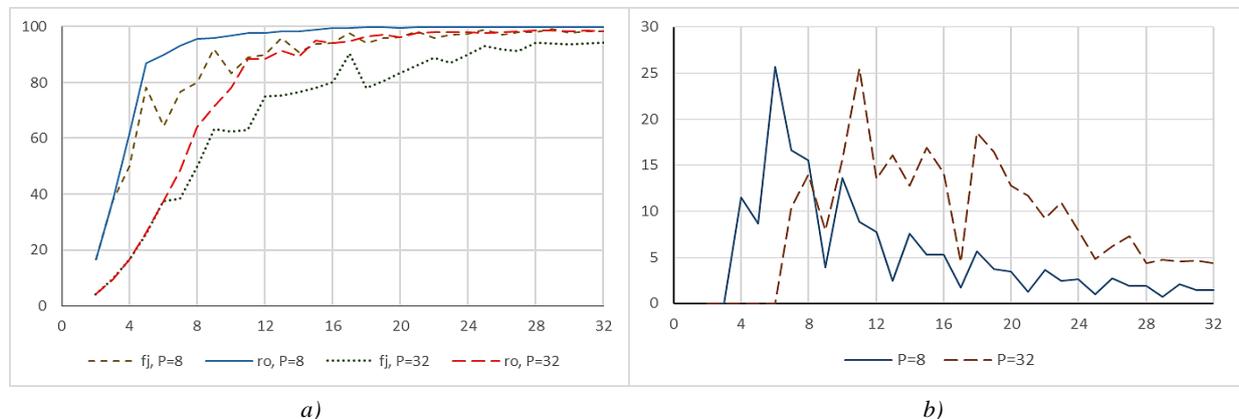


Fig. 1. Processors load in percent (vertical axis) provided by algorithms *PBFW* (fork-join parallelism *fj*) and *PBFW''* (reordering *ro* of calculations) for block-counts $M = 2$ to $M = 32$ (horizontal axis) on 8 and 32 processors P ; and lower bound of speedup *b*) in percent (vertical axis) *PBFW''* has against *PBFW* for block-counts $M = 2$ to $M = 32$ (horizontal axis) on 8 and 32 processors P

The *BFW* and *PBFW* are homogeneous in sense of calculating all blocks with the single *BCA* function. The authors of work [11] extended the algorithms to heterogeneous ones that calculate the blocks of types *D0*, *C1*, *C2* and *P3* using separate functions which operate faster than *BCA*.

Modelling block calculations by actors. A separate actor $A_{r,c}$ is put into accordance with each block $B_{r,c}$ of block-matrix B , which introduces a matrix $A[M \times M]$ of actors. The structure and behavior of actor $A_{r,c}$ depends on two factors: 1) the size $M \times M$ of matrix B ; 2) the location of $B_{r,c}$ in B (diagonal and non-diagonal blocks). Fig. 2 shows diagonal and non-diagonal actors in matrices $A[2 \times 2]$ and $A[3 \times 3]$. The size M influences the number of input ports and the total number of actions in the actor. The block location influences the structure and behavior of the actions incorporated in the actor. In the paper, we assume that the actors have access to two global variables M and B and assume that the actors' ports represent the block calculation levels but the blocks themselves. Actor $A_{r,c}$ may update block $B_{r,c}$ and may not other blocks in B .

In each actor $A_{r,c}$, the number of input ports equals $2 \cdot (M - 1)$ and the number of output ports equals two. The input ports describe calculation levels of other blocks located in row r and column c . Both output ports describe the calculation level of block $B_{r,c}$. The overall number of ports and the actor interface are the same for all actors of A , no matter the actor is diagonal or non-diagonal. Fig. 3 depicts the input and output ports of the actor in matrices $A[2 \times 2]$ and $A[3 \times 3]$.

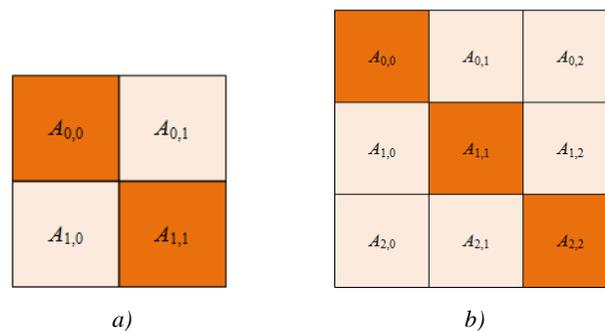


Fig. 2. Matrices a) $A[2 \times 2]$ and b) $A[3 \times 3]$ of diagonal and non-diagonal CAL actors

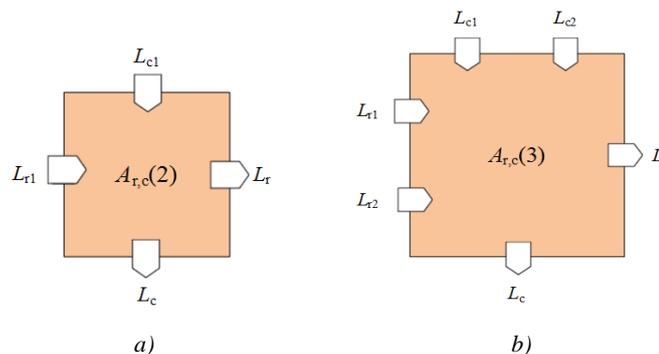


Fig. 3. Interface of actor $A_{r,c}(M)$ that models calculation of block $B_{r,c}$ in matrix $A[M \times M]$:
a) actor $A_{r,c}(2)$ and b) actor $A_{r,c}(3)$

In actors of $A[2 \times 2]$, input port L_{r1} receives the token produced by another actor located in row r , and L_{c1} receives the token produced by another actor in column c . Two output ports L_r and L_c of $A_{r,c}$ send tokens describing the calculation level of $B_{r,c}$ to other actors in row r and column c respectively. In matrix $A[3 \times 3]$, the actor interface has four input ports L_{r1} , L_{r2} , L_{c1} and L_{c2} since the number of blocks in row r and column c is increased to three.

The diagonal and non-diagonal actors have different internal structure and different actions. Algorithm 4 depicts the behavior of a diagonal actor *Block_D* that models the calculation of block $B_{0,0}$ in matrix $A[2 \times 2]$. Input ports L_{0_1} and L_{1_0} describe the calculation level of blocks $B_{0,1}$ and $B_{1,0}$

respectively. Output ports $Lrow$ and $Lcol$ describe the calculation level of block $B_{0,0}$. Variables Lev , Row and Col that describe the calculation level, row, and column of $B_{0,0}$ represent the actor internal state.

The actor contains three actions: diagonal, peripheral, and reset. Action $Dig0$ has no input, but two output tokens (both equal Lev) directed to ports $Lrow$ and $Lcol$. The guard condition requires $Lev = Row = Col$. The action body increments Lev and calls the BCA function to recalculate block $B_{0,0}$ over itself. According to the guard, the action is fired once. Action $Per1$ has three input and no output tokens. Tokens $L01$ and $L10$ arrive from ports L_0_1 and L_1_0 , and the third token is a constant $k = 1$. The guard condition requires $Lev < L01$ and $Lev < L10$. The action body increments Lev and calls the BCA function to recalculate block $B_{0,0}$ over blocks $B_{0,1}$ and $B_{1,0}$. The peripheral action is fired when the input tokens have arrived, and its guard is satisfied. The $Reset$ action sets Lev to 0. It is fired when the block has been recalculated M times.

Algorithm 4: Diagonal actor $Block_D$ for block $B_{0,0}$ in $A[2 \times 2]$

```

actor  $Block\_D$  ( $m$ ) int  $L\_0\_1$ , int  $L\_1\_0$  ==> int  $Lrow$ , int  $Lcol$ :
  int  $Lev := 0$ ; int  $Row := m$ ; int  $Col := m$ ;
   $Dig0$ : action ==>  $Lrow: [Lev], Lcol: [Lev]$  // D0
  guard  $Lev = Row$ 
  do
     $Lev := Lev + 1$ ;
     $BCA$  ( $B[Row, Col], B[Row, Col], B[Row, Col]$ );
  end
   $Per1$ : action  $L\_0\_1: [L01], L\_1\_0: [L10], 1:[k]$  ==> // P3
  guard  $L01 > Lev$  and  $L10 > Lev$ 
  do
     $Lev := Lev + 1$ ;
     $BCA$  ( $B[Row, Col], B[Row, k], B[k, Col]$ ); // P3
  end
   $Reset$ : action ==>
  guard  $Lev = M$  do  $Lev := 0$ ; end
end

```

Algorithm 5: Non-diagonal actor $Block_N$ for block $B_{0,1}$ in $A[2 \times 2]$

```

actor  $Block\_N$  ( $v, u$ ) int  $L\_0\_0$ , int  $L\_1\_1$  ==> int  $Lrow$ , int  $Lcol$ :
  int  $Lev := 0$ ; int  $Row := r$ ; int  $Col := c$ ;
   $Crs0$ : action  $L\_0\_0: [L00]$  ==>  $Lcol: [Lev]$  // C2
  guard  $Lev = L00 - 1$ 
  do
     $Lev := Lev + 1$ ;
     $BCA$  ( $B[Row, Col], B[Row, Row], B[Row, Col]$ );
  end
   $Crs1$ : action  $L\_1\_1: [L11]$  ==>  $Lrow: [Lev]$  // C1
  guard  $Lev = L11 - 1$ 
  do
     $Lev := Lev + 1$ ;
     $BCA$  ( $B[Row, Col], B[Row, Col], B[Col, Col]$ ); // P3
  end
   $Reset$ : action ==>
  guard  $Lev = M$  do  $Lev := 0$ ; end
end

```

Algorithm 5 describes a non-diagonal actor $Block_N$ that models the calculation of block $B_{0,1}$ in matrix $A[2 \times 2]$. Two input ports are L_0_0 and L_1_1 . The output ports and state variables are the same as in actor $Block_D$. The actor contains two actions. Action $Crs0$ has input token $L00$ arriving from port L_0_0 and has output token Lev sended to port $Lrow$. Its guard condition requires $Lev = L00 - 1$. The action body increments Lev and calls the BCA function to recalculate block $B_{0,1}$ over diagonal block $B_{0,0}$. $Crs0$ is fired when a token arrives at its input port and its guard condition evaluates to true. The behaviour of $Crs1$ is like those of $Crs0$ except $B_{0,1}$ is recalculated over $B_{1,1}$. Each of actions $Crs0$ and $Crs1$ is fired once.

Algorithms 6 and 7 describe the behavior of diagonal actor $A_{0,0}$ ($Block_D$) and non-diagonal actor $A_{0,1}$ ($Block_N$) that models the calculation of blocks $B_{0,0}$ and $B_{0,1}$ in matrix $A[3 \times 3]$. In $A_{0,0}$, one action is Dig and all others are Per . In $A_{0,1}$, two actions are Crs and all others are Per . Compared to actors of matrix $A[2 \times 2]$, the actors of $A[3 \times 3]$ have four input ports instead of two and have an additional peripheral action each. The output ports and state variables are the same. Unlike the actor of $A[2 \times 2]$, the $Block_D$ of $A[3 \times 3]$ of Algorithm 6 contains two peripheral actions that are competitive in firing and can be fired in arbitrary order. The actions' guards are redundant and removed since their firing is correctly managed by input tokens. Thus, block $B_{0,0}$ can have two firing sequences: 1) $B_{0,0}^0$, $B_{0,0}^1$ and $B_{0,0}^2$; 2) $B_{0,0}^0$, $B_{0,0}^2$ and $B_{0,0}^1$. CAL and its implementations require to resolve such competitions in advance by adding schedule and priorities. We do not follow this way since the order of firings of the actions does not influence the computation result. We perform a relaxation of CAL, omit the schedule and priorities in the actors and create our own multi-threaded implementation of CAL which resolves the competitions by means of an appropriate mechanism of synchronizing concurrent actions.

Moreover, we remove guards of actions $Crs0$ and $Crs1$ because the conditions they describe are fully satisfied by the conditions of arriving tokens on input ports.

Algorithm 6: Diagonal actor $Block_D$ for block $B_{0,0}$ in $A[3 \times 3]$

```

actor  $Block\_D$  ( $m$ ) int  $L_{-0\_1}$ , int  $L_{-0\_2}$ , int  $L_{-1\_0}$ , int  $L_{-2\_0}$  ==>
    int  $Lrow$ , int  $Lcol$ :
    int  $Lev := 0$ ; int  $Row := r$ ; int  $Col := c$ ;
     $Dig0$ : action ==>  $Lrow: [Lev]$ ,  $Lcol: [Lev]$  // D0
    guard  $Lev = Row$ 
    do
         $Lev := Lev + 1$ ;
         $BCA$  ( $D[Row, Col]$ ,  $D[Row, Col]$ ,  $D[Row, Col]$ );
    end
     $Per1$ : action  $L_{-0\_1}: [L01]$ ,  $L_{-1\_0}: [L10]$ ,  $1:[k]$  ==> // P3
    do
         $Lev := Lev + 1$ ;
         $BCA$  ( $D[Row, Col]$ ,  $D[Row, k]$ ,  $D[k, Col]$ );
    end
     $Per2$ : action  $L_{-0\_2}: [L02]$ ,  $L_{-2\_0}: [L20]$ ,  $2:[k]$  ==> // P3
    do
         $Lev := Lev + 1$ ;
         $BCA$  ( $D[Row, Col]$ ,  $D[Row, k]$ ,  $D[k, Col]$ );
    end
    Reset: action ==>
    guard  $Lev = M$  do  $Lev := 0$ ; end
end

```

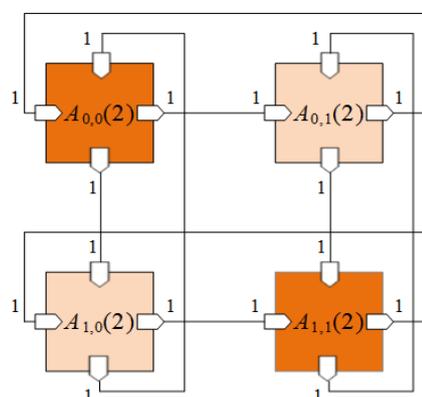
Algorithm 7: Non-diagonal actor $Block_N$ for block $B_{0,1}$ in $A[3 \times 3]$

```

actor  $Block\_N$  ( $v, u$ ) int  $L_{-0\_0}$ , int  $L_{-0\_2}$ , int  $L_{-1\_1}$ , int  $L_{-2\_1}$ 
    ==> int  $Lrow$ , int  $Lcol$ :
    int  $Lev := 0$ ; int  $Row := r$ ; int  $Col := c$ ;
     $Crs0$ : action  $L_{-0\_0}: [L00]$  ==>  $Lcol: [Lev]$  // C2
    do
         $Lev := Lev + 1$ ;
         $BCA$  ( $D[Row, Col]$ ,  $D[Row, Row]$ ,  $D[Row, Col]$ );
    end
     $Crs1$ : action  $L_{-1\_1}: [L11]$  ==>  $Lrow: [Lev]$  // C1
    do
         $Lev := Lev + 1$ ;
         $BCA$  ( $D[Row, Col]$ ,  $D[Row, Col]$ ,  $D[Col, Col]$ );
    end
     $Per2$ : action  $L_{-0\_2}: [L02]$ ,  $L_{-2\_1}: [L21]$ ,  $2:[k]$  ==> // P3
    do
         $Lev := Lev + 1$ ;
         $BCA$  ( $D[Row, Col]$ ,  $D[Row, k]$ ,  $D[k, Col]$ );
    end
    Reset: action ==>
    guard  $Lev = M$  do  $Lev := 0$ ; end
end

```

Parallel dataflow networks of actors for shortest paths search. Composing actors into a network, setting connections among their input and output ports, and allocating buffers to the connections establish a dataflow network. The shortest paths search network structure depends on M . Fig. 4 shows a graphical view and a CAL-code of the NW_{22} network constructed on matrix $A[2 \times 2]$.



a)

```

network  $NW_{22} () \Rightarrow$  :
entities
     $A00 = Block\_D$  ( $0,0$ );
     $A01 = Block\_N$  ( $0,1$ );
     $A10 = Block\_N$  ( $1,0$ );
     $A11 = Block\_D$  ( $1,1$ );
structure
     $A00.Lrow \rightarrow A01.L_{-0\_0}$ ;
     $A00.Lcol \rightarrow A10.L_{-1\_1}$ ;
     $A01.Lrow \rightarrow A00.L_{-0\_1}$ ;
     $A01.Lcol \rightarrow A11.L_{-1\_0}$ ;
     $A10.Lrow \rightarrow A11.L_{-0\_1}$ ;
     $A10.Lcol \rightarrow A00.L_{-1\_0}$ ;
     $A11.Lrow \rightarrow A10.L_{-0\_0}$ ;
     $A11.Lcol \rightarrow A01.L_{-1\_1}$ ;
end

```

b)

Fig. 4. Dataflow network NW_{22} constructed on matrix $A[2 \times 2]$ has 4 actors and 8 channels with buffers on them:
a) graphical view; b) CAL-code

NW_{22} consists of two diagonal and two non-diagonal actors, twelve actions and eight channels annotated with produced and consumed token rates. All rates are 1. Every action of every actor is fired once during the network operation. The diagonal Dig_m action of actor $A_{m,m}$ that is guarded with $Lev = m$ is fired once. It produces tokens, which are transferred to actions of cross non-diagonal actors $A_{m,v}$ and $A_{u,m}$ on row m and column m . Since the tokens are produced once, the cross actions Crs_v and Crs_u of the actors are fired once. The cross actions produce once and transfer tokens to actions of all actors outside the cross, therefore, all the peripheral actions can be fired once.

To prove the rate consistency of NW_{22} , we construct a combined balance equation for each channel $(A_v.p_i, A_u.p_j)$ connecting output port p_i of actor A_v with token rate $R(A_v.p_i)$ to input port p_j of actor A_u with token rate $R(A_u.p_j)$:

$$F(A_v.p_i) \cdot R(A_v.p_i) = F(A_u.p_j) \cdot R(A_u.p_j), \quad (11)$$

where $F(A_v.p_i)$ is the number of firings of A_v that produce tokens at p_i , and $F(A_u.p_j)$ is the number of firings of A_u that consume tokens at p_j . For NW_{22} , the system of balance equations (11) is described by (12)

1. $F(A_{0,0}.Lrow) \cdot 1 = F(A_{0,1}.L_0_0) \cdot 1$
2. $F(A_{0,0}.Lcol) \cdot 1 = F(A_{1,0}.L_1_1) \cdot 1$
3. $F(A_{0,1}.Lrow) \cdot 1 = F(A_{0,0}.L_0_1) \cdot 1$
4. $F(A_{0,1}.Lcol) \cdot 1 = F(A_{1,1}.L_1_0) \cdot 1$
5. $F(A_{1,0}.Lrow) \cdot 1 = F(A_{1,1}.L_0_1) \cdot 1$
6. $F(A_{1,0}.Lcol) \cdot 1 = F(A_{0,0}.L_1_0) \cdot 1$
7. $F(A_{1,1}.Lrow) \cdot 1 = F(A_{1,0}.L_0_0) \cdot 1$
8. $F(A_{1,1}.Lcol) \cdot 1 = F(A_{0,1}.L_1_1) \cdot 1$.

(12)

Equations 1 and 2 of (12) are satisfied because the single firing of action Dig_0 of actor $A_{0,0}$ (we denote $A_{0,0}.Dig_0$) produces at ports $Lrow$ and $Lcol$ the tokens consumed by single firing of $A_{0,1}.Crs_0$ and single firing $A_{1,0}.Crs_1$ respectively. Equation 4 is satisfied as $A_{0,1}.Crs_0$ is fired producing at $Lcol$ a token that is consumed by $A_{1,1}.Per_1$. Equation 5 is satisfied as $A_{1,0}.Crs_1$ is fired producing at $Lrow$ a token that is consumed by $A_{1,1}.Per_1$. Equations 7 and 8 are satisfied because the single firing of action Dig_0 of actor $A_{1,1}$ produces at ports $Lrow$ and $Lcol$ tokens consumed by firings $A_{1,0}.Crs_0$ and $A_{0,1}.Crs_1$ respectively. Equation 6 is satisfied by the firing of $A_{1,0}.Crs_0$ producing at $Lcol$ the token that is consumed by $A_{0,0}.Per_1$. Equation 3 is satisfied by the firing of $A_{0,1}.Crs_0$ producing at $Lcol$ the token that is consumed by $A_{0,0}.Per_1$.

The *Reset* action of all diagonal and non-diagonal actors sets Lev to initial state 0 after firing of all other actions. This guarantees the boundedness of FIFO buffers in NW_{22} . The following schedule proves the liveness of network NW_{22}

$$A_{0,0}.Dig_0, A_{0,1}.Crs_0, A_{1,0}.Crs_1, A_{1,1}.Per_1, A_{1,1}.Dig_0, A_{0,1}.Crs_1, A_{1,0}.Crs_0 \text{ and } A_{0,0}.Per_1.$$

The firing of actions of different actors can proceed in series and in parallel. Fig. 5 depicts a rate-consistent dataflow network constructed on matrix $A[3 \times 3]$, which obtains the properties of boundedness and liveness. Therefore, the network supports the synchronous dataflow model (SDF) of computation.

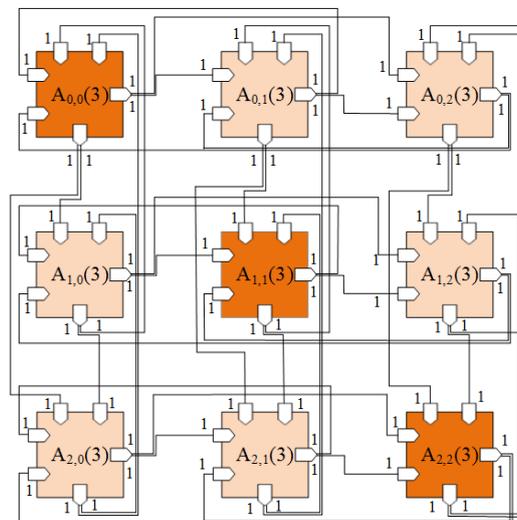


Fig. 5. Graphical view of dataflow network NW_{33} constructed of 9 actors and 27 communication channels from matrix $A[3 \times 3]$

Generation of parallel dataflow networks of actors. Given a blocked matrix $B[M \times M]$, our goal is to generate a matrix $A[M \times M]$ of actors and to establish connections between their input and output ports, thus generating a dataflow network NW_{MM} for shortest paths search. The key new result of the section is a procedure of generating actors of two types and generating a network NW_{MM} for various M . In NW_{MM} , the number of actors is M^2 . Each actor has $2 \cdot (M - 1)$ input ports, 2 output ports, and M actions.

Algorithm 8 generates the diagonal and non-diagonal actors for block $B_{r,c}$ of matrix $B[M \times M]$. It adds the input and output ports and the state variables to each actor. Then, it creates actions. For the diagonal actor *Block_D* modelling $A_{r,r}$ ($r = c$), it creates one diagonal action Dig_r and $M - 1$ peripheral actions Per_k , $k = 0 \dots M - 1$, $k \neq r$. Action Dig_r produces tokens at output ports *Lrow* and *Lcol* that are transferred to $2 \cdot (M - 1)$ input ports of other actors from row r and column c of matrix A . The peripheral actions of the diagonal actor are divided into two groups: $G_D^1 = \{Per_0, \dots, Per_{r-1}\}$ and $G_D^2 = \{Per_{r+1}, \dots, Per_{M-1}\}$.

Algorithm 8 adds one cross action of type *C1*, one cross action of type *C2* and $M-2$ peripheral actions to the non-diagonal actor *Block_N* representing $A_{r,c}$, $r \neq c$. Cross action Crs_r calculates block $B_{r,c}$ of type *C2* through diagonal block $B_{r,r}$ and produces a token at output port *Lcol* which is transferred to $M - 1$ input ports of actors on column c . Cross action Crs_c calculates block $B_{r,c}$ of type *C1* over diagonal block $B_{c,c}$ and produces a token at output port *Lrow* which is transferred to $M - 1$ input ports of actors on row r . Let $r < c$. Peripheral actions Per_k , $k = 0 \dots M - 1$, $k \neq r$ and $k \neq c$ of actor $A_{r,c}$ are divided into three groups: $G_N^1 = \{Per_0 \dots Per_{r-1}\}$, $G_N^2 = \{Per_{r+1} \dots Per_{c-1}\}$ and $G_N^3 = \{Per_{c+1} \dots Per_{M-1}\}$.

