

ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ, МЕХАНИКИ И ОПТИКИ

2023, ТОМ 23, НОМЕР 1 (ЯНВАРЬ–ФЕВРАЛЬ)

ISSN 2226-1494 (PRINT), 2500-0373 (ONLINE)

ИЗБРАННЫЕ МАТЕРИАЛЫ XXXII ШКОЛЫ ПО ГОЛОГРАФИИ

Часть III

Характеризация голографического фотополимера Bayfol HX в инфракрасной области спектра

Исследование реакции кровеносных сосудов на локальный нагрев методом визуализирующей фотоплетизмографии

Передача 3D-голографической информации по радиоканалу методом, близким к SSB

НАНОСТРУКТУРЫ. СВОЙСТВА И ПРИМЕНЕНИЯ

Влияние параметров анодирования на оптические свойства нанопористого оксида алюминия, сформированного поверх островковой пленки серебра

АВТОМАТИЧЕСКОЕ УПРАВЛЕНИЕ И РОБОТОТЕХНИКА

Анализ точности оценивания состояний асинхронного электропривода алгоритмами Люенбергера и Калмана

Метод оптимизации структуры в иерархических распределенных системах управления

Метод идентификации параметров синусоидального сигнала с неизвестной переменной амплитудой

НОВЫЕ МАТЕРИАЛЫ И НАНОТЕХНОЛОГИИ

Улучшение процесса автоматической стабилизации температуры в криовакуумной установке (на англ. яз.)

Исследование ударных свойств и износостойкости гибридных композитов Al6061 (SiC + Al₂O₃) и Al7075 (SiC + Al₂O₃) (на англ. яз.)

КОМПЬЮТЕРНЫЕ СИСТЕМЫ И ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ

Вычислительные методы повышения быстродействия дискретного вейвлет-преобразования на базе FPGA

Диалоговая система на основе устных разговоров с доступом к неструктурированной базе знаний

ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ И КОГНИТИВНЫЕ ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ

Определение аналитических моделей динамических систем в форме дифференциальных уравнений на основе многокритериальной эволюционной оптимизации (на англ. яз.)

Вероятностный критерий оценки предсказуемости временных рядов (на англ. яз.)

Программный фреймворк для оптимизации гиперпараметров тематических моделей с аддитивной регуляризацией

Ценностно-ориентированное моделирование принятия экономических решений в условиях нестационарности внешней среды

Методика организации и проведения исследования по оценке потребительской способности

Автоматизированный кластерный анализ коммуникативных стратегий образовательных telegram-каналов

МАТЕМАТИЧЕСКОЕ И КОМПЬЮТЕРНОЕ МОДЕЛИРОВАНИЕ

Компьютерное моделирование немарковских процессов на основе принципа баланса «комплексных вероятностей»

Метод объективизации значений весовых коэффициентов для принятия решений в многокритериальных задачах

Оценка границ применимости и методов модуляции ближнепольной магнитной связи

Численно-аналитическое моделирование пропульсивного крыла и фюзеляжа аэротакси

Применение обратной связи переменной точности для повышения быстродействия контура тока в инверторах на базе GaN-транзисторов

Моделирование диффузионных процессов при электротермической обработке реакционных тиглей системы железо–олово

Борисов В.Н., Зверев А.Д., Камынин В.А., Копьева М.С.,
Окунь Р.А., Цветков В.Б. 1

Белавенцева А.В., Подолян Н.П., Волынский М.А.,
Зайцев В.В., Саковская А.В., Мамонтов О.В.,
Ромашко Р.В., Камшилин А.А. 14

Пазоев А.Л., Шойдин С.А. 21

Никитин И.Ю., Набиуллина Р.Д., Нащекин А.В.,
Старовойтов А.А., Гладских И.А. 28

Букреев В.Г., Шандарова Е.Б., Перевоицков Ф.В. 35

Онуфрей А.Ю., Разумов А.В., Какаев В.В. 44

Нгуен Х.Т., Власов С.М. 54

Голиков О.Ю., Ережеев Д., Соколов Д.Ю. 62

Равичандаран Р., Сельварасу С., Гопал С., Рамачандран Р. 68

Сай С.В., Зинкевич А.В. 79

Маслюхин С.М. 88

Масляев М.А., Хватов А.А. 97

Кованцев А.Н. 105

Ходорченко М.А., Бутаков Н.А., Насонов Д.А.,
Фирулик М.Ю. 112

Гулева В.Ю., Кованцев А.Н., Суриков А.Г., Чунаев П.В.,
Горнова Г.В., Бухановский А.В. 121

Алейников С.А., Гофман О.О., Басов О.О. 136

Низомутдинов Б.А., Углова А.Б., Богдановская И.М. 142

Гусеница Я.Н., Ширямов О.А. 150

Соловьев Д.С. 161

Гришенцев А.Ю., Горошков В.А., Чернов Р.И. 169

Булат П.В., Курнухин А.А., Продан Н.В. 178

Анучин А.С., Гуляева М.А., Лашкевич М.М., Жарков А.А.,
Чен Х., Дианов А.Н. 192

Фомин В.Е., Тукмакова А.С., Болкунов Г.А.,
Новотельнова А.В., Бочканов Ф.Ю., Карпенков Д.Ю. 202

Главный редактор — В.О. Никифоров, д.т.н., профессор

Editor-in-Chief — Vladimir O. Nikiforov, D.Sc., Professor

РЕДАКЦИОННАЯ КОЛЛЕГИЯ

Беззатеев С.В. — д.ф.-м.н., доцент (Россия)
Беликов А.В. — д.ф.-м.н., профессор (Россия)
Бессмертный И.А. — д.т.н., профессор (Россия)
Бобцов А.А. — д.т.н., профессор (Россия)
Бухановский А.В. — д.т.н. (Россия)
Волков К. Н. — д.ф.-м.н. (Великобритания)
Дукельский К.В. — к.т.н., доцент (Россия)
Ефимов Д.В. — PhD (Франция)
Заколдаев Д.А. — к.т.н., доцент (Россия)
Иорш И.В. — д.ф.-м.н. (Россия)
Кармановский Н.С. (заместитель главного редактора) — к.т.н. доцент (Россия)
Комаров И.И. — к.ф.-м.н., доцент (Россия)
Никоноров Н.В. — д.т.н., профессор (Россия)
Полищук Г.С. — к.т.н. (Россия)
Романов А.Е. — д.ф.-м.н., профессор (Россия)
Сидоркина И.Г. — д.т.н., профессор (Россия)
Степанов О.А. — д.т.н., профессор (Россия)
Татарникова Т.М. — д.т.н., доцент (Россия)
Тимофеев А. В. — д.т.н. (Казахстан)
Успенская М.В. — д.т.н., профессор (Россия)
Цыпкин А. Н. — д.ф.-м.н. (Россия)

EDITORIAL BOARD

Sergey V. Bezzateev — D.Sc., Associate Professor (Russia)
Andrey V. Belikov — D.Sc., Professor (Russia)
Igor A. Bessmertny — D.Sc., Professor (Russia)
Alexey A. Bobtsov — D.Sc., Professor (Russia)
Alexandr V. Boukhanovsky — D.Sc. (Russia)
Konstantin N. Volkov — D.Sc. (Great Britain)
Konstantin V. Dukelskii — PhD, Associate Professor (Russia)
Denis V. Efimov — PhD (France)
Danil A. Zakoldaev — PhD, Associate Professor (Russia)
Ivan V. Iorsh — D.Sc. (Russia)
Nikolai S. Karmanovskiy (Deputy Chief Editor) — PhD, Associate Professor (Russia)
Igor I. Komarov — PhD, Associate Professor (Russia)
Nikolay V. Nikonorov — D.Sc., Professor (Russia)
Grigorii S. Polishchuk — PhD (Russia)
Alexey E. Romanov — D.Sc., Professor (Russia)
Irina G. Sidorkina — D.Sc., Professor (Russia)
Oleg A. Stepanov — D.Sc., Professor (Russia)
Tatiana M. Tatarnikova — D.Sc., Associate Professor (Russia)
Andrey V. Timofeev — Dr. habil. Sc. (Kazakhstan)
Mayya V. Uspenskaya — D.Sc., Professor (Russia)
Anton N. Tsyppkin — D.Sc., Professor (Russia)

Ответственный секретарь — И.В. Малькова

Executive secretary — Irina V. Malkova

Учредитель: Университет ИТМО

Издание зарегистрировано в Федеральной службе по надзору в сфере связи, информационных технологий и массовых коммуникаций (свидетельство ПИ №ФС77-67990 от 6 декабря 2016 г.).

Языки журнала: русский, английский.

ISSN 2226-1494 (print version), ISSN 2500-0373 (online version)

Англоязычное название: «Scientific and Technical Journal of Information Technologies, Mechanics and Optics». Транслитерированное название: «Nauchno-Tekhnicheskii Vestnik Informatsionnykh Tekhnologii, Mekhaniki i Optiki». Сокращенное название: «Sci.Tech. J. Inf. Technol. Mech. Opt.»

Журнал включен в каталог периодических изданий Ulrich's Periodical Directory.

Журнал входит в крупнейшие международные реферативные и наукометрические базы Scopus, EBSCO, ProQuest, EastView, IBIS и другие.

Журнал входит в утвержденный Высшей аттестационной комиссией «Перечень ведущих рецензируемых научных журналов и изданий, выпускаемых в Российской Федерации, в которых должны быть опубликованы основные научные результаты диссертаций на соискание ученой степени доктора и кандидата наук».

Периодичность издания — 6 выпусков в год.

Плата за публикации и редактирование не взимается.

Редакторы: Ю.В. Михайлова, Л.Н. Точилина, Т.В. Точилина
Перевод — Н.А. Корнев
Компьютерная верстка — Е.С. Егорова

Founder: ITMO University

Publication is registered by the Federal Inspectorate Service for Communication, Information Technologies and Communication Media, certificate PI FS77-67 990 dated December, 6th, 2016.

Languages of the journal: Russian, English

ISSN 2226-1494 (print version), ISSN 2500-0373 (online version)

The English title is "Scientific and Technical Journal of Information Technologies, Mechanics and Optics".

Transliterated title is "Nauchno-Tekhnicheskii Vestnik Informatsionnykh Tekhnologii, Mekhaniki i Optiki".

Abbreviated parallel title: "Sci.Tech. J. Inf. Technol. Mech. Opt."

The journal is catalogued in Ulrich's Periodical Directory.

The journal is included in the largest international abstract and scientometric databases Scopus, EBSCO, ProQuest, East View, EVIS and others.

The journal is included in the "List of leading scientific journals and periodicals under review in the Russian Federation, where the main scientific results of theses for the PhD and doctor of sciences degree must be published" approved by the Higher Attestation Commission.

Publication frequency is 6 times a year.

Publication and editing are free of charge.

Editors: Yulia V. Mikhailova, Lyubov N. Tochilina, Tatiana V. Tochilina
English Language Technical Editor — Nikolay A. Kornev
Computer layout — Ekaterina S. Egorova

Оф. 2136, Университет ИТМО, ул. Ломоносова, д. 9,
Санкт-Петербург, Российская Федерация, 191002

of. 2136, ITMO University, Lomonosova Street, 9,
St. Petersburg, 191002, Russia

Телефон (812) 480 02 75
<http://ntv.ifmo.ru>
E-mail: ntvitmo@itmo.ru

Phone (812) 480 02 75
<http://ntv.ifmo.ru/en>
E-mail: ntvitmo@itmo.ru

Статьи Журнала доступны по лицензии Creative Commons
«Attribution-NonCommercial» 4.0 Всемирная.



Подписано к печати 01.02.2023 Тираж 350 экз. Заказ № 91 (143)

Издание Университета ИТМО

Отпечатано в ООО «Университетские телекоммуникации»
Адрес: 199034, Санкт-Петербург, Биржевая линия, д. 16

© Университет ИТМО

ИЗБРАННЫЕ МАТЕРИАЛЫ XXXII ШКОЛЫ ПО ГОЛОГРАФИИ. ЧАСТЬ III PROCEEDINGS OF THE XXXII SCHOOL ON HOLOGRAPHY. PART III

doi: 10.17586/2226-1494-2023-23-1-1-13

УДК 535.421, 535.15

Характеризация голографического фотополимера Bayfol HX в инфракрасной области спектра

Владимир Николаевич Борисов¹, Андрей Дмитриевич Зверев²,
Владимир Александрович Камынин³, Мария Сергеевна Копьева⁴,
Роман Александрович Окунь⁵, Владимир Борисович Цветков⁶

¹ Независимый исследователь, Санкт-Петербург, Российская Федерация

^{2,3,4,5,6} Институт общей физики им. А.М. Прохорова Российской академии наук, Москва, 119991, Российская Федерация

⁴ Российский университет дружбы народов, Москва, 117198, Российская Федерация

¹ pvlc2013@gmail.com, <https://orcid.org/0000-0003-3525-5698>

² izverevad@gmail.com, <https://orcid.org/0000-0002-3383-906X>

³ kamyninva@gmail.com, <https://orcid.org/0000-0002-4265-3478>

⁴ mashutka_kopyova@mail.ru, <https://orcid.org/0000-0002-2113-6607>

⁵ romaokun@rambler.ru, <https://orcid.org/0000-0003-2461-9978>

⁶ tsvetkov@lsk.gpi.ru, <https://orcid.org/0000-0002-1483-3308>

Аннотация

Предмет исследования. Рассмотрена возможность создания дифракционных элементов на основе голографического фотополимера Bayfol HX, работающих в ближнем инфракрасном диапазоне спектра. Исследованы динамический диапазон показателя преломления фотополимера и амплитудно-фазовые характеристики голограмм в инфракрасном диапазоне. Изучено влияние параметров записи (плотности мощности записывающего излучения, времени записи) на распределение динамического диапазона показателя преломления между гармониками решетки. **Метод.** Анализ амплитудно-фазового характера голограмм выполнен с помощью измерения спектров пропускания исследуемого фотополимера после реакции фотополимеризации. Проведена оценка динамического диапазона показателя преломления фотополимера в спектральном диапазоне от 405 до 2099 нм. Предварительно осуществлены измерение и анализ контуров угловой селективности голограмм с периодами от 414 до 2100 нм, оптимизированных под разные части исследуемого спектрального диапазона. Выполнен анализ влияния параметров записи на распределение динамического диапазона показателя преломления между гармониками решетки. Произведен расчет амплитуд первой и второй гармоник модуляции показателя преломления из экспериментально измеренных контуров угловой селективности голограмм, сформированных при разных длительностях записи при постоянной дозе облучения. **Основные результаты.** Показано, что динамические диапазоны показателя преломления фотополимера в ближнем инфракрасном диапазоне и в длинноволновой части видимого диапазона отличаются на величину, не превышающую точности измерений. Продemonстрировано ярко выраженное нарушение закона взаимозаместимости при масштабировании интерференционной картины и/или изменении плотности мощности записывающего излучения. **Практическая значимость.** Найдены оптимальные условия записи голограмм в исследуемом фотополимере для их применения в инфракрасном диапазоне спектра. Исследованный голографический материал может найти применение в телекоммуникационной оптике.