Algorithm 8: Generation of diagonal and non-diagonal actors (*Generate_Actor*)

Input: a number M of blocks (actors) in row (column)
Input: a row number r
Input: a column number c
Output: a generated actor *Actor*

```

if  $r = c$  then
    Actor ← CreateActor ("Block_D",  $r, c$ );
else
    Actor ← CreateActor ("Block_N",  $r, c$ );
for  $k \leftarrow 0$  to  $M - 1$  do
    if  $k \neq c$  then Actor.AddInPort ( $r, k$ );
for  $k \leftarrow 0$  to  $M - 1$  do
    if  $k \neq r$  then Actor.AddInPort ( $k, c$ );
Actor.AddOutPort ("Lrow"); Actor.AddOutPort ("Lcol"); Actor.AddStateVar ("Lev", 0);
Actor.AddStateVar ("Row",  $r$ ); Actor.AddStateVar ("Col",  $c$ );
if  $r = c$  then
    ActionD0 ← CreateAction ("Dig",  $r$ ); Actor.AddAction (ActionD0);
    for  $k \leftarrow 0$  to  $M - 1$  do
        if  $k \neq r$  then ActionP3 ← CreateAction ("Per",  $k$ ); Actor.AddAction (ActionP3);
else
    ActionC1 ← CreateAction ("Crs",  $c$ ); Actor.AddAction (ActionC1);
    ActionC2 ← CreateAction ("Crs",  $r$ ); Actor.AddAction (ActionC2);
    for  $k \leftarrow 0$  to  $M - 1$  do
        if  $k \neq r$  and  $k \neq c$  then ActionP3 ← CreateAction ("Per",  $k$ ); Actor.AddAction (ActionP3);
return Actor;

```

Once the actors are created, Algorithm 9 generates connections among them. It traverses all actors identified by *ID_Dest* and *Actor_Dest* and considers them as destinations. For each destination actor, the algorithm takes every input port *Port_Dest* and selects a single source that is an actor

(ID_Source , $Actor_Source$) and its output port ($Port_Source$). The first loop along variable p goes over destination ports and their sources of column c , and the second loop along variable p goes over destination ports and sources of row r .

The generated dataflow network NW_{MM} has the properties of rate consistency, boundedness, and liveness. The behavior of actors and actions in the network is correctly synchronized. Each actor calculates its own block; therefore, no conflicts occur between the actors. In the diagonal and non-diagonal actors, all actions are connected to distinct input ports, which guarantees that the same token cannot be consumed by different actions; it leads to the independent firing of actions.

Algorithm 9: Generation of connections between actors (*Connect_Actors*)

Input: a number M of blocks in row (column)

Output: a CAL network NW

for $r \leftarrow 0$ **to** $M - 1$ **do**

for $c \leftarrow 0$ **to** $M - 1$ **do**

$ID_Dest \leftarrow r \times M + c$;

$Actor_Dest \leftarrow AcatorName(r, c)$;

for $p \leftarrow 0$ **to** $M - 1$ **do**

if $p \neq r$ **then**

$Port_Dest \leftarrow PortName(p, c)$; $ID_Source \leftarrow p \times M + c$;

$Actor_Source \leftarrow AcatorName(p, c)$; $Port_Source \leftarrow \text{"Lcol"}$;

$NW.AddConnect(ID_Dest, Actor_Dest, Port_Dest, ID_Source, Actor_Source, Port_Source)$;

for $p \leftarrow 0$ **to** $M - 1$ **do**

if $p \neq c$ **then**

$Port_Dest \leftarrow PortName(r, p)$; $ID_Source \leftarrow r \times M + p$;

$Actor_Source \leftarrow AcatorName(r, p)$; $Port_Source \leftarrow \text{"Lrow"}$;

$NW.AddConnect(ID_Dest, Actor_Dest, Port_Dest, ID_Source, Actor_Source, Port_Source)$;

return NW ;

If block B_1^k of level k is used directly (or over other blocks) for calculating block B_2^l to level l , we denote it with precedence $B_1^k \Rightarrow B_2^l$. It can be observed from Algorithm 2 that in BFW the precedence $B_{v,u}^m \Rightarrow B_{v,u}^{m+1}$ holds for all $v, u, m = 0 \dots M - 1$. In NW_{MM} , a precedence relation between block calculations exists, which determines a partial order of firing actions.

Claim 3. In the diagonal CAL-actor $Block_D$ processing block $B_{m,m}$, actions of G_D^1 are fired before action Dig_m and actions of G_D^2 are fired after Dig_m . Actions of G_D^1 as well as actions of G_D^2 can be fired in any order with respect to each other. Then NW_{MM} correctly computes the shortest paths between all pairs of vertices.

Proof. In BFW , $B_{m,m}^k \Rightarrow B_{m,m}^{k+1}$, $k = 0 \dots m$ hold. Actions of G_D^1 calculate block $B_{m,m}$ from level $B_{m,m}^0$ to levels $B_{m,m}^1 \dots B_{m,m}^m$. Action Dig_m calculates the block to level $B_{m,m}^{m+1}$ and actions of G_D^2 calculate it to levels $B_{m,m}^{m+2} \dots B_{m,m}^M$ respectively. The diagonal block calculations hold the following:

1. For block type $D0$, block $B_{m,m}$ must be calculated to level m before calculating $B_{m,m}^{m+1}$. It can be only done by calculating $B_{m,m}^1 \dots B_{m,m}^m$ through $B_{m,0}^1, B_{0,m}^1 \dots B_{m,m-1}^m, B_{m-1,m}^m$ while considering the block as of type $P3$. Then in NW_{MM} , $B_{m,m}^1 \Rightarrow B_{m,m}^m \dots B_{m,m}^{m-1} \Rightarrow B_{m,m}^m$ hold. The actions of G_D^1 are fired before action Dig_m .

2. The precedencies $B_{m,m}^{m+1} \Rightarrow B_{m,m}^{m+2} \dots B_{m,m}^{m+1} \Rightarrow B_{m,m}^M$ are derived from the fact that $B_{m,m}^{m+2} \dots B_{m,m}^M$ are directly or indirectly calculated through $B_{m,m}^{m+1}$.

3. The CAL network can reorder $B_{m,m}^k$ and $B_{m,m}^{k+1}$, $k = 1 \dots m - 1$, for three reasons: a) the block calculations are independent and, therefore, do not precede each other since $B_{m,m}$ is calculated at levels k and $k+1$ by accounting for paths between vertices of V_m passing through vertices of non-intersected subsets V_k and V_{k+1} ; b) Claim 2 allows the calculation of $B_{m,m}^k$ through $B_{m,k}^{k'}$ and $B_{k,m}^{k''}$ of higher levels of $k' > k$ and $k'' > k$; c) the operation of choosing a minimum of two numbers is commutative and associative.

4. Assertions like those of point 3 are proved for the case when $k = m + 1 \dots M - 1$.

Points 1 and 2 prove that in *Block_D* actions of G_D^1 are fired before action Dig_m and actions G_D^2 are fired after Dig_m . Points 3 and 4 prove that actions of G_D^1 (as well as actions of G_D^2) can be fired in any order with respect to each other in *Block_D*. The claim is proved.

Claim 4. In the non-diagonal CAL-actor *Block_N*, which processes block $B_{v,u}$, $v \neq u$, actions of G_N^1 are fired before action Crs_v , actions of G_N^2 are fired after Crs_v and before Crs_u , and actions of G_N^3 are fired after Crs_u . In each of three subsets G_N^1 , G_N^2 and G_N^3 the actions can be fired in any order with respect to each other. Then NW_{MM} correctly computes the shortest paths between all pairs of vertices.

Proof. Let $v < u$. Actions of G_N^1 calculate block $B_{v,u}$ from level $B_{v,u}^0$ to levels $B_{v,u}^1 \dots B_{v,u}^v$. Action Crs_v calculates the block to level $B_{v,u}^{v+1}$ and actions of G_N^2 calculate it to levels $B_{v,u}^{v+2} \dots B_{v,u}^u$. Action Crs_u calculates the block to level $B_{v,u}^{u+1}$ and actions of G_N^3 calculate it to levels $B_{v,u}^{u+2} \dots B_{v,u}^M$. The following precedencies hold for the non-diagonal block calculations:

1. Since the block type *C2* establishes precedence $B_{v,u}^v \Rightarrow B_{v,u}^{v+1}$, block $B_{v,u}$ must be calculated to level v before calculating $B_{v,u}^{v+1}$. The only way is to perform calculations $B_{v,u}^1 \dots B_{v,u}^v$ by considering the block as of type *P3*. In this case, $B_{v,u}^1 \Rightarrow B_{v,u}^v \dots B_{v,u}^{v-1} \Rightarrow B_{v,u}^v$ hold.

2. Since the block type *C1* establishes precedence $B_{v,u}^u \Rightarrow B_{v,u}^{u+1}$, block $B_{v,u}$ must be calculated to level u before calculating $B_{v,u}^{u+1}$. The only way is to perform calculations of $B_{v,u}^{v+2} \dots B_{v,u}^u$ after calculating $B_{v,u}^{v+1}$ by considering the block as of type *P3*. Therefore, precedencies $B_{v,u}^{v+2} \Rightarrow B_{v,u}^u \dots B_{v,u}^{u-1} \Rightarrow B_{v,u}^u$ hold.

3. The precedencies $B_{v,u}^{u+1} \Rightarrow B_{v,u}^{u+2} \dots B_{v,u}^{u+1} \Rightarrow B_{v,u}^M$ are derived from the fact that $B_{v,u}^{u+2} \dots B_{v,u}^M$ can only be calculated through $B_{v,u}^{u+1}$.

4. NW_{MM} can refuse the precedence $B_{v,u}^m \Rightarrow B_{v,u}^{m+1}$, $m = 1 \dots v - 1$ and can reorder $B_{v,u}^m$ and $B_{v,u}^{m+1}$ for three reasons: a) the calculations are independent and, therefore, do not precede each other since $B_{v,u}$ is calculated at levels m and $m+1$ by accounting for paths between vertices of subsets V_v and V_u passing through vertices of non-intersected subsets V_m and V_{m+1} ; b) Claim 2 allows the calculation of $B_{v,u}^m$ through $B_{v,u}^{m'}$ and $B_{v,u}^{m''}$ at higher levels of $m' > m$ and $m'' > m$; c) the operation of choosing a minimum of two numbers is commutative and associative.

5. Similar assertions are proved for the case when $m = v + 1 \dots u - 1$ and when $m = u + 1 \dots M - 1$.

Points 1, 2 and 3 prove that in $Block_N$ the actions of G_N^1 are fired before action Cr_{S_v} , the actions of G_N^2 are fired after Cr_{S_v} and before Cr_{S_u} , and the actions of G_N^3 are fired after action Cr_{S_u} . Points 4 and 5 prove that the actions of G_N^1 (as well as of G_N^2 and G_N^3) can be fired in any order with respect to each other in $Block_N$. The claim is proved.

The number of possible firing sequences of actions of diagonal actors is $|G_D^1|! \cdot |G_D^2|!$ and is $|G_N^1|! \cdot |G_N^2|! \cdot |G_N^3|!$ of non-diagonal actors, where $|G|!$ is factorial of G 's cardinality. The number rapidly increases with the growth of M .

CAL-network development tool and tunable dataflow CAL-engine in C/C++. Based on the C/C++ language we have developed a tool for creating parallel networks of dataflow actors and have developed a tunable multithreaded CAL-based engine (fig. 6) for multicore systems. The tool and engine were used for the realization of the proposed dataflow CAL-actors and parallel networks, which solve the all-pairs shortest path problem. In the current version of engine, any action of any actor is implemented by a separate thread, although we consider other solutions of mapping actors to threads. Since many concurrently and asynchronously operating actions (threads) can simultaneously update shared resources, synchronization primitives protect the resources. The concurrent asynchronous behavior is a source of increasing the throughput of the networks and speeding up the shortest paths search against OpenMP.

Fig. 6 depicts a flexible architecture of the tool and engine. It provides methods of specifying and generating actors, connections and whole network for the problem under solving. The generators process the specifications and elaborate a parallel dataflow network in an internal format. A multithreaded CAL-based engine is firstly tuned to the network and then implements it on a multi-core system.

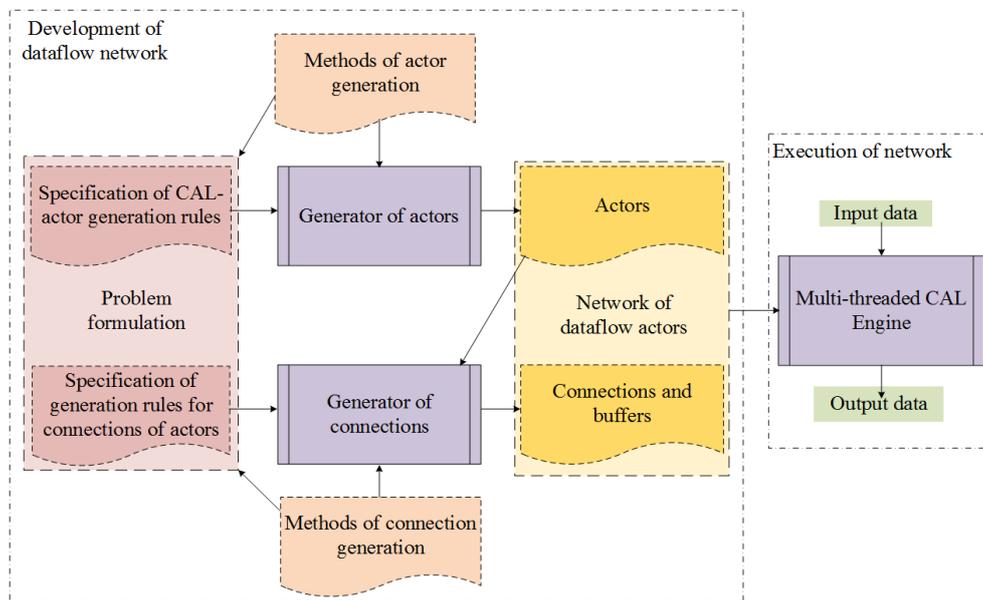


Fig. 6. Architecture of CAL-based development tool and tunable CAL-engine implemented in C/C++

Although the actions may be fired (executed) within one actor only in series, the engine parallelizes the actions' implementations regarding their readiness and selection before firing. It can execute the action in parallel with checking if another action is ready to be fired next (the FSM state, the availability of action's input tokens, the guard condition, and the availability of output ports to receive the produced tokens are considered). To have this property, the CAL-engine implements each action by a separate thread.

The advantage of our CAL-based implementation is the integration of CAL in C/C++ in such a way that all facilities of C/C++ become available for the implementation. C/C++ classes of objects implement all components of the CAL-actor and network. As a result, the network and each of its actors are instantiated over complex data structures and sets of methods written in C/C++. The body of each action is represented as a function in C/C++. To ensure that the implementation is consistent with the CAL model of computation, we have developed a tool for checking and validating the structure.

Experimental results. In the paper, we report results given by the implementations of the dataflow parallel networks and OpenMP based implementations of the *PBFW* algorithm on multi-core systems. The networks of dataflow actors were generated from various block-matrix configurations and block sizes and implemented in C/C++ with the threaded CAL engine. The results are obtained on randomly generated simple complete weighted directed graphs of 1200, 2400, 3600 and 4800 vertices on four Intel(R) Core(TM) processors i3-550, i5-5200U, i7-9750h and i7-10700. The graphs provide high computational load which gives a correct comparison of the implementations. Table describes processors' parameters.

Parameters of four multi-core processors

Processor	Cache L1, KB	Cache L2, KB	Cache L3, MB	Frequency, GHz	Cores	Logical processors
i7-10700	8 × 64	8 × 256	16.0	2.90	8	16
i7-9750h	6 × 64	6 × 256	12.0	2.60	6	12
i5-5200U	2 × 64	2 × 256	3.0	2.20	2	4
i3-550	32 + 64	2 × 256	4.0	3.20	2	

Fig. 7 and 8 show the speedup the dataflow CAL-networks and their multi-threaded implementations have given against matching single-thread implementations of the Floyd–Warshall Algorithm 1. The block count M in matrix B varied in the range 2 to 10. The number of actors varied in the range 4 to 100, the number of actor input ports varied in the range 2 to 18, and the number of output ports was 2 for all actors. The number of actions within actor varied in the range 2 to 10, therefore, the number of threads in the network implementations reached up to 1000. The optimal number M_{opt} of blocks has given the highest speedup of the CAL-networks.

On the 2-core i3-550 processor and the graph of 1200 vertices (fig. 7, a), the CAL-network has given the speedup of 2.57 at $M_{opt} = 4$. For larger M the speedup has reduced to 1.61. On the graph of 2400 vertices, the highest speedup of 2.45 is also obtained at $M_{opt} = 4$. For both graphs, the speedup is larger than the number of cores. For the 2-core i5-5200U processor and the graph of 1200 vertices (fig. 7, b) we can observe the similar pattern, where the highest speedup of 2.51 is obtained at $M_{opt} = 4$. The increase in the graph size to 2400 and 3600 shifts the value of M_{opt} from 4 to 8 (speedup is 2.54) and then to 10 (speedup is 2.55). We can mainly explain this as the CAL-networks exploit the processor hierarchical memory and caches more efficiently for lower size of block.

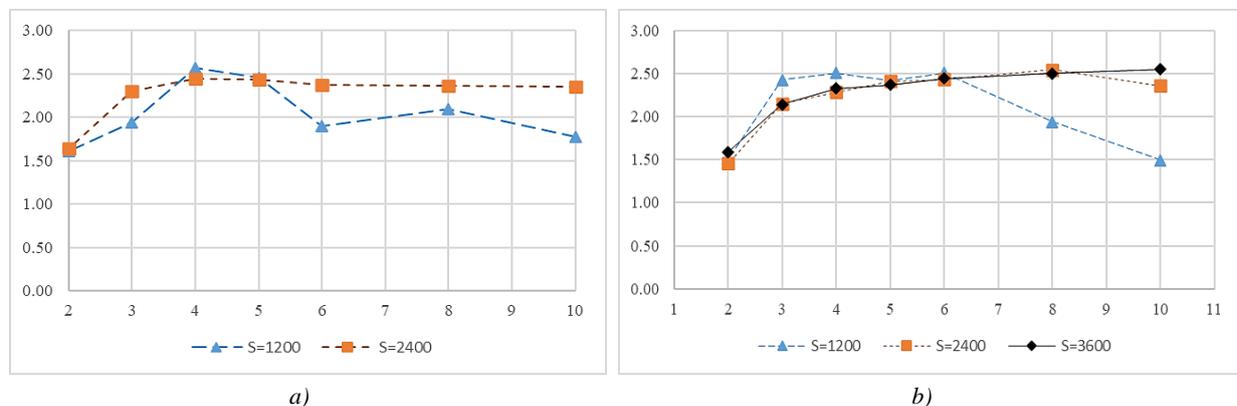


Fig. 7. Speedup (vertical axis) of multi-threaded CAL-networks against single-thread *FW* vs. block count M (horizontal axis) on a) i3-550 and b) i5-5200U processors for three sizes of graphs: 1200, 2400 and 3600 vertices

On the 6-core i7-9750h processor and the graph of 1200 vertices (fig. 8, a), the CAL-networks have given the maximum speedup of 4.98 at $M_{opt} = 6$, which is smaller than the number of cores due to the insufficient potential parallelism and low useful load (see fig. 1, a). On the graphs of 2400 and 3600 vertices, the highest speedups of 6.59 and 6.96 are obtained at $M_{opt} = 8$ and $M_{opt} = 6$ respectively. In both cases, the speedup exceeds the number of cores. On the 8-core i7-10700 processor, the speedup patterns by the CAL-networks are very close for graphs of 2400, 3600 and 4800 vertices (fig. 8, b). For all graphs the maximum speedup of 9.78, 9.37 and 9.34 is obtained at $M_{opt} = 10$. Fig. 8, b also shows the speedup given by OpenMP on the three graph-sizes, which is significantly less against the networks. The speedup is being decreased with the growth of the graph-size and is being increased with the growth of the number of blocks. It can be observed that the CAL-networks convincingly gain the *BFW* OpenMP implementations with respect to runtime.

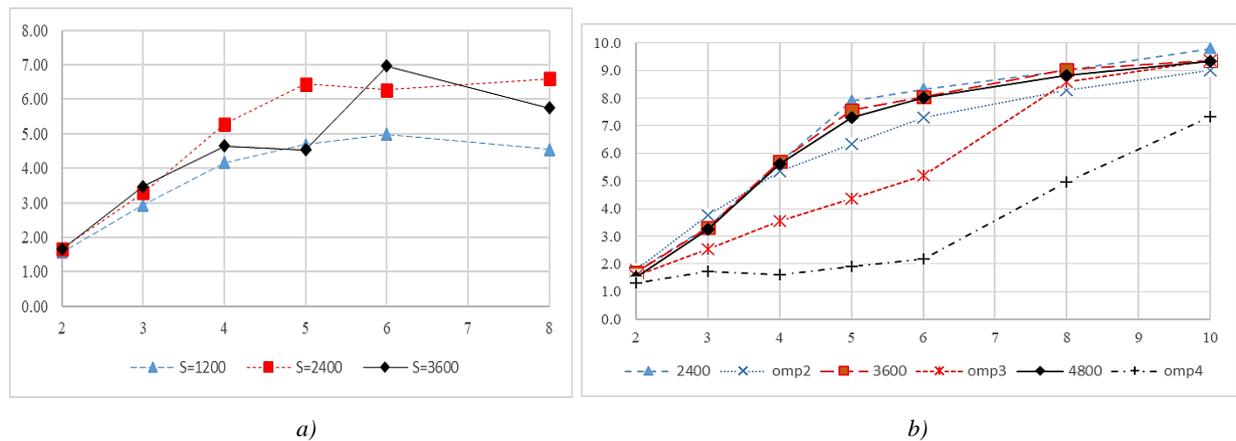


Fig. 8. Speedup (vertical axis) of multi-threaded CAL-networks and OpenMP-*BFW* implementations against single-thread *FW* vs. block count M (horizontal axis) on a) i7-9750h and b) i7-10700 processors for four sizes of graphs (omp2 – 2400, omp3 – 3600 and omp4 – 4800 vertices)

According to (1), the estimated speedup of the block-parallel Floyd – Warshall algorithm implemented in the fork-join style (OpenMP) is 1.33, 3.00, 4.00, 6.25, 5.14, 6.40 and 6.66 for $M = 2, 3, 4, 5, 6, 8, 10$ respectively on 8 cores. For comparison, the networks have given on i7-10700 and on the graph of 2400 vertices much higher speedup of 1.72, 3.35, 5.68, 7.90, 8.31, 9.01 and 9.78 respectively for the same values of M . We can mainly explain this by efficient exploitation of caches and advantages of the networks and their threaded parallel implementations due to highly asynchronous behaviour.

The graphics depicted in fig. 7 and 8 have found out the patterns as follows:

1. There is an optimal number M_{opt} of blocks for which the speedup by the multithreaded CAL-networks is the highest compared to the single-thread *FW*.
2. The highest speedup given by the dataflow networks exceeds the number of cores, which is a very good result for the blocked algorithm with strong data dependences between blocks.
3. M_{opt} depends on the number P of cores, the block-matrix size M , the graph size N , and the scheduler of threads of the operating system.
4. The increase in the size M of matrices B and A increases the amount of parallelism in the CAL-networks, which leads to the growth of computation speedup.
5. The larger number P of cores requires more parallelism and therefore larger M_{opt} .
6. The growth of the graph size N usually leads to the increase of M_{opt} as the processor caches operate more efficiently at smaller block sizes [13].
7. The CAL-networks give the speedup which is higher than that OpenMP gives.

It should be noted that the scheduler of threads of the operating system influences the order of executions of threads, and the increase in M increases the number of threads in the CAL-network implementations which increases the workload of the operating system.

Conclusion. Nowadays, the blocked Floyd – Warshall algorithm is typically parallelized in the fork-join style with OpenMP where each block is calculated in a loop level-by-level. The paper has proven that the block calculations can be reordered, thus increasing the load of cores in the multi-core system. The simulation tool has shown that the reordering can speed up the shortest paths search up to 25 %. The paper has proposed a novel method of generating dataflow networks of CAL-actors, where the management of actor and action firing is carried out over the block calculation levels. The new feature of the networks is that in each actor the executions of actions are ordered partially. The multi-threaded tuneable CAL-engine accounts for the feature and implements the networks in C/C++. The experiments on large complete directed graphs and four multi-core processors have shown that at optimal block count the networks speed up computations against the single-threaded implementations by the following figures: i3-550 (2 cores) – 2.57 (28.5 % higher than core count); i5-5200U (2 cores) – 2.55 (27.5 % higher than core count); i7-9750h (6 cores) – 6.96 (16.0 % higher than core count); i7-10700 (8 cores) – 9.78 (22.3 % higher than core count).