Ключевые слова

голография, Bayfol, фотополимер, инфракрасное излучение, дифракционная оптика

Благодарности

Работа выполнена на базе Научного центра мирового уровня «Фотоника» при финансовой поддержке Министерства науки и высшего образования Российской Федерации (грант № 075-15-2020-912). Авторы благодарят доктора Фридриха-Карла Брудера за предоставление голографической среды Bayfol HX, а также Василия Валерьевича Лесниченко за консультации по вопросам анализа экспериментальных результатов.

Ссылка для цитирования: Борисов В.Н., Зверев А.Д., Камынин В.А., Копьева М.С., Окунь Р.А., Цветков В.Б. Характеризация голографического фотополимера Bayfol HX в инфракрасной области спектра // Научно-технический вестник информационных технологий, механики и оптики. 2023. Т. 23, № 1. С. 1–13. doi: 10.17586/2226-1494-2023-23-1-1-13

Characterization of the holographic photopolymer Bayfol HX in the IR region

Vladimir N. Borisov¹, Andrey D. Zverev², Vladimir A. Kamynin³, Maria S. Kopyeva⁴,
Roman A. Okun⁵, Vladimir B. Tsvetkov⁶

¹ Independent Researcher, Saint Petersburg, Russian Federation

^{2,3,4,5,6} Prokhorov General Physics Institute of the Russian Academy of Sciences, Moscow, 119991, Russian Federation

⁴ RUDN University, Moscow, 117198, Russian Federation

¹ pvlc2013@gmail.com, <https://orcid.org/0000-0003-3525-5698>

² izverevad@gmail.com, <https://orcid.org/0000-0002-3383-906X>

³ kamyninva@gmail.com, <https://orcid.org/0000-0002-4265-3478>

⁴ mashutka_kopyeva@mail.ru, <https://orcid.org/0000-0002-2113-6607>

⁵ romaokun@rambler.ru, <https://orcid.org/0000-0003-2461-9978>

⁶ tsvetkov@lsk.gpi.ru, <https://orcid.org/0000-0002-1483-3308>

Abstract

The possibility of creating holographic optical elements operating in the near infrared spectral range based on the Bayfol HX holographic photopolymer has been considered. The dynamic range of the refractive index of the photopolymer and the amplitude-phase nature of the holograms in the infrared range have been studied. The influence of recording parameters (power density of recording radiation, recording time) on the distribution of the dynamic range of the refractive index between grating harmonics has been studied. The analysis of the amplitude-phase nature of holograms was carried out by measuring the transmission spectra of the studied photopolymer after the photopolymerization reaction. The dynamic range of the refractive index of a photopolymer evaluated in the spectral range from 405 nm to 2099 nm. For this purpose, the angular selectivity contours of holograms with periods from 414 nm to 2100 nm, optimized for different parts of the specified spectral range, were measured and analyzed. The influence of recording parameters on the distribution of the dynamic range of the refractive index between the grating harmonics was analyzed by calculating the amplitudes of the first and second harmonics of the refractive index modulation from the experimentally measured angular selectivity contours of holograms recorded with different recording time at a constant irradiation dose. It was shown that the dynamic range of the refractive index of the photopolymer in the near infrared spectral range, as compared with the long-wavelength part of the visible region of the spectrum, differs by a value that does not exceed the measurement accuracy. A pronounced violation of the reciprocity was demonstrated with scaling of the interference pattern or with changing of the power density of the recording radiation. The optimal recording conditions for holograms calculated for the infrared spectral range for the studied photopolymer were found. The possibility of using of the studied holographic material in telecommunication optics has been demonstrated.

Keywords

holography, Bayfol, photopolymer, IR radiation, diffractive optics

Acknowledgements

This research was funded with the financial support of the Ministry of Science and Higher Education of the Russian Federation, grant number 075-15-2020-912, and carried out on the basis of the World-Class Research Center “Photonics”. The authors thank Dr. Friedrich-Karl Bruder for providing the Bayfol HX holographic medium, and Lesnichii Vasilii Valerievich for consultations on the analysis of experimental results.

For citation: Borisov V.N., Zverev A.D., Kamynin V.A., Kopyeva M.S., Okun R.A., Tsvetkov V.B. Characterization of the holographic photopolymer Bayfol HX in the IR region. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 2023, vol. 23, no. 1, pp. 1–13 (in Russian). doi: 10.17586/2226-1494-2023-23-1-1-13

Введение

В настоящее время голографические оптические элементы активно применяются для управления излучением в широком спектральном диапазоне: от ближнего ультрафиолетового [1] до среднего инфракрасного (ИК) [2]. При решении различных задач возникает потребность как в высокоселективных брэгговских зеркалах, так и в низкоселективных спектральных фильтрах и диспергирующих элементах. На сегодняшний день наиболее распространенный и коммерчески успешный голографический фотополимер, позволяющий получать низкоселективные голографические структуры — Bayfol HX [3]. Данный материал используется

для решения максимально широкого круга задач [4–12]. Голографическая запись в фотополимере происходит при использовании видимого излучения, тем самым в большинстве случаев ограничивая спектральную область применения. В работе [13] предложена модификация материала, позволяющая производить голографическую запись в ближнем ИК диапазоне, что может расширить область применения фотополимера. Но на данный момент характеристики оригинального Bayfol HX 200 в ближнем ИК диапазоне (длина волны 1–2,1 мкм) не исследованы.

В настоящей работе исследованы спектральные характеристики пропускания фотополимера Bayfol HX 200 в ближней ИК области спектра. Сделан вывод

об амплитудно-фазовом характере голограмм в этой области спектра. В фотополимере записаны несколько серий голограмм. Параметры голограмм лежат в диапазонах: периоды голограмм — от 414 до 2100 нм, плотность мощности записывающего излучения — от 10 до 500 мВт/см², время записи — от 32 до 1000 с. Голограммы записаны в видимом диапазоне излучения. Выполнен анализ модуляции показателя преломления в диапазоне длин волн от 405 до 2100 нм. Рассмотрено влияние условий записи (период интерференционной картины, плотность мощности записывающего излучения) на распределение динамического диапазона (Δn) фотополимера между гармониками голографической решетки. Исследовано нарушение закона взаимозаместимости при изменении условий записи, а также влияние кислорода на процесс записи.

Материалы и методы

Голографическая среда. Голографический материал Bayfol HX 200 реализует механизм записи с ортогональной химией [14]: исходная композиция содержит по два типа мономеров и инициаторов полимеризации. В материале использованы два механизма полимеризации (две пары «мономер-инициатор»), которые не подвержены перекрестным реакциям. Первый механизм имеет низкую скорость и инициируется за счет теплового воздействия, а второй — запускается благодаря фотореакции, и его компоненты — ключевые для голографической записи. Исходная композиция нанесена слоем толщиной 16 мкм на полимерную подложку из триацетата целлюлозы толщиной 60 мкм и подвергнута тепловому воздействию, благодаря чему происходит полимеризация по первому механизму. Полимеризовавшаяся часть композиции выполняет функцию связующего (биндер [15]) — жесткой матрицы, в которой находится оставшаяся часть композиции. В таком виде материал готов к голографической записи.

Механизм записи голограмм в материале Bayfol HX основан на радикальной фотополимеризации и фотоиндуцированной диффузии [16] (другое название — фотоиндуцированное разделение фаз [17]). В простейшем варианте данный эффект может быть рассмотрен для двухкомпонентных систем, включающих в себя мономер и химически нейтральную компоненту. За счет фотополимеризации при голографической записи образуется периодический градиент химического потенциала, и создается диффузионный поток мономера в зоны полимеризации. Таким образом, химически нейтральная компонента вынужденно вымещается из зон полимеризации для сохранения объема. В результате голографической записи возникает противofазное периодическое распределение двух химических фаз: полимера и химически нейтральной компоненты. Данный эффект можно в грубом приближении назвать обратным осмосом [18].

Методы исследования и аппаратура. Важную информацию об амплитудно-фазовом характере голограмм и о динамическом диапазоне показателя преломления голографического материала в разных спектральных областях может показать его спектр эк-

стинкции — мнимая часть комплексного показателя преломления:

$$n' = n - i \times \kappa,$$

где n — действительная часть показателя преломления; κ — коэффициент экстинкции, который имеет вид [19]

$$\kappa = \frac{\alpha \lambda}{4\pi}, \quad (1)$$

где λ — длина волны пробного излучения; α — натуральный показатель поглощения среды

$$\alpha = \frac{1}{l} \ln \left(\frac{1}{T} \right),$$

где l — протяженность среды; T — коэффициент пропускания среды.

Проведем измерения спектра экстинкции исследуемого материала в спектральном диапазоне 800–2300 нм при помощи спектрофотометра Shimadzu UV-3101PC. Из-за малой толщины (16 мкм) и высокой эластичности спектр экстинкции непосредственно фотополимера измерить сложно. Для получения информации о пропускании материала в ИК области используем три образца:

образец 1 — голографический фотополимер до обесцвечивания (в светочувствительном состоянии) на подложке из триацетата целлюлозы;

образец 2 — голографический фотополимер после обесцвечивания (в несветочувствительном состоянии) на подложке из триацетата целлюлозы;

образец 3 — подложка из триацетата целлюлозы, с которой фотополимер был полностью удален механически.

Из полученного спектра образца 2 вычтем спектр образца 3 и, таким образом, получим спектр экстинкции фотополимера после реакции фотополимеризации. Также найдем разность спектров образцов 1 и 2 и покажем отличие спектров экстинкции фотополимера до и после реакции фотополимеризации.

Анализ модуляции показателя преломления исследуемого фотополимера осуществим с помощью записи голографических решеток и измерения их контуров угловой селективности. Запись голограмм проведем с использованием двулучевого интерферометра. Для записи применим одномодовый (TEM⁰⁰) непрерывный одночастотный твердотельный лазер с диодной накачкой (Cobolt Bolero DPSS) с длиной волны излучения $\lambda = 639,6$ нм и выходной мощностью 500 мВт. На время записи фотополимер расположим на темном стекле с оптической плотностью на длине волны записи $D \approx 7$ для исключения отражения света от выходной грани образца и записи побочных голограмм. Отметим, что Bayfol HX обладает хорошими адгезионными свойствами, благодаря чему имеется возможность накатывать его на стеклянные поверхности, создавая границу раздела сред фотополимер — стекло. После записи голограммы были обесцвечены белым некогерентным светом с плотностью мощности около 20 мВт/см² в течение 60 мин [16] и перенесены на прозрачную стеклянную подложку.

Для измерения угловой селективности голограмм используем лазерные диоды с длинами волн: 405, 450, 635, 850, 980, 1062, 1549 нм, вторая гармоника твердотельного Nd:YAG лазера с длиной волны излучения 532 нм и волоконный гольмиевый лазер с длиной волны излучения 2099 нм [20]. Лазерное излучение s-поляризовано при помощи поляризатора с соотношением более 100:1. Голограммы установим на электромеханическом поворотном столике Thorlabs HDR50/M с заявленным минимальным шагом порядка 0'0.1", шаг между измерениями постоянен и находится для разных серий образцов в пределах от 3' до 4'12".

Контуры угловой селективности, полученные на разных длинах волн, с учетом Френелевского отражения, аппроксимируем при помощи теории связанных волн Когельника [21]. Построение теоретического контура угловой селективности выполним отдельно для каждой гармоники голограммы, видимой в экспериментальном контуре, по теории связанных волн для случая фазовых пропускающих решеток. Далее осуществим суммирование отдельных контуров пропорционально значениям эффективности (рис. 1, а). Полученную зависимость вычтем из единицы. Теоретический контур сравним с экспериментальным, и путем подбора параметров голограммы (толщины, модуляции показателя преломления) получим минимальные отличия контуров (рис. 1, б). Аппроксимация экспериментального контура угловой селективности теорией связанных волн позволила однозначно определить модуляцию показателя преломления голограммы. В результате полученная точность составила до $\Delta n_1 = \pm 0,001$.

Анализ каждой голограммы осуществлялся в определенном спектральном диапазоне, обладающим длинноволновой и коротковолновой границами. Минимальная длина волны восстановления обусловлена требованием величины угла дифракции менее 50°, поскольку при больших углах восстановления существенно изменяется апертура исследуемой области голограм-

мы из-за наклонного падения пучка. Максимальная длина волны восстановления — требованием углового расстояния между исследуемыми порядками дифракции более 14°, для минимизации влияния порядков дифракции друг на друга.

Экспериментальные результаты

Спектры экстинкции фотополимера в ИК диапазоне. Исследуем амплитудно-фазовый характер голограмм в ИК диапазоне, записанных в Bayfol HX 200, для этого по формуле (1) рассчитаем спектр экстинкции фотополимера после реакции фотополимеризации, который соответствует спектру восстановления записанной голограммы (рис. 2, а). Видно, что на приведенном спектре отсутствуют полосы поглощения, а коэффициент экстинкции соответствует значению $\approx 0,5 \cdot 10^{-4}$. Можно сделать вывод о преимущественно фазовом характере голограмм в данном спектральном диапазоне. Также показано отличие спектров экстинкции фотополимера до и после реакции фотополимеризации (рис. 2, б). Наблюдаемые биения на приведенных спектрах связаны с интерференцией света в образцах, представляющих собой пленки толщинами 60 мкм (образец 3) и 76 мкм (образцы 1 и 2).