References

1. Floyd R. W. Algorithm 97: Shortest path. *Communications of the ACM*, 1962, vol. 5, no. 6, p. 345.
2. Madkour A, Aref W. G., Rehman F. U., Rahman M. A., Basalamah S. A. *Survey of Shortest-Path Algorithms*, 2017, 26 p. Available at: <https://arxiv.org/abs/1705.02044> (accessed 23.11.2022).
3. Anu P., Kumar M. G. Finding all-pairs shortest path for a large-scale transportation network using parallel Floyd-Warshall and parallel Dijkstra algorithms. *Journal of Computing in Civil Engineering*, 2013, vol. 27, no. 3, pp. 263–273.
4. Prihozhy A. A., Mattavelli M., Mlynek D. Evaluation of parallelization potential for efficient multimedia implementations: dynamic evaluation of algorithm critical path. *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 15, no. 5, 2005, pp. 593–608.
5. Singh A., Mishra P. K. Performance analysis of Floyd Warshall algorithm vs rectangular algorithm. *International Journal of Computer Applications*, 2014, vol. 107, no. 16, pp. 23–27.
6. Venkataraman G. A., Sahni S., Mukhopadhyaya S. Blocked all-pairs shortest paths algorithm. *Journal of Experimental Algorithmics (JEA)*, 2003, vol. 8, pp. 857–874.
7. Park J., Penner M., Prasanna V. K. Optimizing graph algorithms for improved cache performance. *IEEE Transactions on Parallel and Distributed Systems*, 2004, vol. 15, no. 9, pp. 769–782.
8. Madduri K., Bader D. A., Berry J. W., Crobak J. R. An experimental study of a parallel shortest path algorithm for solving large-scale graph instances. *Proceedings of the Nine Workshop on Algorithm Engineering and Experiments, ALENEX 2007, New Orleans, Louisiana, USA, 6 January 2007*. New Orleans, 2007, pp. 23–35.
9. Albalwi E., Thulasiraman P., Thulasiram R. Task level parallelization of all pair shortest path algorithm in OpenMP 3.0. *Advances in Computer Science and Engineering (CSE 2013)*. Los Angeles, Atlantis Press, 2013, pp. 109–112.
10. Tang P. Rapid development of parallel blocked all-pairs shortest paths code for multi-core computers. *IEEE SOUTHEASTCON 2014, Lexington, KY, USA, 13–16 March 2014*. Lexington, 2014, pp. 1–7.
11. Prihozhy A. A., Karasik O. N. *Heterogeneous blocked all-pairs shortest paths algorithm*. *Sistemnyj analiz i prikladnaja informatika [System Analysis and Applied Information Science]*, 2017, no. 3, pp. 68–75 (In Russ.). <https://doi.org/10.21122/2309-4923-2017-3-68-75>
12. Karasik O. N., Prihozhy A. A. Threaded block-parallel algorithm for finding the shortest paths on graph. *Doklady Belorusskogo gosudarstvennogo universiteta informatiki i radioelektroniki [Reports of the Belarusian State University of Informatics and Radioelectronics]*, 2018, no. 2, pp. 77–84 (In Russ.).
13. Karasik O. N., Prihozhy A. A. *Tuning block-parallel all-pairs shortest path algorithm for efficient multi-core implementation*. *System Analysis and Applied Information Science*, 2022, no. 3, pp. 68–75. <https://doi.org/10.21122/2309-4923-2022-3-57-65>
14. Prihozhy A. A. *Simulation of direct mapped, k-way and fully associative cache on all pairs shortest paths algorithms*. *System Analysis and Applied Information Science*, 2019, no. 4, pp. 10–18. <https://doi.org/10.21122/2309-4923-2019-4-10-18>
15. Prihozhy A. A. *Optimization of data allocation in hierarchical memory for blocked shortest paths algorithms*. *System Analysis and Applied Information Science*, 2021, no. 3, pp. 40–50. <https://doi.org/10.21122/2309-4923-2021-3-40-50>
16. Likhoded N. A., Sipeyko D. S. Generalized blocked Floyd – Warshall algorithm. *Journal of the Belarusian State University. Mathematics and Informatics*, 2019, no. 3, pp. 84–92 (In Russ.).

17. Kahn G. The semantics of a simple language for parallel programming. *Information Processing 74: Proceedings of the IFIP Congress 74, Stockholm, Sweden, 5–10 August 1974*. Stockholm, 1974, pp. 471–475.
18. Lee E. A., Messerschmitt D. G. Synchronous dataflow. *Proceedings of the IEEE*, September 1987, vol. 75, no. 9, pp. 1235–1245.
19. Prihozhy A., Mlynek D., Solomennik M., Mattavelli M. Techniques for optimization of net algorithms. *2002 International Conference on Parallel Computing in Electrical Engineering (PARELEC 2002), Warsaw, Poland, 22–25 September 2002*. Warsaw, 2002, pp. 211–216.
20. Eker J., Janneck J. W. *Cal Language Report : Technical Report UCB/ERL M03/48*. University of California at Berkeley, December 2003, 107 p.
21. Bhattacharyya S. S., Brebner G., Janneck J. W., Eker J., Platen C., ..., Raulet M. OpenDF – a dataflow toolset for reconfigurable hardware and multicore systems. *First Swedish Workshop on Multi-Core Computing, MCC, Ronneby, Sweden, 27–28 November 2008*. Ronneby, 2008, pp. 43–49.
22. Murthy P. K., Lee E. A. Multidimensional synchronous dataflow. *IEEE Transactions on Signal Processing*, 2002, vol. 50, no. 8, pp. 2064–2079.
23. Bhattacharya B., Bhattacharyya S. S. Parameterized dataflow modeling for DSP systems. *IEEE Transactions on Signal Processing*, 2001, vol. 49, no. 10, pp. 2408–2421.
24. Bebelis V., Fradet P., Girault A., Lavigueur B. *BPDF: Boolean Parametric Data Flow : Research Report RR-8333*. INRIA, 2013, 21 p.
25. Rahman A.-H. Ab, Prihozhy A., Mattavelli M. Pipeline synthesis and optimization of FPGA-based video processing applications with CAL. *EURASIP Journal on Image and Video Processing*, vol. 2011:19, pp. 1–28. <https://doi.org/10.1186/16875281-2011-19>
26. Prihozhy A., Casale-Brunet S., Bezati E., Mattavelli M. Efficient dynamic optimization heuristics for dataflow pipelines. *2018 IEEE International Workshop on Signal Processing Systems, SiPS 2018, Cape Town, South Africa, 21–24 October 2018*. Cape Town, 2018, pp. 337–342.
27. Prihozhy A. A., Casale-Brunet S., Bezati E., Mattavelli M. Pipeline synthesis and optimization from branched feedback dataflow programs. *Journal of Signal Processing Systems*, Springer Nature, 2020, vol. 92, pp. 1091–1099. <https://doi.org/10.1007/s11265-020-01568-5>

Список использованных источников

1. Floyd, R. W. Algorithm 97: Shortest path / R. W. Floyd // *Communications of the ACM*. – 1962. – Vol. 5, no. 6. – P. 345.
2. Survey of Shortest-Path Algorithms / A. Madkour [et al.]. – 2017. – 26 p. – Mode of access: <https://arxiv.org/abs/1705.02044>. – Date of access: 23.11.2022.
3. Anu, P. Finding all-pairs shortest path for a large-scale transportation network using parallel Floyd-Warshall and parallel Dijkstra algorithms / P. Anu, M. G. Kumar // *J. of Computing in Civil Engineering*. – 2013. – Vol. 27, no. 3. – P. 263–273.
4. Prihozhy, A. A. Evaluation of parallelization potential for efficient multimedia implementations: dynamic evaluation of algorithm critical path / A. A. Prihozhy, M. Mattavelli, D. Mlynek // *IEEE Transactions on Circuits and Systems for Video Technology*. – 2005. – Vol. 15, no. 5. – P. 593–608.
5. Singh, A. Performance analysis of Floyd Warshall algorithm vs rectangular algorithm / A. Singh, P. K. Mishra // *Intern. J. of Computer Applications*. – 2014. – Vol. 107, no. 16. – P. 23–27.
6. Venkataraman, G. A. Blocked all-pairs shortest paths algorithm / G. A. Venkataraman, S. Sahni, S. Mukhopadhyaya // *J. of Experimental Algorithmics (JEA)*. – 2003. – Vol. 8. – P. 857–874.
7. Park, J. Optimizing graph algorithms for improved cache performance / J. Park, M. Penner, V. K. Prasanna // *IEEE Transactions on Parallel and Distributed Systems*. – 2004. – Vol. 15, no. 9. – P. 769–782.
8. An experimental study of a parallel shortest path algorithm for solving large-scale graph instances / K. Madduri [et al.] // *Proc. of the Nine Workshop on Algorithm Engineering and Experiments, ALENEX 2007, New Orleans, Louisiana, USA, 6 Jan. 2007*. – New Orleans, 2007. – P. 23–35.
9. Albalwi, E. Task level parallelization of all pair shortest path algorithm in OpenMP 3.0 / E. Albalwi, P. Thulasiraman, R. Thulasiram // *Advances in Computer Science and Engineering (CSE 2013)*. – Los Angeles : Atlantis Press, 2013. – P. 109–112.
10. Tang, P. Rapid development of parallel blocked all-pairs shortest paths code for multi-core computers / P. Tang // *IEEE SOUTHEASTCON 2014, Lexington, KY, USA, 13–16 Mar. 2014*. – Lexington, 2014. – P. 1–7.
11. Прихожий, А. А. Разнородный блочный алгоритм поиска кратчайших путей между всеми парами вершин графа / А. А. Прихожий, О. Н. Карасик // *Системный анализ и прикладная информатика*. – 2017. – № 3. – С. 68–75. <https://doi.org/10.21122/2309-4923-2017-3-68-75>

12. Карасик, О. Н. Поточный блочно-параллельный алгоритм поиска кратчайших путей на графе / О. Н. Карасик, А. А. Прихожий // Доклады БГУИР. – 2018. – № 2. – С. 77–84.
13. Karasik, O. N. Tuning block-parallel all-pairs shortest path algorithm for efficient multi-core implementation / O. N. Karasik, A. A. Prihozhy // System Analysis and Applied Information Science. – 2022. – No. 3. – P. 68–75. <https://doi.org/10.21122/2309-4923-2022-3-57-65>
14. Prihozhy, A. A. Simulation of direct mapped, k-way and fully associative cache on all pairs shortest paths algorithms / A. A. Prihozhy // System Analysis and Applied Information Science. – 2019. – No. 4. – P. 10–18. <https://doi.org/10.21122/2309-4923-2019-4-10-18>
15. Prihozhy, A. A. Optimization of data allocation in hierarchical memory for blocked shortest paths algorithms / A. A. Prihozhy // System Analysis and Applied Information Science. – 2021. – No. 3. – P. 40–50. <https://doi.org/10.21122/2309-4923-2021-3-40-50>
16. Лиходед, Н. А. Обобщенный блочный алгоритм Флойда – Уоршелла / Н. А. Лиходед, Д. С. Сипейко // Журнал Бел. гос. ун-та. Математика. Информатика. – 2019. – № 3. – С. 84–92.
17. Kahn, G. The semantics of a simple language for parallel programming / G. Kahn // Information Processing 74: Proc. of the IFIP Congress 74, Stockholm, Sweden, 5–10 Aug. 1974. – Stockholm, 1974. – P. 471–475.
18. Lee, E. A. Synchronous dataflow / E. A. Lee, D. G. Messerschmitt // Proc. of the IEEE. – Sept. 1987. – Vol. 75, no. 9. – P. 1235–1245.
19. Techniques for optimization of net algorithms / A. Prihozhy [et al.] // 2002 Intern. Conf. on Parallel Computing in Electrical Engineering (PARELEC 2002), Warsaw, Poland, 22–25 Sept. 2002. – Warsaw, 2002. – P. 211–216.
20. Eker, J. Cal Language Report : Technical Report UCB/ERL M03/48 / J. Eker, J. W. Janneck. – University of California at Berkeley, Dec. 2003. – 107 p.
21. OpenDF – a dataflow toolset for reconfigurable hardware and multicore systems / S. S. Bhattacharyya [et al.] // First Swedish Workshop on Multi-Core Computing, MCC, Ronneby, Sweden, 27–28 Nov. 2008. – Ronneby, 2008. – P. 43–49.
22. Murthy, P. K. Multidimensional synchronous dataflow / P. K. Murthy, E. A. Lee // IEEE Transactions on Signal Processing. – 2002. – Vol. 50, no. 8. – P. 2064–2079.
23. Bhattacharya, B. Parameterized dataflow modeling for DSP systems / B. Bhattacharya, S. S. Bhattacharyya // IEEE Transactions on Signal Processing. – 2001. – Vol. 49, no. 10. – P. 2408–2421.
24. BPDF: Boolean Parametric Data Flow: Research Report RR-8333/ V. Bebelis [et al.]. – INRIA, 2013. – 21 p.
25. Rahman, A.-H. Ab. Pipeline synthesis and optimization of FPGA-based video processing applications with CAL / A.-H. Ab Rahman, A. Prihozhy, M. Mattavelli // EURASIP J. on Image and Video Processing. – 2011. – Vol. 2011:19. – P. 1–28. <https://doi.org/10.1186/16875281-2011-19>
26. Efficient dynamic optimization heuristics for dataflow pipelines / A. Prihozhy [et al.] // 2018 IEEE Intern. Workshop on Signal Processing Systems, SiPS 2018, Cape Town, South Africa, 21–24 Oct. 2018. – Cape Town, 2018. – P. 337–342.
27. Pipeline synthesis and optimization from branched feedback dataflow programs / A. A. Prihozhy [et al.] // J. of Signal Processing Systems, Springer Nature. – 2020. – Vol. 92. – P. 1091–1099. <https://doi.org/10.1007/s11265-020-01568-5>

Information about the author

Anatoly A. Prihozhy, D. Sc. (Eng.), Professor, Belarusian National Technical University.
E-mail: prihozhy@yahoo.com

Информация об авторе

Анатолий Алексеевич Прихожий, доктор технических наук, профессор, Белорусский национальный технический университет.
E-mail: prihozhy@yahoo.com

ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ INFORMATION TECHNOLOGIES



УДК 519.711
<https://doi.org/10.37661/1816-0301-2023-20-2-85-95>

Оригинальная статья
Original Paper

Коррекция запросов в системе информационной поддержки принятия решений

С. Ф. Липницкий

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

Аннотация

Цели. Решается задача математического моделирования и алгоритмизации процессов коррекции запросов в системе информационной поддержки принятия решений. При этом преследуются три основные цели: построение обобщенной модели информационного поиска, разработка алгоритмов допоисковой коррекции запросов и разработка алгоритмов послепоисковой коррекции запросов.

Методы. Используются методы теории множеств и теории вероятностей.

Результаты. Разработана обобщенная модель информационного поиска. В рамках модели формализованы понятия поисковой функции, критерия выдачи, релевантности и пертинентности результатов поиска. Предложены алгоритмы допоисковой и послепоисковой коррекции запросов в системе информационной поддержки принятия решений.

Заключение. Разработанные алгоритмы коррекции запросов могут быть использованы при программной реализации информационных систем поддержки принятия решений. Универсальность алгоритмов обеспечивается благодаря применению тематических корпусов текстов в различных предметных областях. Они играют определяющую роль при планировании архитектуры информационных систем и их отдельных компонентов.

Ключевые слова: алгоритм, информационная поддержка, коррекция запроса, математическая модель, поисковая функция, принятие решений

Для цитирования. Липницкий, С. Ф. Коррекция запросов в системе информационной поддержки принятия решений / С. Ф. Липницкий // Информатика. – 2023. – Т. 20, № 2. – С. 85–95.
<https://doi.org/10.37661/1816-0301-2023-20-2-85-95>

Конфликт интересов. Автор заявляет об отсутствии конфликта интересов.

Поступила в редакцию | Received 14.02.2023
Подписана в печать | Accepted 28.02.2023
Опубликована | Published 29.06.2023

Correction of requests in the information system decision support

Stanislav F. Lipnitsky

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

Abstract

Objectives. The problem of mathematical modeling and algorithmization of the processes of correction of requests in the system of information support for decision-making is being solved. At the same time, three main goals are pursued: building of a generalized information retrieval model, development of algorithms for pre-search query correction and development of algorithms for post-search query correction.

Methods. Methods of set theory and probability theory are used.

Results. A generalized information retrieval model has been developed. Within the framework of the model, the concepts of search function, issuance criterion, relevance and pertinence of search results are formalized. Algorithms for pre-search and post-search correction of queries in the information decision support system are proposed.

Conclusion. A mathematical model for correcting user requests in the information decision support system has been developed. Within the framework of the model, the efficiency of search processes in terms of the relevance and pertinence of the information found has been studied. Necessary and sufficient optimality of search functions are proved.

Keywords: algorithm, information support, query correction, mathematical model, search function, decision making

For citation. Lipnitsky S. F. *Correction of requests in the information system decision support*. Informatika [Informatics], 2023, vol. 20, no. 2, pp. 85–95 (In Russ.). <https://doi.org/10.37661/1816-0301-2023-20-2-85-95>

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

Введение. Эффективность информационного поиска определяется на основе оценок, принципиально различающихся между собой. В одних случаях оценивается релевантность, т. е. степень адекватности текста запросу, а в других – pertinence, соответствующая смысловой близости текста и информационной потребности пользователя [1]. Различают два вида релевантности – смысловую и формальную. Соответствие текста содержанию запроса называют смысловой релевантностью, а соответствие поискового образа этого текста поисковому предписанию – формальной релевантностью. Факт смысловой релевантности устанавливается администратором информационной системы на основе сопоставления запросов и результатов поиска. Формальная релевантность определяется системой алгоритмически с учетом принятого в ней критерия выдачи. Запрос может значительно отличаться от информационной потребности, зафиксированной в сознании пользователя. Поэтому часто он оказывается неспособным точно выразить эту потребность. В подобных случаях необходима коррекция запросов в режиме итерационного поиска. В статье представлены алгоритмы их коррекции в системе информационной поддержки принятия решений. Алгоритмы основываются на использовании моделей представления знаний о предметной области, которые предложены автором в статьях [2, 3].

Модель поиска. Построим обобщенную модель информационного поиска с целью формализации основных его составляющих.

Критерий выдачи. Под критерием выдачи понимают правило, по которому вычисляется степень релевантности информации, найденной в процессе поиска, и принимается решение о выдаче (или невыдаче) соответствующего текста пользователю. Определим формально понятие критерия выдачи.

Пусть T и Q – некоторые множества соответственно входных и внутренних текстов системы информационной поддержки. Обозначим через Z ($Z \subseteq T$) непустое подмножество множества T , элементы которого будем называть запросами. Текст $q \in Q$ назовем поисковым образом произвольного входного текста $t \in T$, если существует такое инъективное отображение $\omega : T \rightarrow Q$, что текст q есть образ текста t при отображении ω . Если текст $z \in Z$ – запрос, то текст $q = \omega(z)$ будем называть поисковым предписанием, соответствующим запросу z . Отображение $\eta : \omega(T) \times \omega(Z) \rightarrow R$ декартова произведения множеств поисковых образов текстов и поисковых предписаний в множество R действительных чисел будем называть критерием выдачи.

Поисковая функция. Пусть $z \in Z$ – произвольный запрос. Один шаг поиска текстов промоделируем в виде частичного мультиотображения (по терминологии из монографии [4, с. 32]) $\pi : Z \rightarrow T$ множества запросов в множество текстов. Частичное мультиотображение π назовем поисковой функцией, если множество $\pi(z)$ включает те и только те тексты $t \in T$, для которых значение критерия выдачи не меньше некоторого η_0 , т. е. $\eta(\omega(t), \omega(z)) \geq \eta_0$.

Релевантность и пертинентность. Определим биективное отображение $\theta : Z \rightarrow IP$, ставящее во взаимно однозначное соответствие множество запросов Z и множество IP соответствующих им информационных потребностей, т. е. выраженных вербально характеристик предметной области. Тогда понятия релевантных и пертинентных текстов можно формально ввести следующим образом. Произвольный текст $t \in T$ назовем релевантным запросу $z \in Z$, если существует сюръективное отображение $\mu : T \times Z \rightarrow \{0, 1\}$, такое, что $\mu(t, z) = 1$. Если же $\mu(t, z) = 0$, то будем считать, что текст t нерелевантен запросу z . Любой текст $t \in T$ будем называть пертинентным информационной потребности $\theta(z)$, если найдется сюръективное отображение $\nu : T \times \theta(Z) \rightarrow \{0, 1\}$, для которого справедливо соотношение $\nu(t, \theta(z)) = 1$, и непертинентным, когда $\nu(t, \theta(z)) = 0$.

Допоисковая коррекция запросов. При реализации поисковых функций будем различать два вида взаимодействия пользователей с информационной системой: допоисковое и послепоисковое. В процессе допоискового взаимодействия пользователю предъявляются сведения, которые используются при формулировании и коррекции запросов. Послепоисковое взаимодействие основано на оценке и использовании на последующих этапах поиска промежуточной выдачи найденной информации.

Пусть имеется запрос

$$z = \{a_1, a_2, \dots | a_i \in W_T, i = 1, 2, \dots\}, \quad (1)$$

где W_T – множество всех различных словоформ из множества текстов T .

Рассмотрим три варианта допоисковой коррекции запросов: путем дополнения ключевых слов словоизменениями и синонимами; путем применения корпусов текстов для вычисления информативности ключевых слов; путем разбиения исходного запроса на несколько.

Расширение запросов словоизменениями и синонимами ключевых слов. Словоизменения и синонимы ключевых слов содержат специальные лингвистические словари:

– словарь словоизменительных парадигм

$$Dic_{par} = \{(a, Par_a) | a \in W_T, a \in Par_a\},$$

состоящий из пар $\langle \text{словоформа}, \text{парадигма} \rangle$, где Par_a – совокупность всех словоизменений словоформы a ;

– словарь синонимичных словоформ

$$Dic_{syn} = \{(a, Syn_a) | a \in W_T, a \in Syn_a\},$$

включающий в себя пары $\langle \text{словоформа}, \text{синонимичные словоформы} \rangle$, в которых каждой словоформе a соответствует множество ее синонимов Syn_a .

Представим запрос (1) в виде конъюнкции ключевых слов:

$$z^{(1)} = a_1 \wedge a_2 \wedge \dots, a_i \in W_T, i = 1, 2, \dots$$

Дополним ключевые слова запроса $z^{(1)}$ найденными в словарях Dic_{par} и Dic_{syn} словоизменениями и синонимами. Запишем полученное выражение в виде конъюнктивной нормальной формы:

$$z^{(2)} = D_1 \wedge D_2 \wedge \dots,$$

где $D_1, D_2 \dots$ – ключевые слова или дизъюнкции ключевых слов.

Преобразуем запрос $z^{(2)}$ в дизъюнктивную нормальную форму:

$$z^{(3)} = K_1 \vee K_2 \vee \dots,$$

где $K_1, K_2 \dots$ – конъюнкции ключевых слов.

Каждый конъюнкт $K_i = b_1 \wedge b_2 \wedge \dots$ ($i = 1, 2, \dots$) запроса $z^{(3)}$ представим в виде совокупности $\{b_1, b_2, \dots\}$ всех входящих в него ключевых слов. Таким образом, в результате коррекции исходного запроса (3) получим множество запросов $\{z_1, z_2, \dots\}$.

Коррекция запросов на основе корпусов текстов. В корпусной лингвистике различают статические и динамические корпуса текстов. Примерами статических служат тематические корпуса и полный корпус, являющийся объединением всех тематических. Под динамическим понимают тематический корпус текстов, все документы которого релевантны некоторому тексту или запросу на поиск информации. Динамический корпус создается из релевантных текстов полного корпуса.

Пусть по-прежнему $z = \{a_1, a_2, \dots | a_i \in W_T, i = 1, 2, \dots\}$ – произвольный запрос пользователя информационной системы. Обозначим через Ct_1, Ct_2, \dots тематические корпуса текстов, а через $Cf = Ct_1 \cup Ct_2 \cup \dots$ – полный корпус. Коррекция запроса z на основе корпусов текстов сводится к соотнесению каждого ключевого слова a_i со значением его информативности $I_z^{a_i}$. Запросы являются, как правило, краткими сообщениями. Их объем не позволяет выявить статистические характеристики ключевых слов при вычислении их информативности. Поэтому в данном случае используются релевантные запросам тематические или динамические корпуса текстов.

Поиск релевантного тематического корпуса текстов. Исключим из всех поисковых образцов текстов полного корпуса значения информативности ключевых слов, т. е. поисковый образ каждого текста $t \in T$ представим в виде

$$PO'_{Cf} = \{b_1, b_2, \dots | t \in T, i = 1, 2, \dots\}.$$

При поиске релевантного запросу текста будем использовать векторную модель описания данных [5], а в качестве критерия выдачи – косинус угла между векторами запроса и поискового образа текста [6]. Введем в рассмотрение n -мерное евклидово пространство E ($n = |W_T|$). Для этого лексикографически упорядочим все слова из множества W_T , т. е. представим его в виде кортежа $W_T = \langle c_1, c_2, \dots, c_n \rangle$. Для каждого текста $t \in T$ построим вектор его поискового образа в пространстве E : $\mathbf{F}_t = (p_1, p_2, \dots, p_n)$, где $p_i = 1$, если слово c_i входит в этот поисковый образ, в противном случае $p_i = 0$. Аналогично представим вектор, построенный для запроса z : $\mathbf{F}_z = (q_1, q_2, \dots, q_n)$. Тогда для вычисления косинуса угла между векторами \mathbf{F}_t и \mathbf{F}_z воспользуемся формулой

$$\cos \varphi_{tz} = \frac{\mathbf{F}_t \mathbf{F}_z}{|\mathbf{F}_t| |\mathbf{F}_z|} = \frac{\sum_{i=1}^n p_i q_i}{\sqrt{\sum_{i=1}^n p_i^2} \sqrt{\sum_{i=1}^n q_i^2}}.$$

Обозначим через l количество совпавших ключевых слов поискового образа текста t и запроса z . Пусть также m_t – количество слов в поисковом образе текста t , а m_z – их количество в запросе z . Тогда данный критерий выдачи можно представить в виде

$$\cos \varphi_{tz} = \frac{l}{\sqrt{m_t m_z}}.$$

Приведем описание алгоритма поиска тематического корпуса текстов Ct , релевантного запросу z . На вход алгоритма поступает запрос z , по которому реализуется поиск поисковых образов тематических корпусов текстов в полном корпусе Cf . Результатом поиска считаем корпус Ct , которому соответствует наибольшее из значений критерия выдачи $\cos \varphi_{tz}$, такое, что $\cos \varphi_{tz} \geq \eta_0$. Если корпус Ct не найден, то для текста t необходимо сформировать динамический корпус текстов Dt .