Спектр на рис. 2, б демонстрирует, что существенные изменения поглощения материала в ближнем ИК диапазоне при фотополимеризации отсутствуют. Разностный спектр экстинкции в своих предельных значениях оценен числом $\approx 1,5 \cdot 10^{-4}$, а модуляция показателя преломления исследуемого фотополимера в видимом диапазоне длин волн оценена значением $3 \cdot 10^{-2}$, что в 200 раз больше. Отсюда следует, что модуляция показателя преломления материала не обусловлена изменением поглощения в ИК диапазоне. Таким образом, по спектрам пропускания в видимом диапазоне длин волн (из работ [3, 7] известно, что после фотополимеризации и обесцвечивания в диапазоне длин волн 400–800 нм

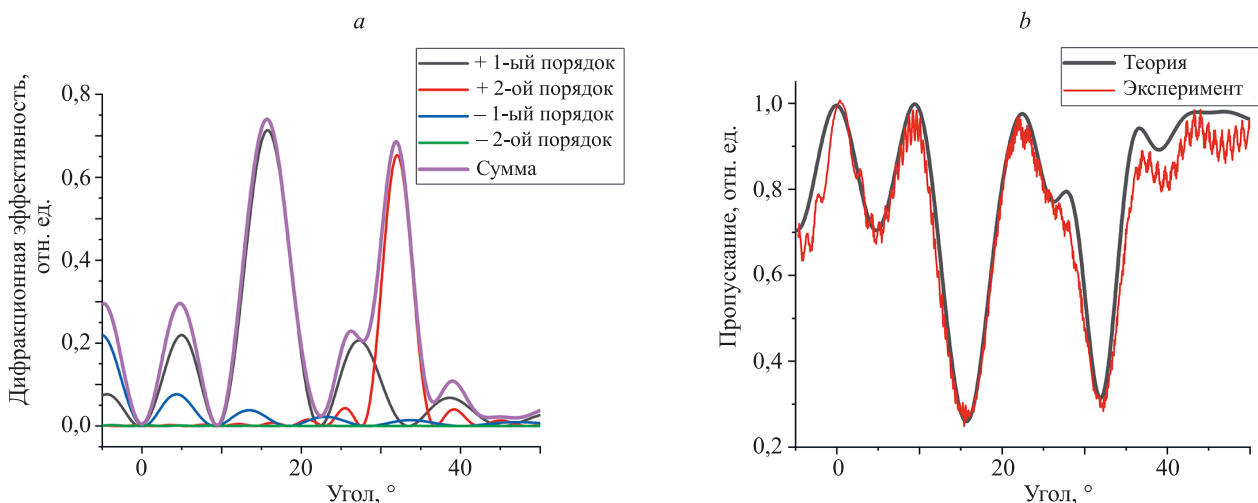


Рис. 1. Пример аппроксимации экспериментального контура теорией связанных волн. Теоретические контуры +1-го, +2-го, -1-го и -2-го порядков дифракции и их сумма (а), теоретический и экспериментальный контуры нулевого порядка дифракции (б)

Fig. 1. An example of approximation of an experimental contour by the Coupled waves theory. Theoretical contours of the +1st, +2nd, -1st and -2nd diffraction orders and their sum (а), theoretical and experimental zero-order diffraction contours (б) are shown

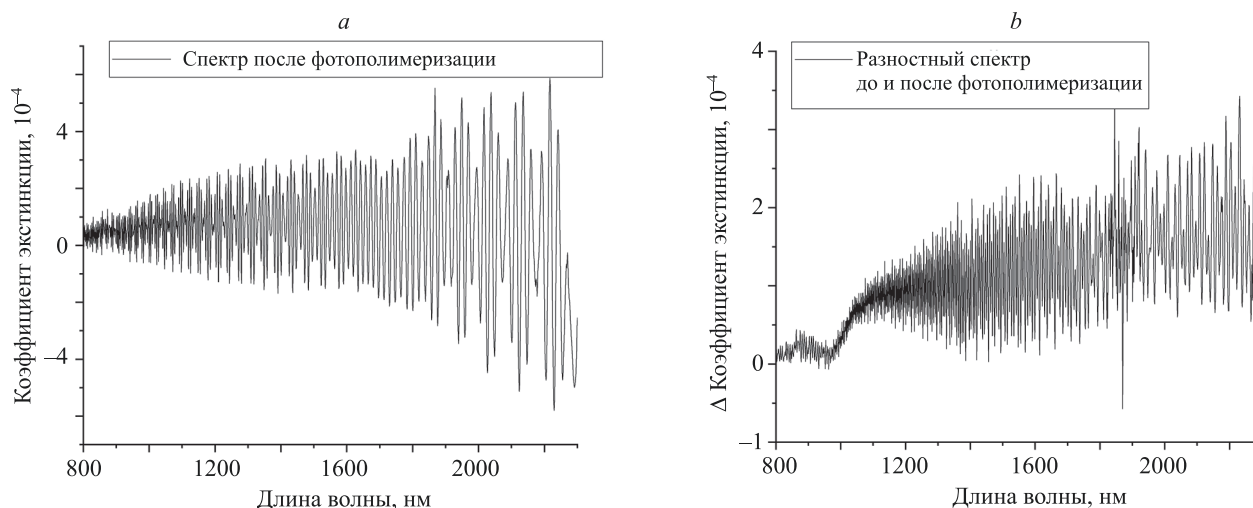


Рис. 2. Спектр экстинкции фотополимера Bayfol HX 200 после фотополимеризации (a) и разность спектров экстинкции до и после фотополимеризации (b)

Fig. 2. Extinction spectrum of Bayfol HX 200 photopolymer after photopolymerization (a) and difference between extinction spectra before and after photopolymerization (b)

материал характеризуется пропусканием более 85 %) и спектрам экстинкции в ИК диапазоне, приведенным в настоящей работе, можно предположить наличие модуляции коэффициента экстинкции только в ультрафиолетовом диапазоне спектра, и, в соответствии с соотношениями Крамерса–Кронига [19] (асимптотическое убывание модуляции показателя преломления с ростом длины волны). При изучении других материалов при помощи соотношений Крамерса–Кронига получено, что модуляция показателя преломления асимптотически уменьшается при отдалении от максимума модуляции коэффициента экстинкции [22].

Спектр модуляции показателя преломления. Основной вопрос, исследуемый в данной работе — возможность достижения высокой модуляции показателя преломления голограмм в ИК диапазоне, находящемся на достаточном спектральном отдалении от максимума модуляции коэффициента экстинкции.

Для сравнения модуляции показателя преломления в видимой и ИК областях спектра в фотополимере Bayfol HX 200 толщиной 16 мкм запишем 6 пропускающих решеток с периодами (d) от 414 до 2100 нм и нормальной ориентацией голографических плоскостей к поверхности элемента (образец 4). Плотность мощности каждого из записывающих пучков в области установки образца составила $P_{\text{зап}} = 500$ мВт/см². Время записи выбрано равным $t = 32$ с, для достижения плотности энергии записывающего излучения (дозы, $E = 2P_{\text{зап}}t \cos\theta$, где θ — угол падения или половинный угол сведения) для каждой голограммы $E > 20$ мДж/см², которая является достаточной для получения максимальной модуляции показателя преломления в фотополимере (табл. 1).

Для построения спектра модуляции показателя преломления измерим контуры угловой селективности голограмм в нулевом порядке дифракции на нескольких длинах волн от 405 до 2099 нм. По результатам анализа контуров голограмм образца 4 получим амплитуды первой (n_1) и второй (n_2) гармоник модуляции показателя

преломления исследуемого фотополимера в разных частях спектрального диапазона (рис. 3).

По полученным результатам видно, что с ростом периода амплитуды гармоник модуляции показателя преломления (n_1 и n_2) снижаются. Сумма n_1 и n_2 для решеток с периодами в диапазоне 661–2100 нм также монотонно снижается со значения 0,054 для решетки с периодом 661 нм до — 0,031 для решетки с периодом 2100 нм. Снижение n_1 и n_2 при увеличении периода решетки является явным нарушением закона взаимозаместимости (то же, что и закон Бунзена–Роско [23]), так как он постулирует зависимость амплитуд всех гармоник модуляции показателя преломления только от дозы и инвариантность условиям ее достижения. Наблюдаемое на рис. 3 снижение n_1 и n_2 может объясняться уменьшением динамического диапазона материала примерно в 1,5 раза, однако в эксперименте отсутствуют предпосылки для такого явления. Также снижение n_1 и n_2 при неизменном Δn может происходить из-за увеличения амплитуд высших гармоник модуляции показателя преломления (n_3 , n_4 и т. д.), т. е. трансформации профиля показателя преломления к несинусоидальной форме и появлению высших порядков дифракции. Такой вывод косвенно подтверждается визуально наблюдаемыми в эксперименте порядками

Таблица 1. Условия записи образца 4
Table 1. Sample 4 recording conditions

Период, нм	Доза, мДж/см ²	Половинный угол сведения
414	20,3	50°34'32"
661	28,0	28°56'4"
826	29,5	22°46'42"
1206	30,9	15°22'38"
1634	31,4	11°17'11"
2100	31,6	8°45'33"

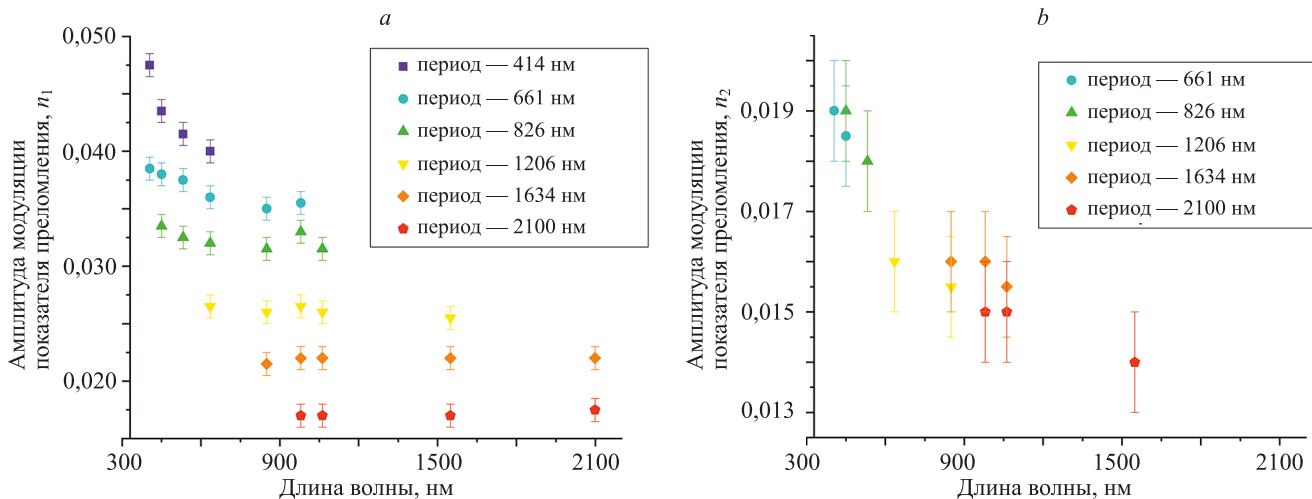


Рис. 3. Спектры амплитуд n_1 (a) и n_2 (b) фотополимера Bayfol HX 200 для голограмм с периодами 414–2100 нм

Fig. 3. Spectra of n_1 (a) and n_2 (b) amplitudes of the Bayfol HX 200 photopolymer for holograms with periods of 414–2100 nm

дифракции вплоть до седьмого (рис. 4) на длине волны восстановления 405 нм для голограммы с периодом 1634 нм (и до 9-го для голограммы с периодом 2100 нм, фото не приведено). Таким образом, можно заключить, что профиль показателя преломления голограмм существенно отличается от синусоиды.

Однозначно проанализировать контуры угловой селективности порядков дифракции выше второго не удалось из-за суперпозиции откликов всех порядков дифракции.

Как видно из результатов эксперимента, модуляция показателя преломления Bayfol HX 200 в ближнем ИК по сравнению с длинноволновой частью видимой области спектра отличается на величину, не превышающую точности измерений. Также наблюдается ярко выраженное нарушение закона взаимозаместимости в зависимости от масштаба интерференционной картины — дифракционная эффективность голограмм уменьшается с увеличением пространственного периода.

Исследование влияния плотности мощности записи на распределение динамического диапазона показателя преломления между гармониками. Выполним анализ соблюдения закона взаимозаместимости при различных скоростях фотополимеризации (которые напрямую связаны с плотностью мощности записывающего излучения [16]). Для проведения анализа распределения динамического диапазона по-

казателя преломления между гармониками запишем 8 голограмм с одинаковыми периодами $d = 1634$ нм, различными плотностями мощности записи ($P_{\text{зап}}$ от 10 до 500 мкВт/см² в одном канале) и одинаковой дозой 20 мДж/см² (условия записи образца 5 приведены в табл. 2). Таким образом, при сохранении неизменной дозы в эксперименте время записи варьируется от нескольких десятков до тысячи секунд.

Для образца 5 измерены контуры угловой селективности на длинах волн 850 и 980 нм и определена дифракционная эффективность в угле Брэгга на — 1549 и 2099 нм. По результатам анализа измеренных дифракционных характеристик голограмм образца 5 построены зависимости n_1 и n_2 (рис. 5). Незначительное снижение модуляции показателя преломления на длине волны 1549 нм, предположительно, связано с низкой степенью коллимации используемого в эксперименте источника.

По полученным графикам видно, что с уменьшением $P_{\text{зап}}$ от 500 до 25 мкВт/см² происходит перераспределение динамического диапазона показателя преломления между n_1 и n_2 . С уменьшением $P_{\text{зап}}$ происходит рост n_1 и подавление n_2 . Исключение из этой тенденции составляет голограмма, записанная при $P_{\text{зап}} = 10$ мкВт/см². Явление может объясняться тем, что при снижении $P_{\text{зап}}$

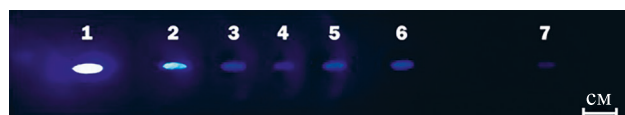


Рис. 4. Фотография семи порядков дифракции голограммы с периодом 1634 нм (длина волны восстановления 405 нм, угол восстановления $60^\circ 10' 11''$, что соответствует углу Брэгга для 7-ой гармоники решетки)

Fig. 4. Photo of seven diffraction orders of a hologram with a period of 1634 nm (reconstruction wavelength was 405 nm, reconstruction angle was $60^\circ 10' 11''$ which corresponds to Bragg angle for the 7th grating harmonic)