Формирование релевантного динамического корпуса текстов. Для создания динамического корпуса текстов в множестве Cf нужно найти все тексты, релевантные запросу z . Пусть $d \in Cf$ – произвольный документ из полного корпуса текстов. Построим вектор \mathbf{F}_d поискового образа документа d по аналогии с вектором $\mathbf{F}_t : \mathbf{F}_d = (r_1, r_2, \dots, r_n)$. В качестве критерия выдачи будем использовать аналог критерия $\cos \varphi_{tz}$:

$$\cos \psi_{dz} = \frac{\mathbf{F}_d \mathbf{F}_z}{|\mathbf{F}_d| |\mathbf{F}_z|} = \frac{\sum_{i=1}^n r_i q_i}{\sqrt{\sum_{i=1}^n r_i^2} \sqrt{\sum_{i=1}^n q_i^2}}.$$

Множество всех текстов из множества Cf , найденных в соответствии с критерием $\cos \psi_{dz}$, образует динамический корпус текстов Dt .

Вычисление информативности ключевых слов запроса. Информативность каждого ключевого слова a запроса z будем вычислять по формуле из статьи [2]

$$I_z^a = \frac{n_{Kt}^a + n_{Kt}^{Par_a} + n_{Kt}^{Syn_a}}{n_{Cf}^a + n_{Cf}^{Par_a} + n_{Cf}^{Syn_a}}, \quad (2)$$

где Kt – релевантный запросу z тематический или динамический корпус текстов, n_{Kt}^a и n_{Cf}^a – частоты встречаемости слова a в корпусах текстов Kt и Cf соответственно;

$n_{Kt}^{Par_a}$ – число вхождений всех слов корпуса Kt , являющихся словоизменениями словоформы a :

$$n_{Kt}^{Par_a} = \sum_{b \in Par_a, b \neq a} n_{Kt}^b;$$

$n_{Kt}^{Syn_a}$ – количество синонимов словоформы a в корпусе Kt :

$$n_{Kt}^{Syn_a} = \sum_{c \in Syn_a, c \neq a} n_{Kt}^c.$$

Аналогичный смысл имеют параметры $n_{Cf}^{Par_a}$ и $n_{Cf}^{Syn_a}$.

В результате коррекции запроса z путем соотнесения его ключевых слов со значениями их информативности получим новый запрос

$$z' = \{(a_1, I_z^{a_1}), (a_2, I_z^{a_2}), \dots | a_i \in W_{Cf}, i = 1, 2, \dots\}.$$

Разбиение исходных запросов. Данный вариант коррекции запросов будем использовать, если они состоят из нескольких предложений.

Классификация предложений запроса. Рассмотрим запрос $z = \langle \rho_1, \rho_2, \dots, \rho_l \rangle$, где $\langle \rho_1, \rho_2, \dots, \rho_l \rangle$ – кортеж предложений. Процессу разбиения кортежа предложений на классы предшествуют процедуры вычисления информативности вербальной ассоциации между словами предложений и между самими предложениями.

Информативность вербальной ассоциации между словами. Обозначим через W множество всех словоформ полного корпуса текстов Cf , а через \prec_w – отношение строгого порядка на W (транзитивное и антирефлексивное бинарное отношение). Определим, кроме того, на множестве W антирефлексивное и антисимметричное бинарное отношение Θ , такое, что любая пара слов (a, b) из множества W является элементом отношения Θ тогда и только тогда, когда слова a и b из этой пары содержатся хотя бы в одном предложении корпуса Cf и выполняется соотношение $a \prec_w b$. Отношение Θ назовем отношением вербальной ассоциации слов в полном корпусе текстов Cf .

Информативность вербальной ассоциации между произвольными словами a и b некоторого предложения определим как вероятность его появления в корпусе Cf . При практической реализации информационной системы под указанной информативностью будем понимать дробь

$$I_{Cf}^{ab} = n_{Cf}^{ab} / N_{Cf}, \quad (3)$$

где n_{Cf}^{ab} – количество всех предложений в полном корпусе текстов Cf , в которых присутствуют слова a и b или их синонимы и словоизменения, а N_{Cf} – количество всех предложений в корпусе Cf . В развернутом виде формулу (3) можно переписать, используя информацию, которую содержат специальные лингвистические словари:

– частотный словарь словоформ

$$Dic_a = \{ \langle a, n_{Cf}^a, n_{Ct_1}^a, n_{Ct_2}^a, \dots, n_{Ct_n}^a \rangle \mid a \in W_{Cf} \},$$

в котором каждой словоформе приписаны частоты ее встречаемости $n_{Cf}^a, n_{Ct_1}^a, n_{Ct_2}^a, \dots, n_{Ct_n}^a$ во всех корпусах текстов;

– словарь вербально-ассоциативных пар слов

$$Dic_{ab} = \{ \langle (a, b), I_{Cf}^{ab} \rangle \mid a, b \in \pi, \pi \in Cf \},$$

в котором каждой паре слов поставлена в соответствие информативность их вербальной ассоциации.

С учетом информации из лингвистических словарей формулу (3) представим в виде

$$I_{Cf}^{ab} = \frac{n_{Cf}^{ab} + n_{Cf}^{Par_{ab}} + n_{Cf}^{Syn_{ab}}}{N_{Cf}}. \quad (4)$$

Параметр $n_{Cf}^{Par_{ab}}$ в формуле (4) указывает на число вхождений всех пар словоформ, являющихся словоизменениями соответственно слов a и (или) b и встречающихся в одном и том же предложении корпуса текстов Cf :

$$n_{Cf}^{Par_{ab}} = \sum_{\substack{c \in Par_a, d \in Par_b, \\ c \neq a \text{ и (или) } d \neq b, \\ (c, d) \in \Theta}} n_{Cf}^{cd}.$$

Аналогичное выражение справедливо для параметра $n_{Cf}^{Syn_{ab}}$:

$$n_{Cf}^{Syn_{ab}} = \sum_{\substack{d \in Syn_a, f \in Syn_b, \\ d \neq a \text{ и (или) } f \neq b, \\ (d, f) \in \Theta}} n_{Cf}^{df}.$$

Информативность вербальной ассоциации между предложениями и текстами. Пусть π и ρ – два предложения (или два текста) из корпуса Cf , а W_π и W_ρ – соответственно множества всех словоформ в этих предложениях, дополненные всеми синонимами и всеми словоизменениями из словарей Dic_{par} и Dic_{syn} . Построим вектор в пространстве E :

$$\mathbf{I}_{Cf}^{\pi\rho} = (I_{Cf}^{a_1b_1}, I_{Cf}^{a_2b_2}, \dots, I_{Cf}^{a_nb_n}). \quad (5)$$

В формуле (5) значение информативности $I_{Cf}^{a_nb_n}$ для любого $i \in \{1, 2, \dots, n\}$ определяется из словаря вербально-ассоциативных пар слов Dic_{ab} , если $(a_i, b_i) \in \Theta$ и выполняется хотя бы одно из двух условий: 1) $a_i \in W_\pi, b_i \in W_\rho$; 2) $b_i \in W_\pi, a_i \in W_\rho$. В противном случае $I_{Cf}^{a_nb_n} = 0$.

С учетом рассмотренных обозначений нормализованную информативность $I_{Cf}^{\pi\rho}$ вербальной ассоциации между предложениями (текстами) π и ρ можно интерпретировать как проекцию вектора $\mathbf{I}_{Cf}^{\pi\rho}$ на направление вектора $\mathbf{e} = (1, 1, \dots, 1)$ размерности n , т. е. отношение скалярного произведения векторов $\mathbf{I}_{Cf}^{\pi\rho}$ и \mathbf{e} к длине вектора \mathbf{e} :

$$I_{Cf}^{\pi\rho} = \frac{\mathbf{I}_{Cf}^{\pi\rho} \cdot \mathbf{e}}{\sqrt{n}} = \frac{\sum_{i=1}^n I_{Cf}^{a_ib_i}}{\sqrt{n}}. \quad (6)$$

При программной реализации алгоритма вычисления информативности вербальной ассоциации между предложениями или текстами удобно пользоваться следующей формулой, полученной из выражения (6):

$$I_{Cf}^{\pi\rho} = \frac{I_1 + I_2 + \dots + I_r}{\sqrt{n}}, \quad (7)$$

где I_1, I_2, \dots, I_r – все отличные от нуля координаты вектора $\mathbf{I}_{Cf}^{\pi\rho}$.

Описание алгоритма классификации предложений. Алгоритм разбиения кортежа $z = \langle \rho_1, \rho_2, \dots, \rho_l \rangle$ на классы работает следующим образом.

На начальном этапе в качестве единственного элемента первого класса S_1 будем рассматривать предложение ρ_1 . Затем формируются множества словоформ предложений ρ_1 и ρ_2 и по формуле (7) вычисляется информативность вербальной ассоциации между ними. Если вычисленное значение не меньше некоторой пороговой величины ρ_0 , то предложение ρ_2 помещается в класс S_1 . Далее аналогичным образом вычисляется информативность вербальной ассоциации между предложениями из пар $(\rho_1, \rho_3), \dots, (\rho_1, \rho_l)$. После завершения процесса формирования класса S_1 точно так же формируются и другие классы. В итоге будем иметь совокупность классов $\{S_1, S_2, \dots, S_m\}$ ($m \leq l$).

Индексирование информативных классов предложений. Среди сформированных классов предложений S_1, S_2, \dots, S_m могут быть неинформативные, использование которых в качестве запросов нецелесообразно. В связи с этим рассмотрим вопросы вычисления информативности классов предложений.

Информативность слов из полнотекстовых документов. Пусть T – полнотекстовый документ, объем которого обеспечивает вычисление статистических характеристик его словоформ и предложений. Информативность I_T^a слова a из текста T будем вычислять по формуле, аналогичной выражению (2):

$$I_T^a = \frac{n_T^a + n_T^{Par_a} + n_T^{Syn_a}}{n_{Cf}^a + N_{Cf}^{Par_a} + N_{Cf}^{Syn_a}}. \quad (8)$$

Информативность слов из полнотекстовых документов. Индексированию краткого сообщения предшествует процесс его расширения за счет включения релевантных предложений из полного корпуса текстов.

Рассмотрим краткое текстовое сообщение Q . Обозначим через W_Q множество всех его словоформ. Вычислим информативность $J_{Cf}^{Q\pi}$ вербальной ассоциации между текстом Q и некоторым предложением π из полного корпуса текстов Cf . Построим вектор $\mathbf{J}_{Cf}^{Q\pi} = (J_{Cf}^{c_1d_1}, J_{Cf}^{c_2d_2}, \dots, J_{Cf}^{a_k b_k})$ в евклидовом пространстве. Для вычисления информативности $J_{Cf}^{Q\pi}$ воспользуемся аналогом формулы (7):

$$J_{Cf}^{Q\pi} = \frac{J_1 + J_2 + \dots}{\sqrt{(J_1)^2 + (J_2)^2 + \dots}}, \quad (9)$$

где J_1, J_2, \dots – все отличные от нуля координаты вектора $\mathbf{J}_{Cf}^{Q\pi}$. Если информативность (9) не меньше некоторого критического значения, то предложение π занесем в текст Q . Аналогично поступим и с другими такими предложениями полного корпуса текстов. В результате получим расширенное множество предложений, которое снова будем считать текстом Q .

Информативность I_Q^a любого слова $a \in W_Q$ вычислим по формуле (8):

$$I_Q^a = \frac{n_Q^a + n_Q^{Par_a} + n_Q^{Syn_a}}{n_{Cf}^a + N_{Cf}^{Par_a} + N_{Cf}^{Syn_a}}. \quad (10)$$

Информативность предложений и текстов. При вычислении информативности предложений текста T будем также исходить из их векторного представления: $\mathbf{\Pi} = (I_\pi^{a_1}, I_\pi^{a_2}, \dots, I_\pi^{a_i})$, где $I_\pi^{a_1}, I_\pi^{a_2}, \dots, I_\pi^{a_i}$ – значения информативности слов произвольного предложения π (компонента вектора $\mathbf{\Pi}$ равна нулю, если соответствующего слова нет в предложении π). Тогда аналогично формуле (9) информативность I_T^π предложения π будем вычислять по формуле

$$I_T^\pi = \frac{I_1 + I_2 + \dots}{\sqrt{(I_1)^2 + (I_2)^2 + \dots}}, \quad (11)$$

где I_T^π, I_T^p, \dots – значения информативности всех предложений документа T .

Описание алгоритма индексирования классов предложений. Алгоритм индексирования классов предложений функционирует в три этапа.

На первом этапе вычисляется информативность каждого из классов предложений S_1, S_2, \dots, S_m по формуле (11). Класс будем считать информативным, если значение информативности не меньше некоторой пороговой величины. В результате выполнения первого этапа получим совокупность информативных классов предложений $\{U_1, U_2, \dots, U_s\}$ ($s \leq m$). Классы, имеющие недостаточный объем для вычисления статистических характеристик словоформ (т. е. являющиеся краткими сообщениями), дополняются релевантными предложениями из полного корпу-

са текстов Cf с использованием формулы (9). Полученные в результате такого расширения новые классы будем использовать в качестве запросов.

На втором этапе вычисляется информативность $I_{U_i}^a$ ($i = \overline{1, s}$) всех словоформ из предложений всех классов U_1, U_2, \dots, U_s по формулам (8) и (10).

На третьем этапе формируются поисковые образы

$$\text{ПП}_i = \{(a, I_{U_i}^a); (b, I_{U_i}^b); \dots | a, b, \dots \in U_i\}, \quad i = \overline{1, s},$$

всех классов U_1, U_2, \dots, U_s предложений. Эти поисковые образы будут использованы в качестве поисковых предписаний, полученных после разбиения исходного запроса.

Послепоисковая коррекция запросов. Пусть по-прежнему $z \in Z$ – некоторый запрос пользователя, $\theta(z)$ – его информационная потребность, а $\pi(z)$ – совокупность найденных текстов по данному запросу.

Коррекция запросов на основе оценок pertinентности результатов поиска. Обозначим через t_z текст, полученный путем объединения предложений из всех текстов множества $\pi(z)$, которые пользователь оценил как pertinентные:

$$t_z = \{\rho | \rho \in t, t \in \pi(z), v(t, \theta(z)) = 1\}.$$

Вычислим информативность всех слов текста t_z по формуле, аналогичной выражению (2):

$$I_{t_z}^a = \frac{n_{t_z}^a + n_{t_z}^{Par_a} + n_{t_z}^{Syn_a}}{n_{Cf}^a + n_{Cf}^{Par_a} + n_{Cf}^{Syn_a}}. \quad (12)$$

Исключим из запроса z все ключевые слова, не входящие в текст t_z , а также слова, информативность которых, вычисленная по формуле (8), меньше некоторого порогового значения. В итоге получим откорректированный запрос

$$z^+ = \{b_1, b_2, \dots | b_i \in z \cup t_z, I_{t_z}^a \geq I^0, i = 1, 2, \dots\}. \quad (13)$$

Синтез поискового предписания из запроса (9) сводится к приписыванию каждому ключевому слову запроса z^+ его информативности:

$$z^{++} = \{(b_1, I_{t_z}^{a_1}), (b_2, I_{t_z}^{a_2}), \dots | b_i \in z \cup t_z, I_{t_z}^a \geq I^0, i = 1, 2, \dots\}.$$

Коррекция запросов на основе замены их pertinентными текстами-образцами. Пусть $z_1 = \{a_1, a_2, \dots | a_i \in W_{Cf}, i = 1, 2, \dots\}$ – запрос пользователя, а $Cf = Ct_1 \cup Ct_2 \cup \dots$ – полный корпус текстов. Послепоисковую коррекцию запроса z_1 реализуем в три этапа.

На первом этапе проведем допоисковую коррекцию запроса z_1 на основе корпусов текстов. В результате получим запрос

$$z_2 = \{(a_1, I_{z_2}^{a_1}), (a_2, I_{z_2}^{a_2}), \dots | a_i \in W_{Cf}, i = 1, 2, \dots\},$$

где $I_{z_2}^{a_1}, I_{z_2}^{a_2}, \dots$ – значения информативности ключевых слов a_1, a_2, \dots соответственно.

Рассмотрим реализацию второго этапа послепоисковой коррекции запроса пользователя. Пусть $Lex = \langle c_1, c_2, \dots, c_n \rangle$ – множество всех различных слов полного корпуса текстов Cf , а E – n -мерное евклидово пространство. Построим в E вектор запроса z_2 : $\mathbf{F}_{z_2} = (r_1, r_2, \dots, r_n)$, где $r_i = I_{z_2}^{a_i}$, если слово r_i входит в запрос z_2 , в противном случае $r_i = 0$. Аналогично для каждого текста $t \in T$ построим вектор его поискового образа в пространстве E : $\mathbf{F}_t = (s_1, s_2, \dots, s_n)$.

На третьем этапе коррекции запроса пользователя осуществим поиск релевантных текстов в соответствии с критерием выдачи

$$\cos \varphi_{tz_2} = \frac{\mathbf{F}_t \mathbf{F}_{z_2}}{|\mathbf{F}_t| |\mathbf{F}_{z_2}|} = \frac{\sum_{i=1}^n s_i r_i}{\sqrt{\sum_{i=1}^n s_i^2} \sqrt{\sum_{i=1}^n r_i^2}}.$$

Множество $\pi(z_2)$ текстов, найденных по запросу z_2 , предъявляется пользователю, который выбирает наиболее пертинентный текст $t^+ \in \pi(z_2)$. Поисковый образ $\omega(t^+)$ текста t^+ приобретает статус откорректированного запроса:

$$z_3 = \{(c_1, I_{t^+}^{c_1}), (c_2, I_{t^+}^{c_2}), \dots\}.$$

Заключение. Разработана математическая модель коррекции запросов пользователей в системе информационной поддержки принятия решений. В рамках модели предложены формулы для вычисления информативности слов, предложений и текстов, а также вербальных ассоциаций между ними. Рассмотрены два вида взаимодействия пользователей с информационной системой при коррекции запросов: допоисковое и послепоисковое. В процессе допоискового взаимодействия пользователю предъявляются сведения, которые используются при формулировании и коррекции запросов. Послепоисковое взаимодействие основано на оценке и использовании на последующих этапах поиска промежуточной выдачи найденной информации.

Разработаны алгоритмы допоисковой коррекции запросов путем дополнения ключевых слов запросов словоизменениями и синонимами, применения корпусов текстов для вычисления информативности ключевых слов и разбиения сложного исходного запроса на несколько простых.

Предложены алгоритмы послепоисковой коррекции запросов на основе оценок пертинентности результатов поиска и на основе замены исходных запросов пертинентными текстами-образцами.

Список использованных источников

1. Савотченко, С. Е. Современные аспекты повышения пертинентности результатов информационного поиска в глобальной сети [Электронный ресурс] / С. Е. Савотченко, Е. А. Проскурина. – Режим доступа: <https://fundamental-research.ru/ru/article/view?id=34639>. – Дата доступа: 08.02.2023.
2. Липницкий, С. Ф. Моделирование информационного поиска на основе динамических корпусов текстов / С. Ф. Липницкий, А. А. Мамчич // Вес. Нац. акад. навук Беларусі. Сер. фіз.-тэхн. навук. – 2011. – № 1. – С. 72–81.
3. Липницкий, С. Ф. Веб-поиск и адресное распространение информации на основе моделирования вербальных ассоциаций / С. Ф. Липницкий // Информатика. – 2019. – № 3. – С. 79–88.
4. Мальцев, А. И. Алгебраические системы / А. И. Мальцев. – М.: Наука, 1970. – 392 с.
5. Ландэ, Д. В. Поиск знаний в Internet. Профессиональная работа : пер. с англ. / Д. В. Ландэ. – М.: Диалектика-Вильямс, 2005. – 272 с.
6. Липницкий, С. Ф. Синтез запросов и поиск альтернатив в системе информационной поддержки принятия решений / С. Ф. Липницкий // Проблемы физики, математики и техники. – 2020. – № 2. – С. 91–95.

References

1. Savotchenko S. E., Proskurina E. A. *Sovremennye aspekty povysheniya pertinentnosti rezul'tatov informacionnogo poiska v global'noj seti. Modern Aspects of Increasing the Pertinence of Information Search Results in the Global Network* (In Russ.). Available at: <https://fundamental-research.ru/ru/article/view?id=34639> (accessed at 02.08.2023).

2. Lipnitsky S. F., Mamchich A. A. *Modeling information retrieval based on dynamic text corpora*. Vestsi Natsyyanal'nai akademii navuk Belarusi. Seryya fizika-technichnykh navuk [*Proceedings of the National Academy of Sciences of Belarus. Physical-technical Series*], 2011, no. 1, pp. 72–81 (In Russ.).
3. Lipnitsky S. F. *Web search and targeted dissemination of information based on the modeling of verbal associations*. Informatika [*Informatics*], 2019, no. 3, pp. 79–88 (In Russ.).
4. Maltsev A. I. *Algebraicheskie sistemy. Algebraic Systems*. Moscow, Nauka, 1970, 392 p. (In Russ.).
5. Lande D. V. *Poisk znaniy v Internet. Professional'naja rabota. Knowledge Search in Internet. Professional Work*. Moscow, Dialektika-Viliams, 2005, 272 p.
6. Lipnitskiy S. F. *Synthesis of queries and search for alternatives in the system of information support for decision-making*. Problemy fiziki, matematiki i tehniki [*Problems of Physics, Mathematics and Technology*], 2020, no. 2, pp. 91–95 (In Russ.).

Информация об авторе

Липницкий Станислав Феликсович, доктор технических наук, главный научный сотрудник, Объединенный институт проблем информатики Национальной академии наук Беларуси.
E-mail: lipn@newman.bas-net.by

Information about the author

Stanislav F. Lipnitsky, D. Sc. (Eng.), Chief Researcher, The United Institute of Informatics Problems of the National Academy of Sciences of Belarus.
E-mail: lipn@newman.bas-net.by



УДК 004.42
<https://doi.org/10.37661/1816-0301-2023-20-2-96-110>

Оригинальная статья
Original Paper

Методика и программное средство поиска аномалий в данных телеметрии солнечной электростанции на основе анализа нормализованных значений мощности

С. В. Валеви́ч¹, К. С. Дик^{1✉}, И. И. Пилецкий¹, И. Крузе³, Р. М. Асимов², В. С. Осипович⁴

¹Белорусский государственный университет информатики и радиоэлектроники, ул. П. Бровки, 6, Минск, 220013, Беларусь
✉E-mail: kanstantind@gmail.com

²ООО «Сенсоэлектроника», ул. Кульман, 9, Минск, 220100, Беларусь

³Sunsniffer, LTD, Людвиг-Фейербах-Штрассе, 69, Нюрнберг, 90489, Германия

⁴Novotech Lab INC, 360 Кинг Стрит, Чарльстон, Южная Каролина, 29401, США

Аннотация

Цели. В связи с увеличением количества солнечных электростанций актуальной задачей становится автоматизация мониторинга их работоспособности. Поиск аномалий в работе солнечных электростанций – одна из основных составляющих мониторинга. Цель исследования заключается в разработке новых методик и программных алгоритмов поиска аномалий в работе солнечных панелей на основе результатов работы цифрового двойника, созданного и обученного по данным телеметрии солнечной электростанции.

Методы. Методика основана на статистических исследованиях отклонений значений мощности в точке максимально эффективной работы солнечной панели, рассчитанной цифровым двойником. Кроме того, введено нормализованное значение мощности максимально эффективной работы солнечной панели для более точной кластеризации и поиска аномалий.

Результаты. С помощью разработанной методики поиска аномалий в данных телеметрии за полгода наблюдений обнаружены 18 аномалий в работе солнечных панелей электростанции. Все случаи проанализированы на предмет причин возникновения аномалий в работе солнечных панелей.

Заключение. Установлено, что при использовании в анализе отклонений нормализованных значений мощностей в точке максимальной мощности P_N возможно обнаружение аномальной работы отдельных панелей. Рассчитан уровень отклонения нормализованных значений в точке максимальной мощности, свидетельствующий о наличии аномалии в работе солнечной панели.

Ключевые слова: солнечная панель, нормализованное значение мощности, поиск аномалий, точка максимальной мощности, солнечная электростанция, телеметрия

Для цитирования. Методика и программное средство поиска аномалий в данных телеметрии солнечной электростанции на основе анализа нормализованных значений мощности / С. В. Валеви́ч [и др.] // Информатика. – 2023. – Т. 20, № 2. – С. 96–110. <https://doi.org/10.37661/1816-0301-2023-20-2-96-110>

Конфликт интересов. Авторы заявляют об отсутствии конфликта интересов.