Таблица 2. Условия записи образца 5

Table 2. Sample 5 recording conditions

Плотность мощности, мкВт/см ²	Время записи, с
10	1020,0
25	408,0
50	204,0
75	136,0
100	102,0
200	51,0
350	29,0
500	20,5

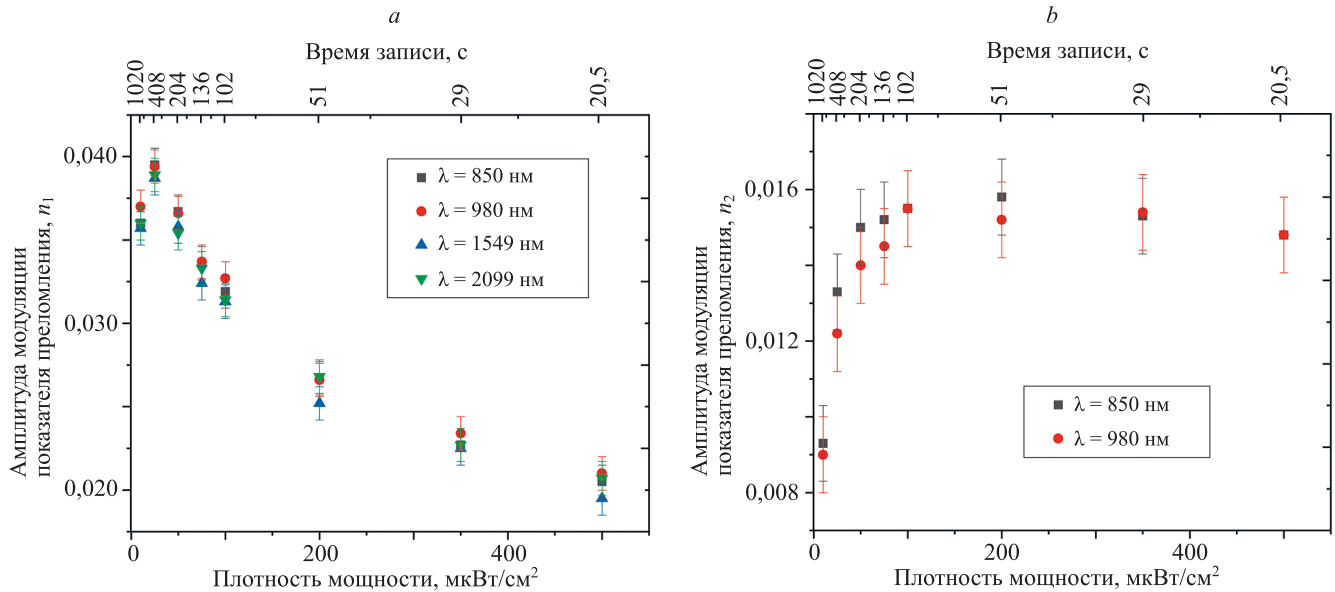


Рис. 5. Зависимости n_1 (a) и n_2 (b) для образца 5 от плотности мощности записи при $d = 1634$ нм

Fig. 5. Dependence of n_1 (a) and n_2 (b) for sample 5 on the recording power density with $d = 1634$ nm

и, соответственно, увеличении времени записи, увеличивается влияние притока кислорода из атмосферы через подложку фотополимера. Постоянно притекающий во время записи через подложку фотополимера кислород окисляет радикализованный фотоинициатор, который не может участвовать в реакции полимеризации, и динамический диапазон фотополимера снижается [16]. При снижении $P_{\text{зап}}$ скорость радикализации фотоинициатора уменьшается, тогда как скорость притока кислорода остается постоянной и из-за увеличения времени записи начинает вносить все больший вклад в уменьшение динамического диапазона показателя преломления фотополимера.

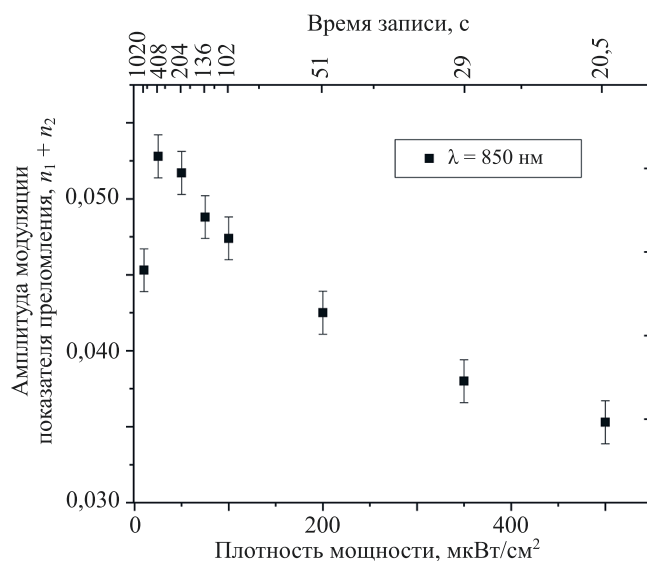


Рис. 6. Зависимость $(n_1 + n_2)$ для образца 5 от плотности мощности записи при $d = 1634$ нм

Fig. 6. Dependence of $(n_1 + n_2)$ for sample 5 on the recording power density with $d = 1634$ nm

Отметим, что сумма n_1 и n_2 для голограмм образца 5, записанных при $P_{\text{зап}}$ 25–500 мкВт/см², уменьшается с ростом $P_{\text{зап}}$ (рис. 6), что косвенно свидетельствует о перераспределении динамического диапазона показателя преломления между низшими и высшими гармониками при изменении $P_{\text{зап}}$.

В эксперименте при восстановлении голограммы с периодом 1634 нм, записанной при $P_{\text{зап}} = 25$ мкВт/см², видны всего пять порядков дифракции (рис. 7, a), тогда как при восстановлении голограммы, записанной при $P_{\text{зап}} = 500$ мкВт/см² — семь порядков (рис. 7, b).

Выполним численный анализ дифракционной эффективности. Для наиболее сильной голограммы с образца 5 ($P_{\text{зап}} = 25$ мкВт/см², $n_1 = 0,039$) дифракционная эффективность на длине волны восстановления 1550 нм

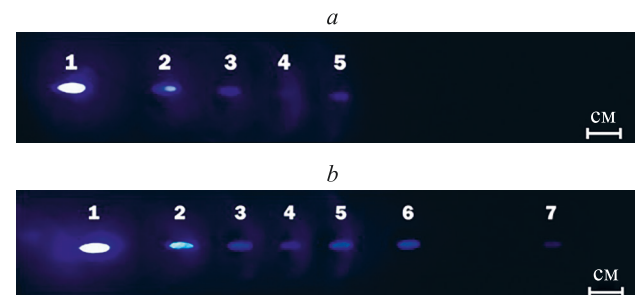


Рис. 7. Фотографии пяти порядков дифракции голограммы с периодом 1634 нм при $P_{\text{зап}} = 25$ мкВт/см² (a) и семи порядков — $P_{\text{зап}} = 500$ мкВт/см² (b). Длина волны восстановления 405 нм, угол восстановления $60^\circ 10' 11''$, что соответствует углу Брэгга для 7-ой гармоники решетки

Fig. 7. Photo of five diffraction orders of a hologram with a period of 1634 nm and a recording power density of 25 $\mu\text{W}/\text{cm}^2$ (a), photo of seven diffraction orders of a hologram with a period of 1634 nm and a recording power density of 500 $\mu\text{W}/\text{cm}^2$ (b), (reconstruction wavelength was 405 nm, reconstruction angle was $60^\circ 10' 11''$, which corresponds to Bragg angle for the 7th grating harmonic)

составила 97 %, а на длине волны 2099 нм — 75 %. Дифракционная эффективность 97 % соответствует потерям на элементе менее 0,135 дБ, что делает Bayfol HX 200 привлекательным материалом для изготовления дифракционной оптики, в том числе для телекоммуникационных применений [24]. Также приведенные цифры свидетельствуют о возможности расширения спектральной области применения дифракционных оптических элементов на базе Bayfol HX 200 до 2100 нм.