Methods and software for anomalies searching in the telemetry data of a solar power plant based on the normalized power analysis

Sergey V. Valevich¹, Konstantine S. Dzick^{1✉}, Ivan I. Pilecki¹, Ingmar Kruse³, Roustam M. Asimov², Vitali S. Asipovich⁴

¹Belarusian State University of Informatics and Radioelectronics,
st. P. Brovki, 6, Minsk, 220013, Belarus

✉E-mail: kanstantind@gmail.com

²Sensotronica LTD,
st. Kulman, 9, Minsk, 220100, Belarus

³Sunsniffer, LTD,
Ludwig-Feuerbach-Straße, 69, Nürnberg, 90489, German

⁴Novotech Lab INC,
360 King Street, Charleston, SC, 29401, USA

Abstract

Objectives. In connection with the increase in the number of solar power plants, the automation of monitoring their performance becomes an urgent task. The search for anomalies in the operation of solar power plants is one of the main components of monitoring. The purpose of the study is to develop new methods and software algorithms for finding anomalies in the operation of solar panels based on the results of a digital twin created and trained according to the telemetry data of a solar power plant.

Methods. The developed technique is based on statistical studies of deviations of power values at the point of maximum efficient operation of the solar panel calculated by the digital twin. In addition, a normalized value of the power in the maximum efficient operation of the solar panel was introduced for more accurate clustering and anomaly search.

Results. Using the developed method of static search for half a year of observations, 18 anomalies were detected in the operation of the solar panels of the power plant. All cases are analyzed for the causes of anomalies in the operation of solar panels.

Conclusion. It has been established that when using normalized power values in the analysis of deviations at the point of maximum power PN, it is possible to detect abnormal operation of individual panels. The level of deviation of the normalized values at the point of maximum power was calculated, indicating the presence of an anomaly in the operation of solar panel.

Keywords: solar panel, normalized power value, anomaly search, maximum power point, solar power plant, telemetry

For citation. Valevich S. V., Dzick K. S., Pilecki I. I., Kruse I., Asimov R. M., Asipovich V. S. *Methods and software for anomalies searching in the telemetry data of a solar power plant based on the normalized power analysis*. Informatika [Informatics], 2023, vol. 20, no. 2, pp. 96–110 (In Russ.).
<https://doi.org/10.37661/1816-0301-2023-20-2-96-110>

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

Введение. Рост объемов выработки электроэнергии посредством солнечных электростанций стимулирует развитие систем мониторинга их технического состояния, а также способов автоматизации поиска неисправных солнечных панелей и фактов их аномального функционирования.

Для мониторинга технического состояния используют облачные ресурсы (URL: <https://www.solarfeeds.com/mag/solar-monitoring-systems-in-the-world/>, <http://www.sunsniffer.de/solution/what-is-sunsniffer.html>, <https://www.solareye.eu/platform/?r=site/page&view=features>, <http://www.pvsyst.com/en/>), которые позволяют собирать данные телеметрии при оснащении солнечных электростанций соответствующим оборудованием, осуществлять через веб-интерфейс анализ основных параметров работы электростанции в целом и конкретных панелей в частности. Как правило, в облачных сервисах сохраняются значения напряжения, температу-

ры, силы тока каждой панели и освещенности для всех панелей электростанции. Обнаружение аномалий в работе конкретных солнечных панелей и их классификация зависят от внимания и квалификации пользователя облачного ресурса, а также от наличия у ресурса функций по обнаружению аномалий в работе и неисправных солнечных панелей.

Для решения задачи поиска аномалий в работе солнечных панелей с использованием данных телеметрии ряд исследователей предлагают осуществлять моделирование солнечных панелей [1–3]. Другие авторы [4–10] используют методики и алгоритмы оценки так называемой точки максимальной мощности (Maximum Power Point, MPP) – комбинации силы тока и напряжения панели, при которых съем энергии является оптимальным, – для солнечных панелей электростанции при определенных условиях. Однако эти методики либо требуют установки дополнительного оборудования с извлечением дополнительных данных для применения методик и алгоритмов, либо не учитывают того факта, что солнечные панели подключены в цепочки последовательно и могут работать не в своем оптимальном режиме.

Отмеченные недостатки методик моделирования работоспособности солнечных панелей устранены авторами [11–14]. Результаты работы цифрового двойника позволяют использовать расчетные значения мощности MPP и вольт-амперные характеристики для каждой панели при поиске аномалий в работе солнечных панелей.

Цель настоящего исследования – разработка методики и программного средства обнаружения аномалий в работе солнечных панелей по результатам телеметрии, представленной облачным ресурсом (URL: <http://www.sunsniffer.de/solution/what-is-sunsniffer.html>), на основании мощности MPP, полученной с использованием цифрового двойника солнечной электростанции.

Методика эксперимента. В ходе исследования использовались результаты телеметрии с электростанции Südstadt-Forum, расположенной в Нюрнберге, Германия. В состав установки входят три инвертора (модели SUN2000-20KTL, Sinvert PVM17 и Sinvert PVM20) с 16 последовательностями (цепочки солнечных панелей) и 287 солнечных панелей. Все цепочки состоят из 18 фотоэлектрических монокристаллических модулей M190 (STORM Energy GmbH, Германия).

Платформа цифрового двойника предоставляет программный интерфейс API (Application Programming Interface) [14, 15], который принимает данные мониторинга за определенный период и возвращает параметры физико-математической модели для всех солнечных панелей. Входные данные для API включают в себя следующие параметры: напряжение, силу тока, температуру в корпусе солнечной панели, уровень освещенности, временную метку. Интервал фиксации перечисленных параметров 2 мин. Набор данных собирался с помощью телеметрии с июня по ноябрь 2019 г. включительно.

Результатом работы цифрового двойника (API) являются параметры, определенные в смоделированных стандартных условиях испытаний (Standard Test Conditions, STC) для каждой солнечной панели электростанции: выходная мощность в MPP, напряжение и сила тока в MPP, последовательное и параллельное электрическое сопротивление, ток короткого замыкания и напряжение холостого хода.

Для реализации и проверки предложенной методики использовался язык программирования Node.JS вместе с распараллеливанием `async/await` для ускорения вычислений с большим количеством солнечных панелей.

Общий анализ с помощью критериальной методики применяется для представления и изучения данных, удобных для обнаружения аномалий. Такая методика может помочь выявить некоторые проблемы на уровне цепочек или конкретных солнечных панелей на основе локальных критериев.

Исходные точки данных, полученные в результате расчета цифровым двойником, имеют следующие особенности: мощность MPP в STC, идентификатор модуля, календарный месяц. Для анализа данных использовались нормализованные значения мощностей в MPP P_N для всех солнечных панелей, которые рассчитывались по формуле

$$P_N = \frac{P_{mpp} - P_m}{P_m} \cdot 100 \%, \quad (1)$$

где P_m – глобальное медианное значение мощности, Вт. Вычисляется как среднее значение по всем медианам строк:

$$P_m = \frac{\sum_{i=1}^n P_S}{n}, \quad (2)$$

где P_S – медианное значение мощности P_{mpp} для отдельной цепочки последовательно подключенных солнечных панелей, n – количество цепочек в электростанции, i – номер цепочки солнечных панелей.

Чтобы структурировать результаты расчетов и однозначно выделить аномалии, значения P_N были разделены на следующие категории:

- минимальное значение;
- 25-й перцентиль;
- 50-й перцентиль;
- 75-й перцентиль;
- максимальное значение.

Для представления 50-го перцентилья распределения использовались два метода оценки: медианы и центра масс. Оба метода были проанализированы визуально по графикам, чтобы оценить, какой из них наиболее информативен. Такое статистическое структурирование данных позволяет визуализировать выбросы при выполнении анализа на уровне цепочки или отдельно взятой солнечной панели.

Для анализа данные были объединены в два представления: одноточечные и линейные. Одноточечные данные содержат только одно агрегированное значение за весь полугодовой период. Линейные данные имеют шесть точек: по одной точке на каждый месяц (июнь – ноябрь 2019 г.). Одноточечные данные позволяют грубо оценить общую производительность конкретной солнечной панели, линейные данные – аномалии в динамике. Помимо расчета глобального P_N по формуле (1) был осуществлен еще расчет локального P_{N1} по этой же формуле. Однако для расчета P_{N1} использовались медианы, рассчитанные по цепочкам панелей, относящимся к конкретному инвертору.

Для дальнейшего анализа данные структурируются по нескольким этапам группировки: начиная со всех модулей вместе, затем по группам, прикрепленным к конкретному типу инвертора, и, наконец, по панелям, соединенным в одну цепочку.

Этапы группировки:

- 1) все модули;
- 2) модули по типу инвертора (группа 1 – Huawei SUN2000-20KTL, группа 2 – Siemens Sinvert PVM17, группа 3 – Siemens Sinvert PVM20);
- 3) модули, подключенные в цепочки.

На первом этапе все точки разделяются на три категории в зависимости от типа инвертора. Следующий этап включает 16 категорий (одна категория на одну цепочку солнечных панелей).

При анализе данных внутри групп используются два типа критериев: глобальные и локальные. Глобальные критерии учитывают все нормализованные значения мощностей в MPP P_N для панелей солнечной электростанции при создании эталонного значения. Напротив, локальные критерии рассматривают только текущую группу (например, конкретную цепочку или инвертор). Кроме того, нормализация данных применяется для устранения смещения номинальных мощностей между солнечными панелями. Это позволяет лучше визуализировать аномалии и определять их количественно в процентах. Медиана распределения P_{mpp} использовалась в качестве эталонного значения для нормализации.

Последовательность группировки, представленная выше, позволяет начать поиск выбросов на уровне всей электростанции, а затем на каждом этапе группирования постепенно сужать область анализа для выявления локальных проблем с отдельными солнечными панелями.

Результаты исследования. Анализ основан на нескольких этапах группировки и представляет локализованные результаты на каждом этапе. На первом этапе осуществляется анализ среднего значения мощности солнечных панелей в МРР и в состоянии STC (P_{mpp}). На рис. 1 красным цветом показана кривая со средними значениями P_{mpp} за полгода.

Среднее значение P_{mpp} для всех точек составляет 176,89 Вт. Порог отклонения 10 Вт от среднего будем считать признаком проявления аномальной работы солнечной панели. (Этот порог установлен экспериментально, разработанное программное средство обеспечивает возможность его изменения.) В результате применения методики обнаружены четыре солнечные панели, выходящие за установленный порог: Module 1.4_1 (200,01 Вт), Module 1.4_8 (188,18 Вт), Module 1.11_15 (158,7 Вт) и Module 2.3_10 (130,5 Вт). Эти две отрицательные и две положительные аномалии видны на графике.

Для того чтобы определить, были обнаруженные аномалии в работе солнечных панелей постоянными или проявились единожды, осуществлен расчет среднего значения P_{mpp} для каждого месяца и проанализированы минимальные и максимальные значения в сопоставлении с общим средним значением (рис. 1).

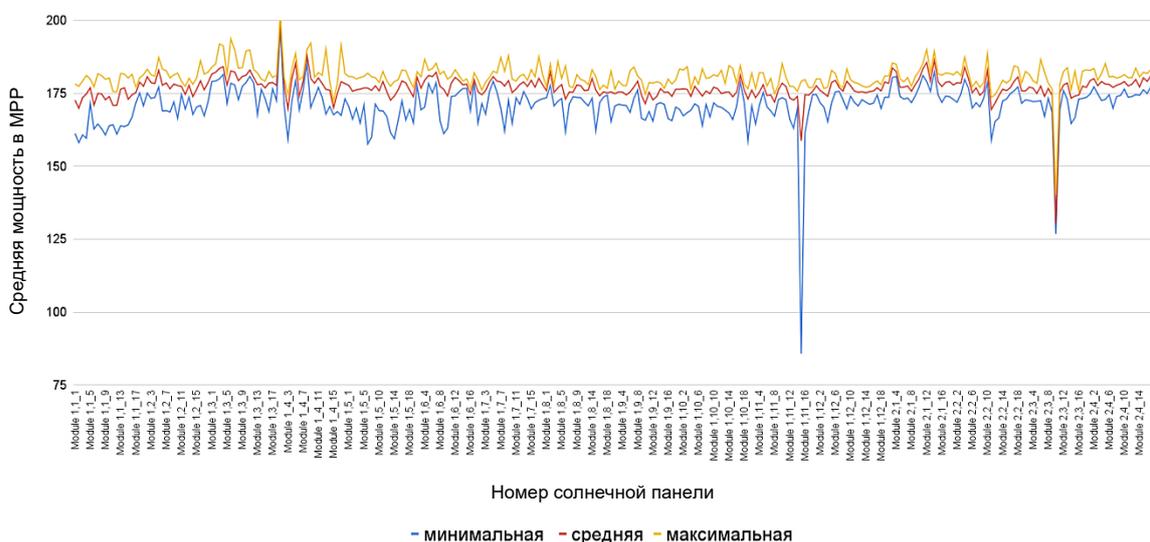


Рис. 1. Средняя мощность в МРР для солнечных панелей с июня по ноябрь 2019 г.
Синим цветом отмечено минимальное среднеемесячное значение, оранжевым – максимальное среднеемесячное значение, красным – среднее за полгода значение

Fig. 1. Average power per point at the MPP for solar panels from June to November 2019. Blue marks the minimum average monthly value, orange – the maximum average monthly value, red – the average value for half a year

Рис. 1 позволяет визуально классифицировать паттерны аномалий за шесть месяцев наблюдения и выработать критерии для алгоритма автоматического поиска аномалий в работе солнечных панелей. Видно, что отклонение в работе солнечных панелей Module 2,3_10, Module 1.4_1 и Module 1.4_8 имеет постоянный характер. Например, среднее значение P_{mpp} для солнечной панели Module 2,3_10 составляет 130,51 Вт, а минимальное значение – 126,78 Вт. Напротив, солнечная панель с Module 1,11_15 демонстрирует аномальное поведение единожды с минимальным значением P_{mpp} в 85,72 Вт, что скорее всего и приводит к среднему значению 158,72 Вт.

Далее был осуществлен анализ линейных данных, где нормированные значения P_N для всех модулей были рассчитаны с помощью формул (1) и (2). Первый подход к кластеризации заключается в отнесении солнечных панелей, отклонение P_N которых от глобальной медианы превышает определенный предел, к отдельному кластеру. Выбор одного порога дает два кластера: нормальные и аномальные солнечные панели. При использовании в кластеризации порогового значения $\pm 10\%$ были получены восемь солнечных панелей, которые относятся к кластеру аномальных (разработанное программное средство обеспечивает изменение этого порогового

значения). Для более детального и наглядного анализа солнечные панели были сгруппированы по типу инвертора, к которому их подключили. Результаты для группы 1 (цепочки панелей, подключенных к инвертору SUN2000-20KTL) представлены на рис. 2. В группе 1 в кластер аномальных попадают пять солнечных панелей: Module 1,1_2 (10,8 % в августе), Module 1,1_4 (- 10 % в августе), Module 1,4_1 (все шесть месяцев, от +17 до +20 %), Module 1,4_1 1.4_3 (- 10,1 % в ноябре) и Module 1,4_1 1.5_6 (- 11,1 % в ноябре). Интересно, что Module 1,4_1 постоянно показывает положительные значения отклонения в течение всех шести месяцев. Это можно объяснить значительно лучшими условиями установки или просто аппаратным артефактом мониторинга.

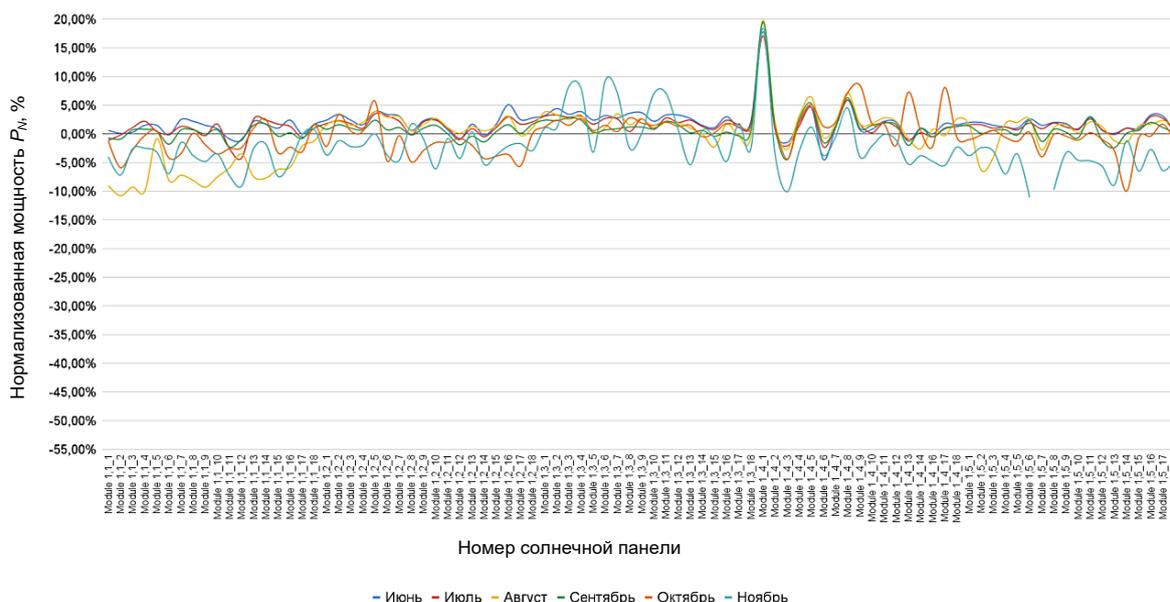


Рис. 2. Нормализованные значения мощностей в MPP P_N для солнечных панелей группы 1 по месяцам в 2019 г.

Fig. 2. Normalized power values in MPP P_N for group 1 solar panels by month in 2019

В группу 2 входят цепочки солнечных панелей, подключенных к инвертору Sinvert PVM17 (рис. 3). Солнечные панели группы 2 не имеют аномалий согласно установленному критерию (отклонение $P_N \pm 10\%$). Некоторые наблюдаемые отклонения могут быть вызваны естественным загрязнением или очисткой модулей.

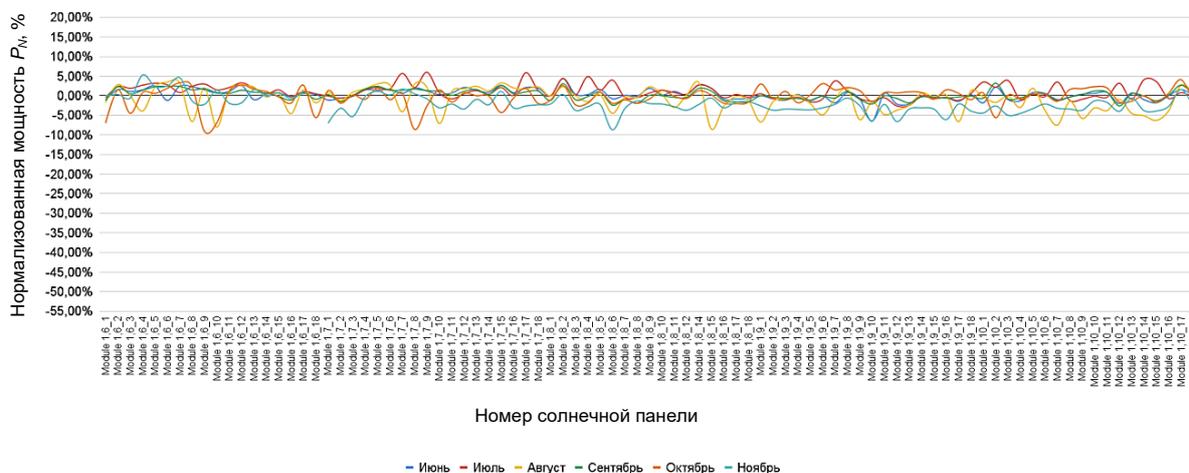


Рис. 3. Нормализованные значения мощностей в MPP P_N для солнечных панелей группы 2 по месяцам в 2019 г.

Fig. 3. Normalized capacities in MPP P_N for group 2 panels by month in 2019

В группу 3 входят цепочки солнечных панелей, подключенных к инвертору Sinvert PVM20 (рис. 4).



Рис. 4. Нормализованные значения мощностей в MPP P_N для солнечных панелей группы 3 по месяцам в 2019 г.

Fig. 4. Normalized power values in MPP P_N for group 3 solar panels by month in 2019

В группе 3 три солнечные панели можно отнести к аномальным: Module 1.11_1 (– 10,4 % в ноябре), Module 1.11_15 (– 51,6 % в ноябре) и Module 2.3_10 (более – 20 % для всех шести месяцев). Следует отметить, что аномалия Module 1,11_15 носит временный характер. Отклонение в – 51,6 % зарегистрировано только в ноябре. Анализ кривых освещенности и силы тока показывает, что причина заключается в попадании солнечной панели в тень практически на весь день во второй половине ноября.

Аномальная работа солнечной панели Module 2,3_10 сохраняется все шесть месяцев (– 28,2 % в июне, – 25,5 % в июле, – 28,5 % в августе, – 28,1 % в сентябре, – 26,9 % в октябре и – 21 % в ноябре).

В целом анализ показывает, что только модули, подключенные к первому и третьему инверторам, имеют аномалии. Производительность модулей двух инверторов все время находится в пределах допустимого диапазона отклонений. Это означает, что группировка по типу инвертора может дать информацию о том, как распределение аномалий соотносится с инверторами.

Результаты анализа отклонений расчетных P_N для конкретных панелей от глобального медианного значения P_N рассмотрим на примере двух цепочек солнечных панелей с номерами 2.3 и 2.4 (рис. 5 и 6) для инвертора группы 3 в 2019 г. Маркеры в виде цветных кругов и квадратов указывают границы выбранных статистических категорий, а цветными линиями отражены нормализованные значения мощности в MPP P_N цепочки солнечных панелей (см. легенду на рис. 5–8).

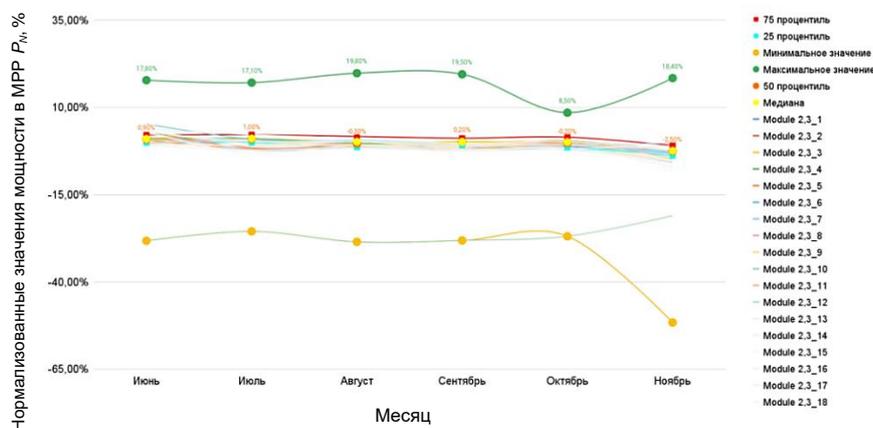


Рис. 5. Нормализованные значения мощности в MPP P_N солнечных панелей цепочки String 2.3

Fig. 5. Normalized power values in MPP P_N photovoltaic modules of String 2.3

Видно, что большинство кривых для солнечных панелей тяготеют к центру распределения, показанному с критериями медианы и центра масс. Кривые солнечных панелей, которые демонстрируют аномальную работу, либо значительно выше, либо ниже медианы и центра масс (например, Module 2,3_10).

Для солнечной панели Module 2,3_10 в цепочке солнечных панелей с номером String 2.3 почти всегда (кроме ноября) расчетная мощность в MPP P_N имеет минимальное значение. Она показывает отклонение относительно центра около 25 % в течение всех шести месяцев.

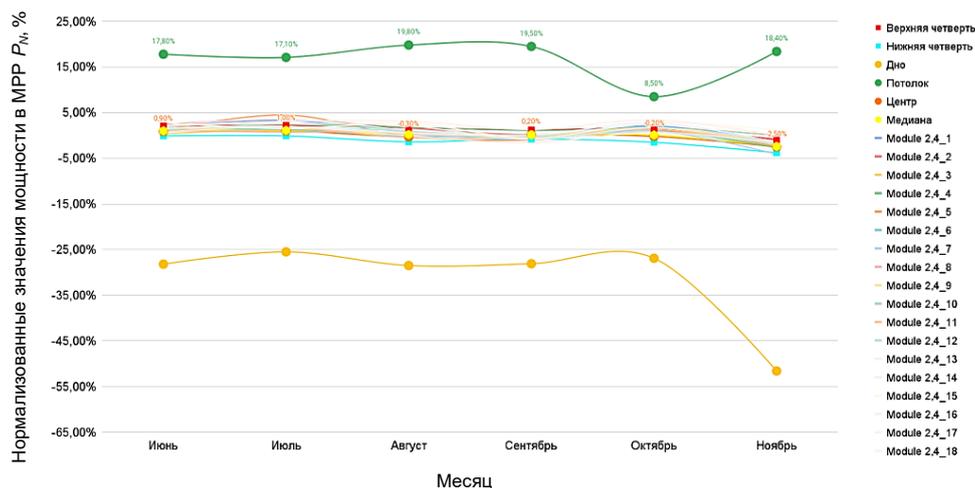


Рис. 6. Нормализованные значения мощности в MPP P_N солнечных панелей цепочки String 2.4

Fig. 6. Normalized power values in MPP P_N photovoltaic modules of String 2.4

Тот же анализ, что и для строки 2.4, показывает типичные результаты для исправной строки. Отклонение всех модулей от центра распределения остается в пределах $\pm 4,5 \%$.

Можно сделать интересное заключение, что почти нет видимой разницы между центрами распределения, определенными критериями медианы и центра масс. Это можно объяснить наличием значительного количества модулей, используемых для расчета глобальных критериев, что дает статистически надежный результат.

Две цепочки были проанализированы путем расчета нормализованных значений мощности в MPP P_{N1} с использованием локальных медиан – значений, полученных от солнечных панелей группы 3. Результаты представлены на рис. 7 и 8. Маркеры указывают границы выбранных статистических категорий для инвертора группы 3 в 2019 г.

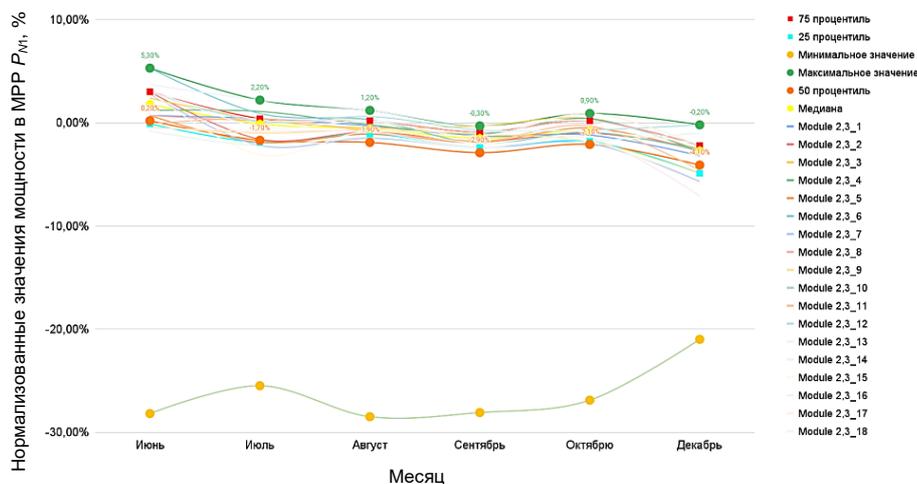


Рис. 7. Нормализованные значения мощности в MPP P_{N1} солнечных панелей цепочки String 2.3

Fig. 7. Normalized power values in MPP P_{N1} photovoltaic modules String 2.3

На рис. 7 видно, что при отмеченном способе анализа неисправная солнечная панель Module 2,3_10 находится прямо у критериев минимального значения и показывает отклонения порядка – 25 % за весь период наблюдения.

Следует заметить, что центр распределения, определяемый средним значением, остается в пределах большинства кривых, хотя расположение центра масс значительно смещено в сторону кривой солнечной панели, которая демонстрирует аномальное поведение (Module 2,3_10). Поэтому использование медианы предпочтительнее для анализа локальных критериев из-за меньшей статистической выборки.

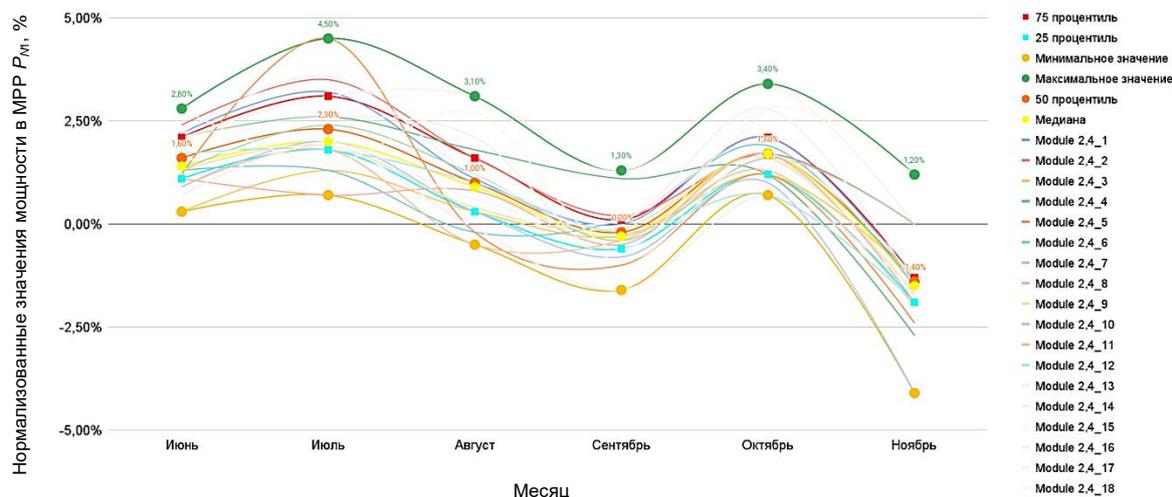


Рис. 8. Нормализованные значения мощности в MPP P_{M1} солнечных панелей цепочки String 2.4

Fig. 8. Normalized power values in MPP P_{M1} photovoltaic modules String 2.4

Анализ локальных критериев для цепочки 2.4, в которой солнечные панели работают одинаково, дает результаты с малым отклонением от результатов расчета глобальных критериев. Отклонение всех солнечных панелей от центра распределения остается в пределах $\pm 4,5\%$.

Подчеркнем, что все исправные цепочки в электростанции демонстрируют поведение, подобное описанному выше. Таким образом, исправная цепочка характеризуется относительным разбросом производительности в пределах $\pm 4,5\%$ независимо от глобальных или локальных критериев, использованных при расчетах. Характер кривых на рис. 8 позволяет судить о количестве ясных дней в том или ином месяце. Отклонение кривой в отрицательную сторону свидетельствует о снижении количества ясных дней относительно других месяцев.

Результаты анализа, проведенного для всех цепочек солнечных панелей электростанции, представлены в таблице и на рис. 9. В рассматриваемом тестовом примере восемь солнечных панелей демонстрируют аномальное поведение, всего было обнаружено 18 случаев аномальной работы. Группирование модулей в соответствии с их инверторными и цепными соединениями подчеркивает корреляцию аномалий с конкретным электрическим соединением или физической областью.

Проанализируем кривые напряжения, силы тока и температуры, измеренные для солнечных панелей (таблица), с целью проверки наличия неисправностей в их работе и предварительного анализа возможных причин этих неисправностей. На рис. 10 изображены кривые напряжения в зависимости от времени измерения, снятые датчиками в солнечных панелях Module 2,3_9 и Module 2,3_10. Кривые показывают, что солнечная панель Module 2,3_10 имеет напряжение на 10 В ниже, чем у других солнечных панелей цепочки. Это может свидетельствовать об открытии одного из защитных диодов в солнечной панели, которые предназначены для отключения от цепи частей солнечной панели.

Результаты поиска аномалий в течение шести месяцев (июнь – ноябрь) 2019 г.

Anomaly results for six months (June – November) 2019

Вид анализа <i>Type of analysis</i>	Количество аномалий <i>Number of anomalies</i>	Номер солнечной панели <i>PV module number</i>	Месяц <i>Month</i>	Максимальное отклонение <i>Maximum deviation</i>
Средняя мощность в МРР	4 солнечные панели / 19 аномалий	Module 1,4_1	Весь период. Среднее значение	+23,03 W
		Module 1,4_8	Весь период. Среднее значение	+11,28 W
		Module 1,11_15	Ноябрь	-18,17 W
		Module 2,3_10	Весь период. Среднее значение	-46,38 W
Общий статический анализ по критериям на основе нормализованных значений мощности в МРР: глобальный и локальный	В общем: 8 солнечных панелей / 18 аномалий По группам: в группе 1 5 солнечных панелей / 10 аномалий; в группе 2 0 солнечных панелей / 0 аномалий; в группе 3 3 солнечных панели / 8 аномалий	Module 1,1_2	Август	-10,8 %
		Module 1,1_4	Август	-10 %
		Module 1,4_1	Июнь	+17,8 %
			Июль	+17,1 %
			Август	+19,8 %
			Сентябрь	+19,5 %
			Октябрь	+19,1 %
		Module 1,4_3	Ноябрь	-10,1 %
		Module 1,5_6	Ноябрь	-11,1 %
		Module 1,11_1	Ноябрь	-10,4 %
		Module 1,11_15	Ноябрь	-51,6 %
		Module 2,3_10	Июнь	-28,2 %
			Июль	-25,5 %
			Август	-28,5 %
			Сентябрь	-28,1 %
Октябрь	-26,9 %			
Ноябрь	-21 %			

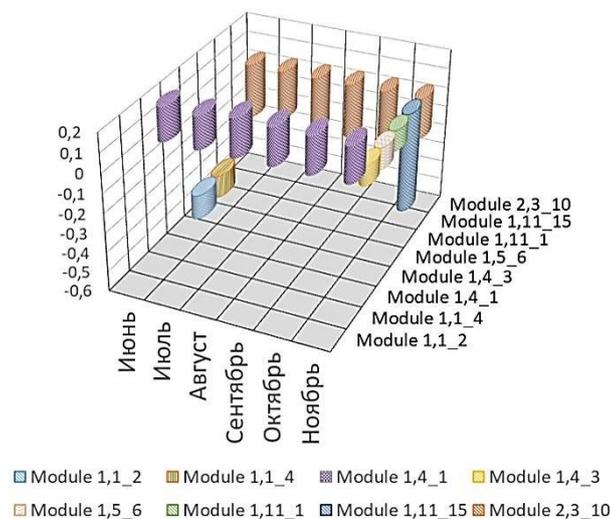


Рис. 9. Группирование аномалий в работе солнечных панелей в течение шести месяцев (июнь – ноябрь) 2019 г.

Fig. 9. Grouping anomalies in the operation of solar panels for six months (June – November) 2019

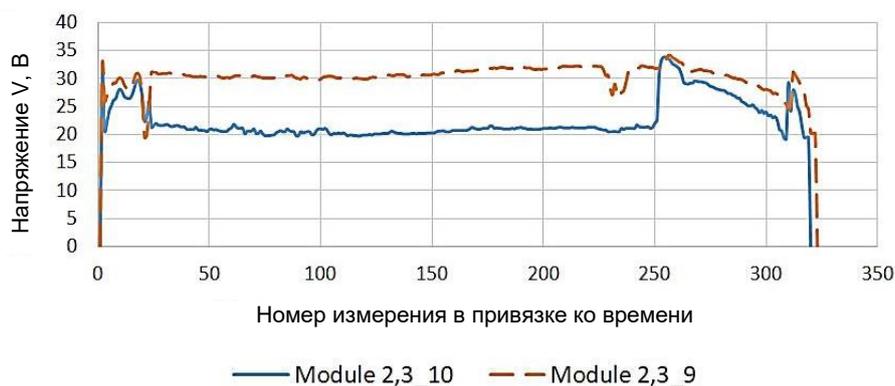


Рис. 10. Изменение напряжения в солнечных панелях Module 2,3_9 и Module 2,3_10 в зависимости от времени

Fig. 10. Change of voltage in solar panels Module 2,3_9 and Module 2,3_10 depending on time

На рис. 11 изображены кривые температуры в зависимости от времени измерения, снятые датчиками в солнечных панелях Module 1,1_2 и Module 1,1_3 в течение ясного солнечного дня. Кривые показывают, что солнечная панель Module 1,1_2 нагревается больше, чем панель Module 1,1_3, на 7–9 °С. Избыточный нагрев может свидетельствовать о неполном или поврежденном контакте при подключении солнечной панели в цепь. Солнечная панель Module 1,1_4 демонстрирует аналогичное поведение температурной кривой в течение дня.

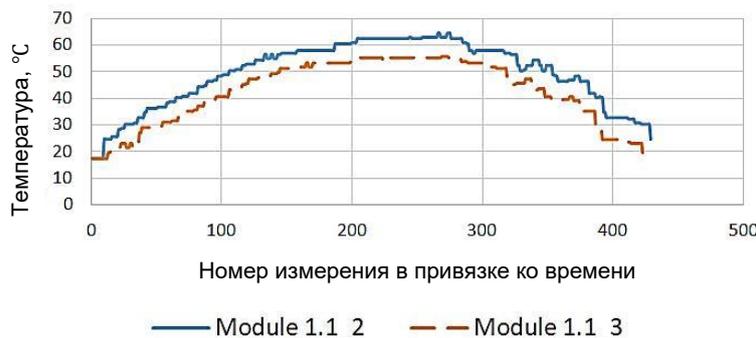


Рис. 11. Изменение температуры в солнечных панелях Module 1,1_2 и Module 1,1_4 в зависимости от времени

Fig. 11. Temperature change in solar panels Module 1,1_2 and Module 1,1_4 depending on time

Анализ кривых температуры для солнечных панелей Module 1,4_3 и Module 1,5_6 в ясные дни в ноябре показал, что они тоже работали с повышением температуры, измеряемой датчиком, на 8–11 °С.

На рис. 12 отражены кривые напряжения, силы тока и температуры в зависимости от времени измерения, снятые датчиками в солнечных панелях Module 1,11_1, Module 1,11_9 и Module 1,11_15 в течение ясного солнечного дня. Кривые напряжения показывают, что данные с датчика солнечной панели Module 1,11_15 отсутствуют при выходе панели на интенсивный рабочий режим. Такое может происходить при срабатывании всех трех защитных диодов и фактическом отключении солнечной панели вместе с датчиком от цепи.

Кривые температуры показывают, что солнечная панель Module 1,11_1 нагревается больше, чем панель Module 1,11_9, на 9–12 °С, т. е. имеет место случай, аналогичный панели Module 1,1_3.

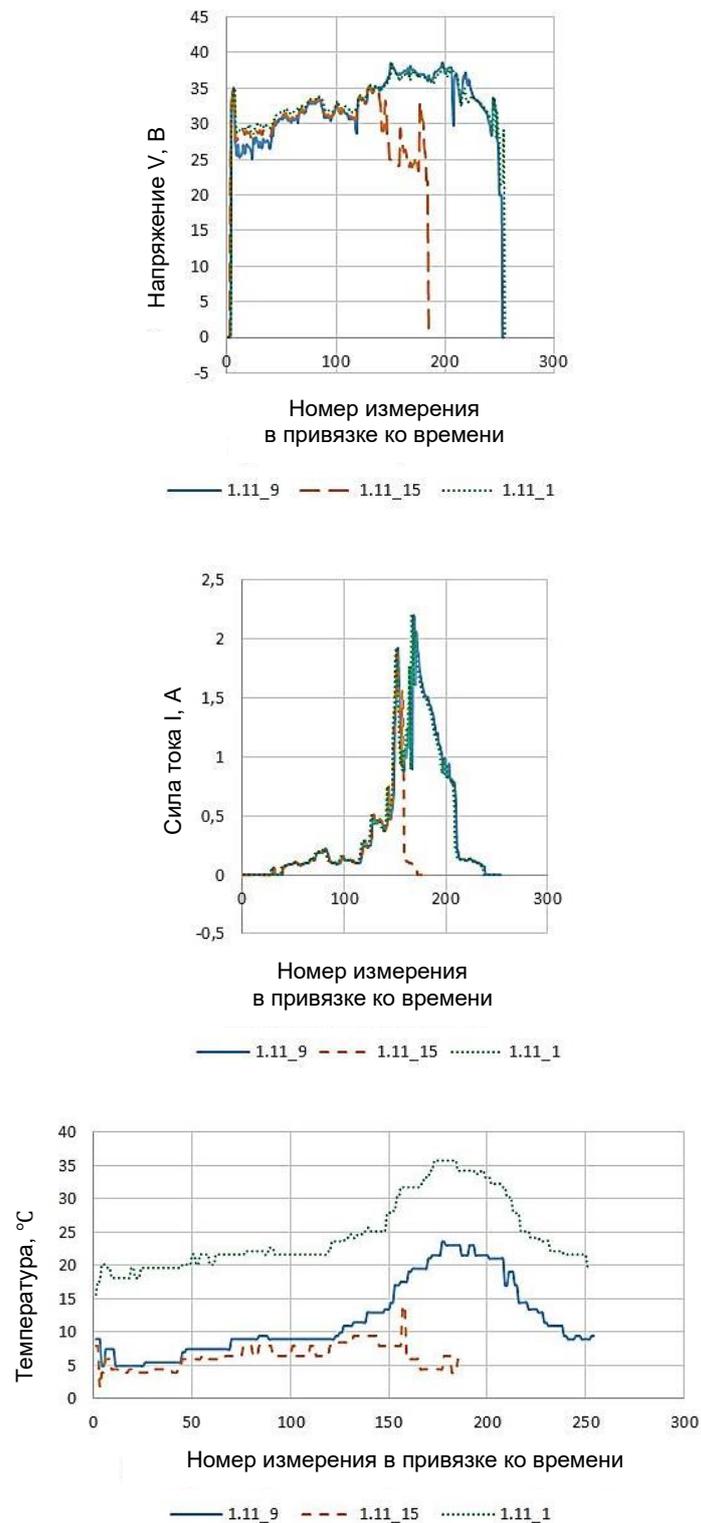


Рис. 12. Изменение напряжения, силы тока и температуры в солнечных панелях Module 1,11_1, Module 1,11_9 и Module 1,11_15 в зависимости от времени
 Fig. 12. Change of voltage, current and temperature in solar panels Module 1,11_1, Module 1,11_9 and Module 1,11_15 depending on time

На основании проведенных исследований было разработано программное средство, алгоритм работы которого представлен на рис. 13. Входными данными для программного средства являются объекты json от API цифрового двойника, а выходными данными – результаты расче-

та нормализованных значений мощностей в МРР для каждой солнечной панели, квартили, медианное и среднее значения P_N , список номеров солнечных панелей, в работе которых обнаружены аномалии, и критерии найденных аномалий.

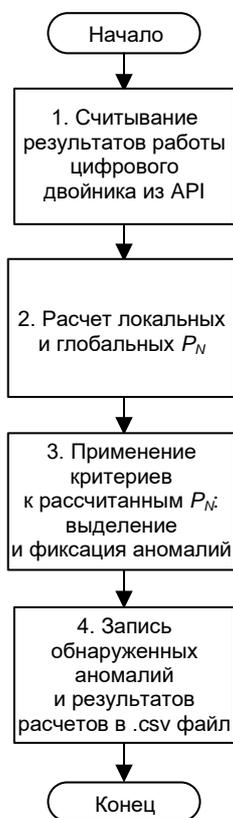


Рис. 13. Блок-схема алгоритма программного обеспечения для поиска аномалий в работе солнечных панелей, разработанного на основе предложенной методики

Fig. 13. Block diagram of the software algorithm for detecting anomalies in the operation of solar panels, developed on the basis of the proposed methodology

Заключение. В исследовании установлено, что, основываясь на нормализованных значениях мощностей в МРР P_N для всех панелей солнечной электростанции, полученных с использованием цифрового двойника, можно обнаруживать аномальную работу отдельных панелей. При использовании в анализе отклонений нормализованных значений мощностей в МРР P_N локальных критериев уровня инвертора об аномальной работе солнечной панели свидетельствует отклонение 9 % и выше. Это подтверждает непосредственный анализ результатов телеметрии, полученных в ходе работы солнечной электростанции.

Таким образом, предлагаемые методика и алгоритм поиска аномалий в работе электростанции смогут оперативно помочь обнаружению солнечных панелей, которые имеют отклонения в режиме работы, что повысит скорость реагирования на появление дефектов в солнечных панелях, а значит и эффективность работы солнечной электростанции.

Вклад авторов. С. В. Валевич разработал алгоритмы и программное обеспечение для выборки и фильтрации данных в соответствии с методикой исследования. К. С. Дик провел сравнительный анализ кривых напряжения силы тока и температуры солнечных панелей, в которых обнаружены аномалии, работал с графическим материалом, редактировал текст рукописи. И. И. Пилецкий сформулировал выводы по результатам исследования. И. Крузе интерпретировал полученные результаты и провел экспертизу результатов исследования. Р. М. Асимов разработал методику поиска аномалий на основе результатов работы цифрового двойника солнечных па-

нелей. В. С. Осипович разработал методику поиска аномалий на основе результатов работы цифрового двойника солнечных панелей. Все авторы принимали участие в формулировании гипотез, целей и задач исследования.

Список использованных источников

1. Dorin, P. Modelling and simulation of photovoltaic cells / P. Dorin, C. Farcas, I. Ciocan // *ACTA Technica Napocensis*. – 2008. – Vol. 49, no. 1. – P. 42–47.
2. Modeling and simulation of energy recovery from a photovoltaic solar cell / Adeniyi O. D. [et al.] // *Nigerian J. of Technological Research*. – 2016. – Vol. 11. – P. 26–31.
3. MATLAB/Simulink based modelling of solar photovoltaic cell / T. Salmi [et al.] // *Intern. J. of Renewable Energy Research*. – 2012. – Vol. 2, no. 2. – P. 213–218.
4. Tina, G. Monitoring and diagnostics of photovoltaic power plants / G. Tina, F. Cosentino, C. Ventura // *Renewable Energy in the Service of Mankind* / ed. A. Sayigh. – Springer, Cham, 2016. – Vol. 2. – P. 505–516.
5. Ibbini, M. Analysis and design of a maximum power point tracker for a stand-alone photo voltaic system using Simscape / M. Ibbini, A. Adawi // *Intern. J. of Advanced Trends in Computer Science and Engineering*. – 2019. – Vol. 8, no. 1. – P. 54–57.
6. Rashid, Md. M. Design and construction of the solar photovoltaic simulation system with the implementation of MPPT and boost converter using Matlab/Simulink / Md. M. Rashid, A. Habib, M. Mahdi Hasan // *Asian J. of Current Research*. – 2018. – Vol. 3, no. 1. – P. 27–36.
7. Flatness-based control for the maximum power point tracking in a photovoltaic system / G.-A. Leopoldo [et al.] // *Energies*. – 2019. – Vol. 12. – P. 1843–1862. <https://doi.org/10.3390/en12101843>
8. Modeling of PV module with consideration of environmental factors / N. Kishor [et al.] // *2010 IEEE PES Innovative Smart Grid Technologies Conf. Europe (ISGT Europe)*, Gothenburg, Sweden, 11–13 Oct. 2010. – Gothenburg, 2010. – P. 121–126.
9. Simulation of PV array output power for modified PV cell model / N. Kishor [et al.] // *2010 IEEE Intern. Conf. on Power and Energy*, Kuala Lumpur, Malaysia, 29 Nov. – 01 Dec. 2010. – Kuala Lumpur, 2010. – P. 533–538.
10. Steady-state model and power flow analysis of grid-connected photovoltaic power system / Y. Wang [et al.] // *2008 IEEE Intern. Conf. on Industrial Technology*, Chengdu, China, 21–24 Apr. 2008. – Chengdu, 2008. – P. 1–6.
11. Virtual laboratory for testing of solar power plants in big data analysis / R. M. Asimov [et al.] // *Collection of Materials of the V Intern. Scientific and Practical Conf. "Big Data and Advanced Analytics"*, Minsk, 13–14 Mar. 2019. – Minsk, 2019. – P. 61–65.
12. Asipovich, V. S. Digital twin in the analysis of a big data / V. S. Asipovich, R. M. Asimov, S. V. Chernoshey // *Collection of Materials of the IV Intern. Scientific and Practical Conf. "Big Data and Advanced Analytics"*, Minsk, 3–4 May 2018. – Minsk, 2018. – P. 69–78.
13. Digital twin for PV module fault detection / S. Valevich [et al.] // *J. of Engineering Science*. – 2020. – Vol. XXVII, no. 4. – P. 80–87.
14. Digital twin for PV plant's power generation analysis / R. M. Asimov [et al.] // *Collection of Materials of the VI Intern. Scientific and Practical Conf. "Big Data and Advanced Analytics"*, Minsk, 20–21 May 2020. – Minsk, 2020. – P. 78–88.
15. Информационное обеспечение мониторинга технического состояния солнечных электростанций / С. В. Валеvич [и др.] // *Информационные технологии*. – 2020. – Т. 26, № 10. – С. 594–601.

References

1. Dorin P., Farcas C., Ciocan I. Modelling and simulation of photovoltaic cells. *ACTA Technica Napocensis*, 2008, vol. 49, no. 1, pp. 42–47.
2. Adeniyi O. D., Ali D. A., Olutoye M. A., Adeniyi M. I., Azeez O. S., ..., Eniafe B. O. Modeling and simulation of energy recovery from a photovoltaic solar cell. *Nigerian Journal of Technological Research*, 2016, vol. 11, pp. 26–31.
3. Salmi T., Bouzguenda M., Gastli A., Masmoudi A. MATLAB/Simulink based modelling of solar photovoltaic cell. *International Journal of Renewable Energy Research*, 2012, vol. 2, no. 2, pp. 213–218.
4. Tina G., Cosentino F., Ventura C. Monitoring and diagnostics of photovoltaic power plants. *Renewable Energy in the Service of Mankind*. In A. Sayigh (ed.), Springer, Cham, 2016, vol. 2, pp. 505–516.

5. Ibbini M., Adawi A. Analysis and design of a maximum power point tracker for a stand-alone photo voltaic system using simscape. *International Journal of Advanced Trends in Computer Science and Engineering*, 2019, vol. 8, no. 1, pp. 54–57.
6. Rashid Md. M., Habib A., Mahdi Hasan M. Design and construction of the solar photovoltaic simulation system with the implementation of MPPT and boost converter using Matlab/Simulink. *Asian Journal of Current Research*, 2018, vol. 3, no. 1, pp. 27–36.
7. Leopoldo G.-A., Belem S., Otniel P.-R., Juan C. Á.-V., Pánfilo R. M.-R., Rigoberto M.-M. Flatness-based control for the maximum power point tracking in a photovoltaic system. *Energies*, 2019, vol. 12, pp. 1843–1862. <https://doi.org/10.3390/en12101843>
8. Kishor N., Villalva M., Mohanty S. R., Filho E. Modeling of PV module with consideration of environmental factors. *2010 IEEE PES Innovative Smart Grid Technologies Conference Europe (ISGT Europe), Gothenburg, Sweden, 11–13 October 2010*. Gothenburg, 2010, pp. 121–126.
9. Kishor N., Mohanty S. R., Villalva M., Filho E. Simulation of PV array output power for modified PV cell model. *2010 IEEE International Conference on Power and Energy, Kuala Lumpur, Malaysia 29 November – 01 December 2010*. Kuala Lumpur, 2010, pp. 533–538.
10. Wang Y., Chun-Sheng W., Hua L., Hong-Hua X. Steady-state model and power flow analysis of grid-connected photovoltaic power system. *2008 IEEE International Conference on Industrial Technology, Chengdu, China, 21–24 April 2008*. Chengdu, 2008, pp. 1–6.
11. Asimov R. M., Valevich S. V., Kruse I., Asipovich V. S. Virtual laboratory for testing of solar power plants in big data analysis. *Collection of Materials of the V International Scientific and Practical Conference "Big Data and Advanced Analytics", Minsk, 13–14 March 2019*. Minsk, 2019, pp. 61–65.
12. Asipovich V. S., Asimov R. M., Chernoshey S. V. Digital twin in the analysis of a big data. *Collection of Materials of the IV International Scientific and Practical Conference "Big Data and Advanced Analytics", Minsk, 3–4 May 2018*. Minsk, 2018, pp. 69–78.
13. Valevich S., Asimov R., Kruse I., Asipovich V. Digital twin for PV module fault detection. *Journal of Engineering Science*, 2020, vol. XXVII, no. 4, pp. 80–87.
14. Asimov R. M., Valevich S. V., Kruse I., Asipovich V. S. Digital twin for PV plant's power generation analysis. *Collection of Materials of the VI International Scientific and Practical Conference "Big Data and Advanced Analytics", Minsk, 20–21 May 2020*. Minsk, 2020, pp. 78–88.
15. Valevich S. V., Asipovich V. S., Kruse I., Asimov R. M. *Information support for monitoring the technical condition of solar power plants*. *Informacionnie tehnologii [Information Technology]*, 2020, vol. 26, no. 10, pp. 594–601 (In Russ.).

Информация об авторах

Валевич Сергей Владимирович, магистр технических наук, аспирант, Белорусский государственный университет информатики и радиоэлектроники.
<https://orcid.org/0000-0001-9324-8790>

Дик Константин Сергеевич, аспирант, Белорусский государственный университет информатики и радиоэлектроники.
<https://orcid.org/0009-0006-8750-0738>

Пилецкий Иван Иванович, кандидат физико-математических наук, доцент, доцент кафедры информатики, Белорусский государственный университет информатики и радиоэлектроники.
<https://orcid.org/0009-0002-2252-0835>

Крузе Ингмар, директор, Sunsniffer LTD.
<https://orcid.org/0000-0001-7074-6172>

Асимов Рустам Мустафьевич, кандидат технических наук, директор, ООО Сенсотроника.
<https://orcid.org/0000-0002-4033-4238>

Осипович Виталий Семенович, кандидат технических наук, доцент, технический директор, Novotech Lab LTD.
<https://orcid.org/0000-0001-9658-2866>

Information about the authors

Sergey V. Valevich, M. Sc. (Eng.), Postgraduate Student, Belarusian State University of Informatics and Radioelectronics.
<https://orcid.org/0000-0001-9324-8790>

Konstantine S. Dzick, Postgraduate Student, Belarusian State University of Informatics and Radioelectronics.
<https://orcid.org/0009-0006-8750-0738>

Ivan I. Pilecki, Ph. D. (Phys.-Math.), Associate Professor, Associate Professor Department of Informatics, Belarusian State University of Informatics and Radioelectronics.
<https://orcid.org/0009-0002-2252-0835>

Kruse Ingmar, CEO, Sunsniffer LTD.
<https://orcid.org/0000-0001-7074-6172>

Roustam M. Asimov, Ph. D. (Eng.), CEO, Sensotronica LTD.
<https://orcid.org/0000-0002-4033-4238>

Vitali S. Asipovich, Ph. D. (Eng.), Associate Professor, CTO, Novotech Lab LTD.
<https://orcid.org/0000-0001-9658-2866>

МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ MATHEMATICAL MODELING



УДК 519.6
<https://doi.org/10.37661/1816-0301-2023-20-2-111-120>

Оригинальная статья
Original Paper

Решение смешанной краевой задачи для уравнения Пуассона в нерегулярных двумерных областях

М. М. Чуйко^{1✉}, О. М. Королёва²

¹Институт математики НАН Беларуси,
ул. Сурганова, 11, Минск, 220072, Беларусь
✉E-mail: mikhail.chuiko@gmail.com

²Белорусский национальный технический университет,
пр. Независимости, 65, Минск, 220013, Беларусь

Аннотация

Цели. Построение конечно-разностного вычислительного алгоритма решения смешанной краевой задачи для уравнения Пуассона, заданной в нерегулярных двумерных областях.

Методы. Для решения задачи используются обобщенные криволинейные координаты. Физическая область отображается в расчетную (единичный квадрат) в пространстве обобщенных координат. Исходная задача записывается в обобщенных криволинейных координатах и аппроксимируется на равномерной сетке в расчетной области. Полученные результаты отображаются на неравномерную разностную сетку, сгенерированную в физической области.

Результаты. Построены аппроксимации второго порядка смешанных краевых условий Неймана – Дирихле для уравнения Пуассона в пространстве обобщенных криволинейных координат. Для повышения порядка аппроксимаций условия Неймана используется аппроксимация уравнения Пуассона на границе области.

Заключение. Для решения смешанной краевой задачи для уравнения Пуассона в нерегулярных двумерных областях построен вычислительный алгоритм второго порядка точности с использованием обобщенных криволинейных координат. Приведены результаты численных экспериментов, подтверждающие второй порядок точности вычислительного алгоритма.

Ключевые слова: эллиптический оператор, смешанные производные, обобщенные криволинейные координаты, краевая задача Неймана – Дирихле, конечно-разностные методы, разностные схемы

Для цитирования. Чуйко, М. М. Решение смешанной краевой задачи для уравнения Пуассона в нерегулярных двумерных областях / М. М. Чуйко, О. М. Королёва // Информатика. – 2023. – Т. 20, № 2. – С. 111–120. <https://doi.org/10.37661/1816-0301-2023-20-2-111-120>

Конфликт интересов. Авторы заявляют об отсутствии конфликта интересов.

Поступила в редакцию | Received 02.03.2023
Подписана в печать | Accepted 06.04.2023
Опубликована | Published 29.06.2023

Solution of the mixed boundary problem for the Poisson equation on two-dimensional irregular domains

Mikhail M. Chuiko^{1✉}, Olga M. Korolyova²

¹*Institute of Mathematics of the National Academy of Sciences of Belarus,
st. Surganova, 11, Minsk, 220072, Belarus*

✉E-mail: mikhail.chuiko@gmail.com

²*Belarusian National Technical University,
av. Nezavisimosty, 65, Minsk, 220013, Belarus*

Abstract

Objectives. A finite-difference computational algorithm is proposed for solving a mixed boundary-value problem for the Poisson equation given in two-dimensional irregular domains.

Methods. To solve the problem, generalized curvilinear coordinates are used. The physical domain is mapped to the computational domain (unit square) in the space of generalized coordinates. The original problem is written in curvilinear coordinates and approximated on a uniform grid in the computational domain. The obtained results are mapped on non-uniform boundary-fitted difference grid in the physical domain.

Results. The second order approximations of mixed Neumann-Dirichlet boundary conditions for the Poisson equation in the space of generalized curvilinear coordinate are constructed. To increase the order of Neumann condition approximations, an approximation of the Poisson equation on the boundary of the domain is used.

Conclusions. To solve a mixed boundary value problem for the Poisson equation in two-dimensional irregular domains, the computational algorithm of second-order accuracy is constructed. The generalized curvilinear coordinates are used. The results of numerical experiments, which confirm the second order accuracy of the computational algorithm, are presented.

Keywords: elliptic operator, mixed derivatives, generalized curvilinear coordinates, Neumann – Dirichlet boundary problem, finite-difference methods, difference schemes

For citation. Chuiko M. M., Korolyova O. M. *Solution of the mixed boundary problem for the Poisson equation on two-dimensional irregular domains*. Informatika [Informatics], 2023, vol. 20, no. 2, pp. 111–120 (In Russ.). <https://doi.org/10.37661/1816-0301-2023-20-2-111-120>

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

Введение. Одним из подходов к приближенному решению уравнений в частных производных в областях сложной формы является введение обобщенных криволинейных координат, отображающих исходную область в физическом пространстве в параметрический квадрат в расчетном пространстве криволинейных координат [1]. Исходные уравнения преобразуются в новые независимые переменные и решаются в расчетном пространстве на прямоугольной разностной сетке, что позволяет использовать классические методы теории разностных схем. Построение невырожденного преобразования исходной области в параметрический квадрат в дискретном случае заключается в генерации в исходной области подходящей регулярной (четырёхугольной) разностной сетки. Одна из проблем такого подхода – усложнение исходных уравнений в связи с появлением смешанных производных. Для решения этой проблемы необходимо разработать и использовать монотонные консервативные разностные схемы.

Консервативные разностные схемы для уравнений со смешанными производными были предложены А. А. Самарским и В. Б. Андреевым [2, 3]. В более поздних работах А. А. Самарского, П. П. Матуса, Г. И. Шишкина [3–5] для эллиптических и параболических уравнений со смешанными производными были разработаны монотонные разностные методы на основе двух консервативных схем [2]. Построенные вычислительные алгоритмы удовлетворяют одновременно свойствам монотонности и консервативности, только если коэффициенты при смешанных производных либо положительные, либо отрицательные.

В работе [6] для эллиптических и параболических уравнений со знакопеременными коэффициентами при смешанных производных построены разностные схемы второго порядка аппроксимации, удовлетворяющие одновременно свойствам монотонности и консервативности. Свой-

ство монотонности гарантирует хорошую обусловленность системы линейных алгебраических уравнений, так как элементы матрицы удовлетворяют свойству диагонального преобладания, а свойство консервативности гарантирует выполнение законов сохранения при преобразовании координат.

При задании краевых условий Неймана важно сохранить второй порядок аппроксимации по пространственным переменным. Для повышения порядка аппроксимации условий Неймана используются два подхода. Первый подход заключается в построении разностных сеток с фиктивными узлами и аппроксимации краевых условий с выходом за границу области [7, 8], второй подход – в использовании аппроксимации исходного уравнения на границе области.

В настоящей работе для решения в сложных двумерных областях краевой задачи Неймана – Дирихле для уравнения Пуассона используется переход к обобщенной криволинейной системе координат. Повышение порядка аппроксимации краевых условий Неймана до второго производится с привлечением аппроксимации уравнения Пуассона на границе области.

Постановка задачи. Рассмотрим уравнение Пуассона в нерегулярной двумерной области Ω_{xy} с граничными условиями первого и второго рода:

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = f(x, y), \quad (x, y) \in \Omega_{xy}, \quad (1)$$

$$u(x, y)|_{\Gamma_1} = u_0(x, y), \quad \Gamma_1 \in \partial\Omega_{xy}, \quad (2)$$

$$\frac{\partial u}{\partial \mathbf{n}} \Big|_{\Gamma_2} = \mu(x, y), \quad \Gamma_2 \in \partial\Omega_{xy} - \Gamma_1, \quad (3)$$

где $\partial\Omega_{xy}$ – граница области Ω_{xy} , \mathbf{n} – внешняя нормаль к границе. Краевые условия (2), (3) могут соответствовать, например, заданному распределению температуры на участке Γ_1 границы $\partial\Omega_{xy}$ и заданному тепловому потоку на участке Γ_2 .

Численное решение задач в областях сложной формы удобно искать в криволинейной системе координат, координатные линии которых совпадают с границами области [1]. Введем в рассмотрение некоторое расчетное пространство $\Omega_{\xi\eta}$, в котором определена криволинейная система координат (ξ, η) . Предположим, что существует невырожденное взаимно однозначное преобразование $\xi = \xi(x, y)$, $\eta = \eta(x, y)$, отображающее физическую область произвольной формы Ω_{xy} в прямоугольник $\Omega_{\xi\eta} = \{(\xi, \eta) : 0 < \xi, \eta < 1\}$ в плоскости обобщенных криволинейных координат (ξ, η) .

Краевую задачу (1)–(3) запишем в виде

$$\frac{\partial}{\partial \xi} \left(B_{11} \frac{\partial u}{\partial \xi} + B_{12} \frac{\partial u}{\partial \eta} \right) + \frac{\partial}{\partial \eta} \left(B_{21} \frac{\partial u}{\partial \xi} + B_{22} \frac{\partial u}{\partial \eta} \right) = |J^{-1}| \tilde{f}(\xi, \eta), \quad (\xi, \eta) \in \Omega_{\xi\eta}, \quad (4)$$

$$u(\xi, \eta) = \tilde{u}_0(\xi, \eta), \quad (\xi, \eta) \in \Gamma_{1,\xi\eta}, \quad (5)$$

$$B_{11} \frac{\partial u}{\partial \xi} + B_{12} \frac{\partial u}{\partial \eta} = g^{1/2} \tilde{\mu}(\xi, \eta), \quad (\xi, \eta) \in \Gamma_{2,\xi\eta}, \quad \xi = 0, 1, \quad (6)$$

$$B_{21} \frac{\partial u}{\partial \xi} + B_{22} \frac{\partial u}{\partial \eta} = g^{1/2} \tilde{\mu}(\xi, \eta), \quad (\xi, \eta) \in \Gamma_{2,\xi\eta}, \quad \eta = 0, 1, \quad (7)$$

где

$$B_{11} = \frac{g_{22}}{|J^{-1}|}, \quad B_{12} = B_{21} = -\frac{g_{12}}{|J^{-1}|}, \quad B_{22} = \frac{g_{11}}{|J^{-1}|},$$

$$g_{11} = \left(\frac{\partial x}{\partial \xi}\right)^2 + \left(\frac{\partial y}{\partial \xi}\right)^2, \quad g_{12} = \frac{\partial x}{\partial \xi} \frac{\partial x}{\partial \eta} + \frac{\partial y}{\partial \xi} \frac{\partial y}{\partial \eta}, \quad g_{22} = \left(\frac{\partial x}{\partial \eta}\right)^2 + \left(\frac{\partial y}{\partial \eta}\right)^2,$$

$$|J^{-1}| = \frac{\partial x}{\partial \xi} \frac{\partial y}{\partial \eta} - \frac{\partial y}{\partial \xi} \frac{\partial x}{\partial \eta}.$$

Аппроксимация краевых условий. В расчетной прямоугольной области $\Omega_{\xi\eta}$ введем равномерную по обоим направлениям разностную сетку $\bar{\omega}_h = \omega_h \cup \gamma_h$:

$$\bar{\omega}_h = \left\{ (\xi_i, \eta_j), \quad \xi_i = ih_1, \quad i = \overline{0, N}, \quad \xi_N = 1, \quad \eta_j = jh_2, \quad j = \overline{0, M}, \quad \eta_M = 1 \right\},$$

где ω_h – множество внутренних узлов, γ_h – множество граничных узлов. На разностной сетке ω_h аппроксимируем выражение (4) следующей разностной схемой [6]:

$$\sum_{\alpha, \beta=1}^2 \Lambda_{\alpha\beta} u = |J^{-1}| \tilde{f}, \quad (8)$$

$$\Lambda_{11} u = (\beta_{11} u_{\xi})_{\xi}, \quad \Lambda_{22} u = (\beta_{22} u_{\eta})_{\eta}, \quad (9)$$

$$\Lambda_{12} u = \frac{1}{2} \left((\beta_{12}^- u_{\bar{\eta}})_{\xi} + (\beta_{12}^- u_{\eta})_{\bar{\xi}} + (\beta_{12}^+ u_{\eta})_{\xi} + (\beta_{12}^+ u_{\bar{\eta}})_{\bar{\xi}} \right), \quad (10)$$

$$\Lambda_{21} u = \frac{1}{2} \left((\beta_{21}^- u_{\bar{\xi}})_{\eta} + (\beta_{21}^- u_{\xi})_{\bar{\eta}} + (\beta_{21}^+ u_{\xi})_{\eta} + (\beta_{21}^+ u_{\bar{\xi}})_{\bar{\eta}} \right), \quad (11)$$

где $\beta^{\pm} = 0,5(\beta \pm |\beta|)$.

В граничных узлах $(\xi_i, \eta_j) \in \gamma_1$ с заданными условиями первого рода краевые условия аппроксимируются точно: $u_{ij} = \tilde{u}_0(\xi_i, \eta_j)$.

Рассмотрим аппроксимацию краевых условий второго рода. Так, для $\eta=0$ из равенства (7) получим

$$B_{21} \frac{\partial u}{\partial \xi} + B_{22} \frac{\partial u}{\partial \eta} = -g_{11}^{1/2} \tilde{\mu}(\xi, 0), \quad (\xi, \eta) \in \Gamma_{2, \xi\eta}, \quad \eta=0. \quad (12)$$

Аппроксимируем (12) следующим образом:

$$\left(\beta_{21} u_{\xi} + \beta_{22} u_{\eta} \right)_{i,0} = \left(B_{21} \frac{\partial u}{\partial \xi} + B_{22} \frac{\partial u}{\partial \eta} + \frac{h_2}{2} \left(B_{22} \frac{\partial^2 u}{\partial \eta^2} \right) \right)_{i,0} + O(h_1^2 + h_2^2). \quad (13)$$

Рассмотрим способ повышения порядка аппроксимации граничного условия (12), используя для этого уравнение (4) на границе. Запишем уравнение (4) в виде

$$\frac{\partial}{\partial \xi} \left(B_{11} \frac{\partial u}{\partial \xi} + B_{12} \frac{\partial u}{\partial \eta} \right) + \frac{\partial B_{21}}{\partial \eta} \frac{\partial u}{\partial \xi} + B_{21} \frac{\partial^2 u}{\partial \xi \partial \eta} + \frac{\partial B_{22}}{\partial \eta} \frac{\partial u}{\partial \eta} + B_{22} \frac{\partial^2 u}{\partial \eta^2} = |J^{-1}| \tilde{f}(\xi, \eta), \quad (\xi, \eta) \in \Omega_{\xi\eta}. \quad (14)$$

Выразим значение $B_{22} \frac{\partial^2 u}{\partial \eta^2}$ из (14) и подставим его в выражение (13). В результате получим равенство

$$\begin{aligned} & \left(\beta_{21} u_{\xi}^0 + \beta_{22} u_{\eta} \right)_{i,0} + \frac{h_2}{2} \left(\frac{\partial}{\partial \xi} \left(B_{11} \frac{\partial u}{\partial \xi} + B_{12} \frac{\partial u}{\partial \eta} \right) + \frac{\partial B_{21}}{\partial \eta} \frac{\partial u}{\partial \xi} + B_{21} \frac{\partial^2 u}{\partial \xi \partial \eta} + \frac{\partial B_{22}}{\partial \eta} \frac{\partial u}{\partial \eta} + \tilde{f} \right)_{i,0} = \\ & = \left(B_{21} \frac{\partial u}{\partial \xi} + B_{22} \frac{\partial u}{\partial \eta} \right)_{i,0} + O(h_1^2 + h_2^2). \end{aligned} \quad (15)$$

Аппроксимируем дифференциальные слагаемые в левой части равенства (15):

$$\begin{aligned} \frac{\partial}{\partial \xi} \left(B_{11} \frac{\partial u}{\partial \xi} \right)_{i,0} &= \left((\beta_{11,i-1/2,1} u_{\xi}^-)_{\xi} \right)_{i,0} + O(h_1^2), \\ \frac{\partial}{\partial \xi} \left(B_{12} \frac{\partial u}{\partial \eta} \right)_{i,0} &= \left((\beta_{12}^+ u_{\eta})_{\xi} + (\beta_{12}^- u_{\eta})_{\bar{\xi}} \right)_{i,0} + O(h_1 + h_2), \\ \left(\frac{\partial B_{21}}{\partial \eta} \frac{\partial u}{\partial \xi} \right)_{i,0} &= \left(\beta_{21,\eta} u_{\xi}^0 \right)_{i,0} + O(h_1^2 + h_2), \\ \left(B_{21} \frac{\partial^2 u}{\partial \xi \partial \eta} \right)_{i,0} &= \left(\beta_{21,i,1/2}^+ u_{\xi\eta} + \beta_{21,i,1/2}^- u_{\bar{\xi}\eta} \right)_{i,0} + O(h_1 + h_2), \\ \left(\frac{\partial B_{22}}{\partial \eta} \frac{\partial u}{\partial \eta} \right)_{i,0} &= \left(\beta_{22,\eta} u_{\eta} \right)_{i,0} + O(h_2) \end{aligned}$$

и запишем (14) в виде

$$\begin{aligned} & \left(\beta_{21} u_{\xi}^0 + \beta_{22} u_{\eta} \right)_{i,0} + \\ & + \frac{h_2}{2} \left((\beta_{11,i-1/2,1} u_{\xi}^-)_{\xi} + (\beta_{12}^+ u_{\eta})_{\xi} + (\beta_{12}^- u_{\eta})_{\bar{\xi}} + \beta_{21,\eta} u_{\xi}^0 + \beta_{21,i,1/2}^+ u_{\xi\eta} + \beta_{21,i,1/2}^- u_{\bar{\xi}\eta} + \beta_{22,\eta} u_{\eta} + \tilde{f} \right)_{i,0} = \\ & = \left(B_{21} \frac{\partial u}{\partial \xi} + B_{22} \frac{\partial u}{\partial \eta} \right)_{i,0} + O(h_1^2 + h_2^2). \end{aligned} \quad (16)$$

Сгруппируем слагаемые в уравнении (16):

$$\begin{aligned} & \left(\left(\beta_{22} + \frac{h_2}{2} \beta_{22,\eta} \right) u_{\eta} \right)_{i,0} = \beta_{22,i,1/2} u_{\eta,i,0}, \\ & \left(\beta_{21} + \frac{h_2}{2} \beta_{21,\eta} \right) u_{\xi,i,0}^0 + \frac{h_2}{2} \left(\beta_{21}^+ u_{\xi\eta} + \beta_{21}^- u_{\bar{\xi}\eta} \right)_{i,0} = \\ & = \left(\beta_{21}^+ + \beta_{21}^- \right)_{i,1/2} \frac{(u_{\bar{\xi}} + u_{\xi})_{i,0}}{2} + \beta_{21,i,1/2}^+ \frac{u_{\xi,i,1} - u_{\xi,i,0}}{2} + \beta_{21,i,1/2}^- \frac{u_{\bar{\xi},i,1} - u_{\bar{\xi},i,0}}{2} = \\ & = \frac{1}{2} \left(\beta_{21,i,1/2}^+ (u_{\xi,i,1} + u_{\bar{\xi},i,0}) + \beta_{21,i,1/2}^- (u_{\bar{\xi},i,1} + u_{\xi,i,0}) \right), \end{aligned}$$

где

$$\beta_{22,i,1/2} = 0,5(\beta_{22,i,0} + \beta_{22,i,1}), \quad \beta_{21,i,1/2} = 0,5(\beta_{21,i,0} + \beta_{21,i,1}).$$

В результате получим аппроксимацию второго порядка для граничного условия (12):

$$\begin{aligned} & \frac{1}{2}(\beta_{21,i,1/2}^+(u_{\xi,i,1} + u_{\bar{\xi},i,0}) + \beta_{21,i,1/2}^-(u_{\bar{\xi},i,1} + u_{\xi,i,0})) + \beta_{22,i,1/2}u_{\eta,i,0} + \\ & + \frac{h_2}{2}((\beta_{11,i-1/2,0}u_{\bar{\xi}})_{\xi} + (\beta_{12}^+u_{\eta})_{\xi} + (\beta_{12}^-u_{\eta})_{\bar{\xi}})_{i,0} = \left(-\mu g_{11}^{1/2} + \frac{h_2}{2}\tilde{f}\right)_{i,0}. \end{aligned}$$

Аналогичным образом строятся аппроксимации краевых условий Неймана для других участков границы криволинейного четырехугольника.

Так, краевое условие

$$B_{21}\frac{\partial u}{\partial \xi} + B_{22}\frac{\partial u}{\partial \eta} = g_{11}^{1/2}\tilde{\mu}(\xi,1), \quad (\xi,\eta) \in \Gamma_{2,\xi\eta}, \quad \eta=1,$$

аппроксимируем с помощью выражения

$$\begin{aligned} & \frac{1}{2}(\beta_{21,i,M-1/2}^+(u_{\bar{\xi},i,M-1} + u_{\xi,i,M}) + \beta_{21,i,M-1/2}^-(u_{\xi,i,M-1} + u_{\bar{\xi},i,M})) + \beta_{22,i,N-1/2}u_{\bar{\eta},i,M} - \\ & - \frac{h_2}{2}((\beta_{11,i-1/2,M}u_{\bar{\xi}})_{\xi} + (\beta_{12}^+u_{\eta})_{\xi} + (\beta_{12}^-u_{\eta})_{\bar{\xi}})_{i,M} = \left(\mu g_{11}^{1/2} - \frac{h_2}{2}\tilde{f}\right)_{i,M}, \end{aligned}$$

а краевые условия

$$B_{11}\frac{\partial u}{\partial \xi} + B_{12}\frac{\partial u}{\partial \eta} = -g_{22}^{1/2}\tilde{\mu}(0,\eta), \quad (\xi,\eta) \in \Gamma_{2,\xi\eta}, \quad \xi=0,$$

и

$$B_{11}\frac{\partial u}{\partial \xi} + B_{12}\frac{\partial u}{\partial \eta} = g_{22}^{1/2}\tilde{\mu}(1,\eta), \quad (\xi,\eta) \in \Gamma_{2,\xi\eta}, \quad \xi=1, \quad -$$

с помощью разностных соотношений

$$\begin{aligned} & \frac{1}{2}(\beta_{12,1/2,j}^+(u_{\eta,1,j} + u_{\bar{\eta},0,j}) + \beta_{12,1/2,j}^-(u_{\eta,0,j} + u_{\bar{\eta},1,j})) + \beta_{11,1/2,j}u_{\xi,0,j} + \\ & + \frac{h_1}{2}((\beta_{22,0,j-1/2}u_{\bar{\eta}})_{\eta} + (\beta_{21}^+u_{\xi})_{\eta} + (\beta_{21}^-u_{\xi})_{\bar{\eta}})_{0,j} = \left(-\mu g_{22}^{1/2} + \frac{h_1}{2}\tilde{f}\right)_{0,j} \end{aligned}$$

и

$$\begin{aligned} & \frac{1}{2}(\beta_{12,N-1/2,j}^+(u_{\eta,N,j} + u_{\bar{\eta},N-1,j}) + \beta_{12,N-1/2,j}^-(u_{\bar{\eta},N,j} + u_{\eta,N-1,j})) + \beta_{11,N-1/2,j}u_{\xi,N,j} - \\ & - \frac{h_1}{2}((\beta_{22,N,j-1/2}u_{\bar{\eta}})_{\eta} + (\beta_{21}^-u_{\xi})_{\eta} + (\beta_{21}^+u_{\xi})_{\bar{\eta}})_{N,j} = \left(\mu g_{22}^{1/2} - \frac{h_1}{2}\tilde{f}\right)_{N,j} \end{aligned}$$

соответственно.

Численный эксперимент. Проверка алгоритма проводилась на точном решении $u(x,y) = \sin\left(\frac{2\pi x}{L_1}\right)\cos\left(\frac{2\pi y}{L_2}\right)$ в областях $\Omega_{1,xy}, \Omega_{2,xy}$, где L_1, L_2 – размеры области по x и y соот-

ответственно. На рис. 1 изображены разностные сетки 32×32 , построенные в областях $\Omega_{1,xy}$, $\Omega_{2,xy}$ и в расчетной области $\Omega_{\xi\eta}$. Матрица системы линейных алгебраических уравнений имеет специальную структуру и является девятидиагональной. Для решения системы уравнений использован модифицированный сильно неявный метод MSIM [9]. В качестве начального приближения решение во всех узлах задавалось нулевым. Критерием окончания итерационного процесса являлось выполнение условия

$$\left| u_{ij}^{m+1} - u_{ij}^m \right| < \varepsilon_1 \left| u_{ij}^m \right| + \varepsilon_2, \quad i = \overline{0, N}, \quad j = \overline{0, M}, \quad \varepsilon_1 = \varepsilon_2 = 10^{-6}.$$

На рис. 2 показаны примеры численного решения смешанной краевой задачи для уравнения Пуассона.

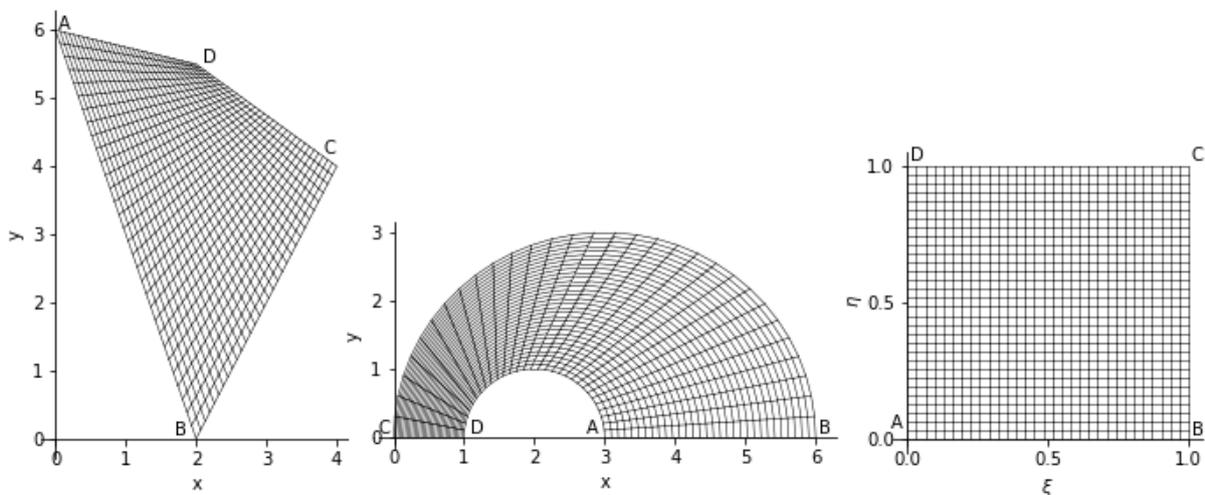


Рис. 1. Физические области $\Omega_{1,xy}$, $\Omega_{2,xy}$ и расчетная область $\Omega_{\xi\eta}$

Fig. 1. Physical domains $\Omega_{1,xy}$, $\Omega_{2,xy}$ and computational domain $\Omega_{\xi\eta}$

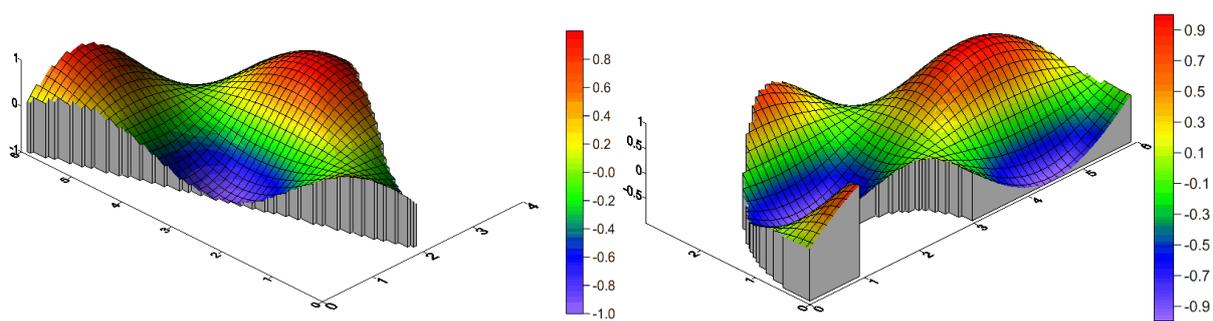


Рис. 2. Численное решение на сетке 32×32

Fig. 2. Numerical solution on grid 32×32

Абсолютные погрешности $z^N = \max |u_{ij} - u(x_{ij}, y_{ij})|$ разностного решения на последовательности разностных сеток $N \times N$ для различных способов задания смешанных краевых условий на границах областей приведены в таблице. В фигурных скобках даны количества итераций метода MSIM для достижения заданной точности. Результаты численных экспериментов подтверждают второй порядок точности алгоритма. Уменьшение шагов сетки приводит к увеличению числа итераций. Анализ эффективности итерационной реализации разностных схем (в том

числе и используемой разностной схемы (8)–(11)), аппроксимирующих уравнение эллиптического типа со смешанными производными второго порядка в методе фиктивных областей, проведен в работе [10].

Погрешности для задач Дирихле (I) и Неймана (II) с различными способами задания граничных условий
Errors for the Dirichlet (I) and Neumann (II) problems with different setting boundary conditions

Краевое условие <i>Boundary condition</i>	$\Omega_{1,xy}$			$\Omega_{2,xy}$		
	16×16	32×32	64×64	16×16	32×32	64×64
I на ABCD	0,01032 {18}	0,00244 {54}	0,00058 {180}	0,06310 {15}	0,01414 {44}	0,00341 {141}
II на AB ($\eta=0$)	0,02234 {26}	0,00523 {80}	0,00130 {263}	0,05475 {21}	0,01253 {68}	0,00303 {226}
II на DC ($\eta=1$)	0,01250 {33}	0,00288 {108}	0,00074 {361}	0,06310 {15}	0,01414 {45}	0,00341 {142}
II на AD ($\xi=0$)	0,01042 {21}	0,00245 {65}	0,00066 {217}	0,06549 {27}	0,01484 {93}	0,00359 {320}
II на BC ($\xi=1$)	0,02303 {25}	0,00538 {81}	0,00130 {280}	0,07113 {32}	0,01550 {99}	0,00366 {297}
II кроме A, B, C, D	0,04624 {129}	0,01136 {527}	0,00293 {2019}	0,06581 {143}	0,01506 {606}	0,00391 {2266}

На рис. 3 показаны отклонения приближенного решения от точного первой краевой задачи и смешанной краевой задачи в области $\Omega_{2,xy}$ на сетке 64×64 .

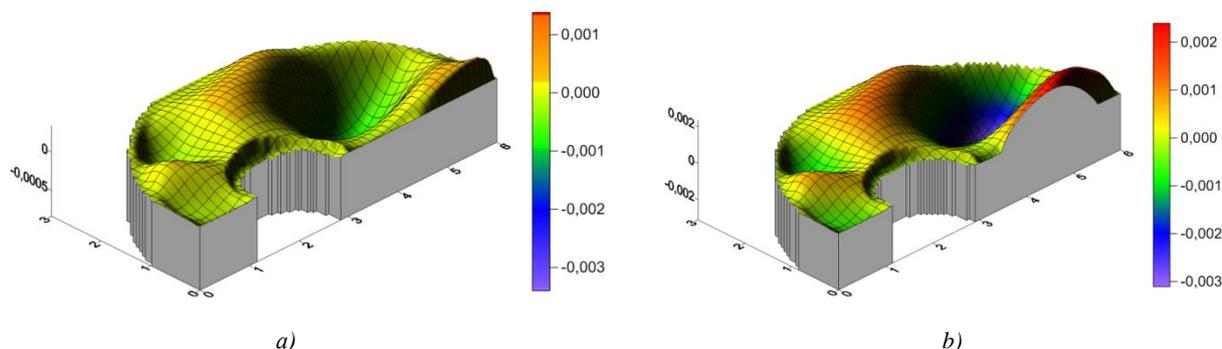


Рис. 3. Отклонения приближенного решения от точного в области $\Omega_{2,xy}$:
a) первой краевой задачи; b) смешанной краевой задачи

Fig. 3. Deviations of the approximate solution from the exact one in the domain $\Omega_{2,xy}$:
a) first boundary value problem; b) mixed boundary problem

Построенные в статье аппроксимации краевых условий допускают одновременное задание условий обоих типов на сторонах криволинейного четырехугольника. Так, в случае смешанной краевой задачи (рис. 3, b) условия Неймана задавались на участке $3,5 < x < 5,5, y=0$ стороны AB криволинейного четырехугольника ABCD. На остальной части границы области $\Omega_{2,xy}$ задавались краевые условия Дирихле.

Заключение. В работе реализованы разностные схемы для решения смешанной задачи Неймана – Дирихле для уравнения Пуассона в сложных двумерных областях на неортогональных разностных сетках с фиксированными узлами, аппроксимирующие дифференциальную задачу второго порядка. Построены аппроксимации граничного условия Неймана второго порядка.

При построении вычислительного алгоритма использовался переход к обобщенной криволинейной системе координат. На дискретном уровне переход к криволинейной системе координат требует генерации в исходной нерегулярной области подходящей четырехугольной разностной сетки.

Для решения полученных систем девятиточечных разностных уравнений реализован вычислительный алгоритм модифицированного сильно неявного метода MSIM. Проведены численные эксперименты по сравнению аппроксимаций различных краевых задач Дирихле и Неймана для уравнения Пуассона в нерегулярных двумерных областях. Результаты экспериментов подтверждают второй порядок точности предложенного вычислительного алгоритма.

Вклад авторов. М. М. Чуйко построил аппроксимации второго порядка краевых условий Неймана в нерегулярной двумерной области с использованием обобщенных криволинейных координат. О. М. Королёва программно реализовала вычислительный алгоритм решения задачи Пуассона в криволинейных координатах, провела серию численных экспериментов, подтвердивших второй порядок точности вычислительного алгоритма.

Список использованных источников

1. Флетчер, К. Вычислительные методы в динамике жидкостей : пер. с англ. / К. Флетчер. – М. : Мир, 1991. – 295 с.
2. Самарский, А. А. Теория разностных схем / А. А. Самарский. – М. : Наука, 1997. – 380 с.
3. Самарский, А. А. Разностные методы для эллиптических уравнений / А. А. Самарский, В. Б. Андреев. – М. : Наука, 1976. – 352 с.
4. Монотонные разностные схемы для уравнений со смешанными производными / А. А. Самарский [и др.] // Математическое моделирование. – 2001. – Т. 13, № 2. – С. 17–26.
5. Monotone difference schemes for equations with mixed derivatives / A. Samarskii [et al.] // *Computers and Mathematics with Applications*. – 2002. – Vol. 44. – P. 501–510. [https://doi.org/10.1016/S0898-1221\(02\)00164-5](https://doi.org/10.1016/S0898-1221(02)00164-5)
6. Matus, P. Monotone difference schemes for equations with mixed derivatives / P. Matus, I. Rybak // *Computational Methods in Applied Mathematics*. – 2009. – Vol. 4, no. 4. – P. 494–505. <https://doi.org/10.2478/cmam-2004-0027>
7. Matus, P. Monotone and economical difference schemes on nonuniform grids for multidimensional parabolic equations with boundary condition of third kind / P. Matus, G. Martynkevich // *Computational Methods in Applied Mathematics*. – 2004. – Vol. 4, no. 3. – P. 350–367. <https://doi.org/10.2478/cmam-2004-0019>
8. Monotone finite difference schemes for quasilinear parabolic problems with mixed boundary conditions / F. J. Gaspar [et al.] // *Computational Methods in Applied Mathematics*. – 2016. – Vol. 16, no. 2. – P. 231–244. <https://doi.org/10.1515/cmam-2016-0002>
9. Schneider, G. E. A modified strongly implicit procedure for the numerical solution of field problem / G. E. Schneider, M. Zedan // *Numerical Heat Transfer*. – 1981. – Vol. 4. – P. 1–19.
10. Волков, В. М. Итерационная реализация разностных схем в методе фиктивных областей для эллиптических задач со смешанными производными / В. М. Волков, Е. В. Прокопина // *Журнал Бел. гос. ун-та. Математика. Информатика*. – 2019. – № 1. – С. 69–76. <https://doi.org/10.33581/2520-6508-2019-1-69-7>

References

1. Fletcher C. A. J. *Computational Techniques for Fluid Dynamics 2*. Springer, 1988, 494 p.
2. Samarskii A. A. *Teoriya raznostnyh shem. The Theory of Difference Schemes*. Moscow, Nauka, 1997, 380 p.
3. Samarskii A. A., Andreev V. B. *Raznostnye metody dlja elipticheskikh uravnenij. Difference Methods for Elliptic Equations*. Moscow, Nauka, 1976, 352 p. (In Russ.).
4. Samarskii A. A., Mazhukin V. I., Matus P. P., Shishkin G. I. *Monotone difference schemes for equation with mixed derivatives*. *Matematicheskoe modelirovanie [Mathematical Models and Computer Simulation]*, 2001, vol. 13, no. 2, pp. 17–26 (In Russ.).
5. Samarskii A., Matus P., Mazhukin V., Mozolevski I. Monotone difference schemes for equations with mixed derivatives. *Computers and Mathematics with Applications*, 2002, vol. 44, pp. 501–510. [https://doi.org/10.1016/S0898-1221\(02\)00164-5](https://doi.org/10.1016/S0898-1221(02)00164-5)
6. Matus P., Rybak I. Monotone difference schemes for equations with mixed derivatives. *Computational Methods in Applied Mathematics*, 2009, vol. 4, no 4, pp. 494–505. <https://doi.org/10.2478/cmam-2004-0027>

7. Matus P., Martsynkevich G. Monotone and economical difference schemes on nonuniform grids for multidimensional parabolic equations with boundary condition of third kind. *Computational Methods in Applied Mathematics*, 2004, vol. 4, no. 3, pp. 350–367. <https://doi.org/10.2478/cmam-2004-0019>

8. Gaspar F. J., Lisbona F. J., Matus P., Tuyen V. T. K. Monotone finite difference schemes for quasilinear parabolic problems with mixed boundary conditions. *Computational Methods in Applied Mathematics*, 2016, vol. 16, no. 2, pp. 231–244. <https://doi.org/10.1515/cmam-2016-0002>

9. Schneider G. E., Zedan M. A modified strongly implicit procedure for the numerical solution of field problem. *Numerical Heat Transfer*, 1981, vol. 4, pp. 1–19.

10. Volkov V. M., Prokonina E. V. *Iterative realization of finite difference schemes in the fictitious domain method for elliptic problems with mixed derivatives*. Zhurnal Belorusskogo gosudarstvennogo universiteta. Matematika. Informatika [Journal of the Belarusian State University. Mathematics and Informatics], 2019, no. 1, pp. 69–76 (In Russ.). <https://doi.org/10.33581/2520-6508-2019-1-69-76>

Информация об авторах

Чуйко Михаил Матвеевич, кандидат физико-математических наук, ведущий научный сотрудник отдела вычислительной математики, Институт математики НАН Беларуси.

E-mail: mikhail.chuiko@gmail.com

Королёва Ольга Михайловна, кандидат физико-математических наук, доцент кафедры высшей математики, Белорусский национальный технический университет.

E-mail: korolyovaola@gmail.com

Information about the authors

Mikhail M. Chuiko, Ph. D. (Phys.-Math.), Leading Researcher of the Department of Computational Mathematics, Institute of Mathematics of the National Academy of Sciences of Belarus.

E-mail: mikhail.chuiko@gmail.com

Olga M. Korolyova, Ph. D. (Phys.-Math.), Associate Professor of the Department of Higher Mathematics, Belarusian National Technical University.

E-mail: korolyovaola@gmail.com

Правила для авторов

Редакция журнала «Информатика» просит авторов руководствоваться приведенными ниже правилами.

I. Статьи принимаются в редакцию через электронную систему подачи по адресу <http://inf.grid.by> в формате файлов текстовых редакторов Microsoft Word. Объем оригинальной статьи – от 8 до 16 стр., включая рисунки, таблицы и достаточное количество наиболее актуальных ссылок; объем обзорной статьи – от 16 до 32 стр., включая все основные ссылки. Текст набирается с переносами, шрифт Times New Roman 11 пт, интервал между строками – одинарный, абзацный отступ 0,5 см, поля по 2,5 см со всех сторон.

Материал статьи должен быть четко структурированным: Введение; основные разделы, в которых изложены цели и задачи, методы, результаты; Заключение (выводы).

II. Статьи о результатах работ, проведенных в научных учреждениях, должны иметь разрешение на публикацию (сопроводительное письмо за подписью руководителя или выписку из заседания ученого совета, отдела или кафедры, акт экспертизы).

III. Статьи в обязательном порядке должны включать аннотацию, ключевые слова, список литературы, информацию об авторах на русском и английском языках.

На титульной странице располагаются следующие метаданные:

1. Индекс по универсальной десятичной классификации (УДК); на русском и английском языках тип статьи (оригинальная или обзорная), название статьи, инициалы и фамилии всех авторов, полное наименование учреждений, где работают авторы, с указанием почтового адреса, при наличии указывается ученая степень и ORCID, e-mail ответственного лица.

2. Аннотация (Abstract) объемом 150–250 слов в оригинальной статье должна быть структурирована отдельными подразделами: Цели, Методы, Результаты, Заключение, а также максимально характеризовать содержательную часть рукописи. Сюда не следует включать впервые введенные термины, аббревиатуры (за исключением общеизвестных), ссылки на литературу.

3. Ключевые слова (Keywords) – наиболее значимые слова или словосочетания по теме работы, отражающие специфику темы, объекты и результаты исследования; перечень ключевых слов должен содержать 5–10 слов.

4. В разделе Благодарности (Acknowledgements) указываются все источники финансирования исследования, а также благодарности людям, которые участвовали в работе над статьей.

5. Автор обязан уведомить редакцию о реальном или потенциальном конфликте интересов, включив информацию в раздел Конфликт интересов (Conflict of interest).

6. Формулы, рисунки, таблицы в статье нумеруются в соответствии с порядком их упоминания в тексте. Ссылки на рисунки и таблицы в тексте обязательны. Рисунки должны быть выполнены с хорошим разрешением в масштабе, позволяющем четко различать надписи и обозначения. Цветные иллюстрации печатаются только в том случае, когда это необходимо для понимания излагаемого материала. Подрисовочные подписи с расшифровкой всех позиций, представленных на рисунке, и названия таблиц набираются шрифтом гарнитуры основного текста размером 9 пт. Перевод подрисовочной подписи и пояснений к рисунку, а также перевод названия таблицы, заголовки строк или столбцов располагаются курсивом после русскоязычной версии.

7. Набор формул выполняется в формульном редакторе Microsoft Equation или Math Type. Прямым шрифтом набираются: греческие и русские буквы; математические символы (\sin , \lg , ∞); символы химических элементов (C, Cl, CH₃); цифры (римские и арабские); индексы (верхние и нижние), являющиеся сокращениями слов. Курсивом набираются латинские буквы, символы физических величин (в том числе и в индексе).

8. Список использованной литературы оформляется в соответствии с требованиями Высшей аттестационной комиссии Республики Беларусь (ГОСТ 7.5–2008). Номер литературной ссылки в тексте дается порядковым номером в квадратных скобках. Ссылаться на неопубликованные работы не допускается.

10. Отдельно оформляется References со следующей структурой: авторы (транслитерация), транслитерированное название монографии, *Перевод названия монографии на английский язык*. Выходные данные с обозначениями на английском языке. От транслитераций названий статей можно отказаться.

Ссылки на учебно-методическую литературу, ГОСТы, авторефераты, статистические отчеты в список не включаются, а оформляются в виде сносок (с подробными рекомендациями можно ознакомиться на сайте журнала в разделе Правила для авторов).

11. В разделе Информация об авторах (Information about the authors) приводятся ФИО авторов полностью, ученая степень, звание, должность, название организации, ORCID (при наличии).

IV. Все поступающие в редакцию рукописи проходят предварительную проверку на соответствие Правилам для авторов. Статья может быть возвращена автору на доработку с просьбой устранить недостатки или дополнить информацию. После проверки на соответствие правилам статья направляется рецензенту с указанием сроков рецензирования.

V. При наличии замечаний рецензента автору предоставляется определенное время на доработку рукописи. Статьи, направляемые на доработку, должны быть возвращены в исправленном виде с ответами на все замечания. Окончательное решение о публикации или отклонении рукописи принимается редколлегией журнала. При положительном заключении рецензента статья передается редактору для подготовки к печати. Редакция оставляет за собой право на редакционные изменения, не искажающие основное содержание статьи.

VI. Редакция журнала предоставляет возможность первоочередного опубликования статей, представленных лицами, которые осуществляют послевузовское обучение (аспирантура, докторантура, соискательство) в год завершения обучения.

VII. Авторы несут ответственность за направление в редакцию статей, уже опубликованных ранее или принятых к публикации другими изданиями.

ИНДЕКСЫ

00827

для индивидуальных
подписчиков

008272

для предприятий
и организаций

2 | 2023

ТОМ ■ VOL. 20

ИНФОРМАТИКА ■ INFORMATICS

ISSN 1816-0301 (Print)



9 771816 030000