Таким образом, при анализе спектра модуляции показателя преломления и исследовании влияния плотности мощности записи на распределение динамического диапазона показателя преломления между гармониками продемонстрировано выраженное нарушение закона взаимозаменяемости как от масштаба интерференционной картины, так и при уменьшении плотности мощности записи.

Исследование влияния условий записи на амплитуду модуляции первой гармоники показателя преломления. Применительно к Bayfol HX, нарушение закона взаимозаменяемости может иметь различные функциональные зависимости в разных диапазонах времени записи вследствие наличия двух конкурирующих процессов, вызванных эффектами массопереноса: фотоиндуцированного разделения фаз [25] и диффузии кислорода [16]. При малом времени записи фотоиндуцированное разделение фаз вносит доминирующий вклад в нарушение закона взаимозаменяемости, что связано с изменением скорости фотополимеризации при изменении плотности мощности излучения [26–28]. Из-за этого происходит перераспределение динамического диапазона показателя преломления между гармониками при увеличении времени записи с сохранением фиксированной дозы. При большом времени записи возрастает влияние диффузии кислорода, который вносит заметный противоположный вклад, вследствие

чего происходит снижение динамического диапазона показателя преломления материала.

Механизм нарушения закона взаимозаменяемости, связанного с фотоиндуцированным разделением фаз, описан моделью полимеризации и линейной диффузии в фотополимерах [26], применимой для Bayfol HX [29]. В явном виде в модели введен параметр R , связывающий материальные константы (скорости фотополимеризации k и диффузии D), и условия записи (среднюю интенсивность записывающего излучения в среде I и пространственную частоту решетки $K = \frac{2\pi}{d}$):

$$R = \frac{K^2 D}{I k}. \quad (2)$$

Значение параметра R определяет амплитуды гармоник модуляции показателя преломления голограммы. Таким образом, сопоставление амплитуды модуляции показателя преломления и величины R может рассматриваться как критерий выбора оптимального соотношения условий записи.

Для анализа описанных нарушений в дополнение к образцам 4 и 5 записаны несколько серий решеток (образцы 6–9) с разными периодами (414, 470 и 1206 нм), неизменной дозой ($E = 20$ мДж/см²), разным временем записи (от 20 до 1365 с) и плотностью мощности записывающего излучения (от 10 до 500 мВт/см²) (табл. 3). А также одна решетка (образец 10) с периодом $d = 661$ нм, дозой $E = 20$ мДж/см², временем записи $t = 457$ с и плотностью мощности записывающего излучения $P_{\text{зап}} = 25$ мВт/см².

Экспериментальные результаты обрабатываем с использованием формулы (2). При этом в настоящей работе эквивалентом средней интенсивности в среде является удвоенная плотность мощности каждо-

Таблица 3. Условия записи образцов 6–9

Table 3. Samples 6–9 recording conditions

Плотность мощности, мВт/см ²	Время записи, с	Плотность мощности, мВт/см ²	Время записи, с
Образец 6: $d = 414$ нм, $E = 20$ мДж/см ²		Образец 7: $d = 1206$ нм, $E = 20$ мДж/см ²	
25	630,0	10	1037,0
50	315,0	25	415,0
75	210,0	50	207,0
100	158,0	75	138,0
200	79,0	100	104,0
350	45,0	200	52,0
500	31,5	350	30,0
—	—	500	20,5
Образец 8: $d = 470$ нм, $E = 20$ мДж/см ²		Образец 9: $d = 470$ нм, $E = 20$ мДж/см ²	
10	1365,0	25	546,0
15	910,0	50	273,0
25	546,0	100	137,0
35	390,0	350	39,0
50	273,0	—	—

го из записывающих пучков умноженная на косинус угла Брэгга, как следствие, I заменим на $2P_{\text{зап}} \cos \theta$. Поскольку значения скоростей полимеризации и диффузии — коммерческая тайна (k и D неизвестны), дальнейшие рассуждения, связанные с параметром R , проводятся с точностью до материальных констант:

$$Rk/D = \frac{K^2}{2P_{\text{зап}} \cos \theta}. \quad (3)$$

Посчитанные значения n_1 для всех голограмм образцов 4–10 сопоставим с условиями записи (рис. 8). При этом, для того чтобы отделить нарушение закона взаимозаместимости, связанное с диффузией кислорода, голограммы с характерным временем записи более 408 с обведены красным овалом, как вероятно подверженные влиянию кислорода. Основанием для такого разделения стали зависимости, полученные для образца 5 на рис. 5 и 6, из которых следует, что голограмма с характерным временем записи более 408 с подвержена влиянию диффузии кислорода.

Из рис. 8 видна зависимость n_1 от условий записи, анализ которой важен для достижения максимальных значений n_1 . Наблюдается закономерность возрастания n_1 при соблюдении условий записи, обеспечивающих более медленный процесс формирования голограммы. Такой характер зависимости прослеживается для голограмм, записанных при значениях $Rk/D < 0,8$ пВт⁻¹, максимальные значения n_1 достигаются в диапазоне значений Rk/D от 0,8 до 1,2 пВт⁻¹. Однако из эксперимента остается неясным характер зависимости при $Rk/D > 1,2$ пВт⁻¹ из-за существенного влияния диффузии кислорода на динамический диапазон фотополимера при записи голограмм с периодами более 400 нм. Запись голограмм с меньшими периодами в используемом интерферометре невозможна в связи с его габаритными ограничениями. В связи с этим для уточнения значений n_1 при $Rk/D > 1,2$ пВт⁻¹ возникает необходимость записи голограмм с ограничением притока кислорода в фотополимер.

Для окончательного отделения влияния диффузии кислорода на величину n_1 запишем образцы 11–14 (параметры голограмм и условия записи приведены в табл. 4). Во время записи образцов 11–14 ограничим приток кислорода в фотополимер. Ограничение реализовано за счет иммерсирования непроницаемой для кислорода, но прозрачной для света, стеклянной пластины (пропускание более 99 %, без учета потерь на отражение) поверх подложки фотополимера на время записи голограммы. Схематично условия записи голограмм показаны на рис. 9.

Выполним обработку полученных результатов в соответствии с моделью полимеризации и линейной диффузии в фотополимерах при помощи формулы (3). Значения n_1 для голограмм образцов 11–14 сопоставлены с условиями записи на рис. 10, а. Также для наглядного сравнения влияния кислорода продублировано распределение значений n_1 для голограмм образцов 4–10 (рис. 10, б)

На приведенной зависимости виден рост n_1 при увеличении Rk/D , однако максимальное значение n_1

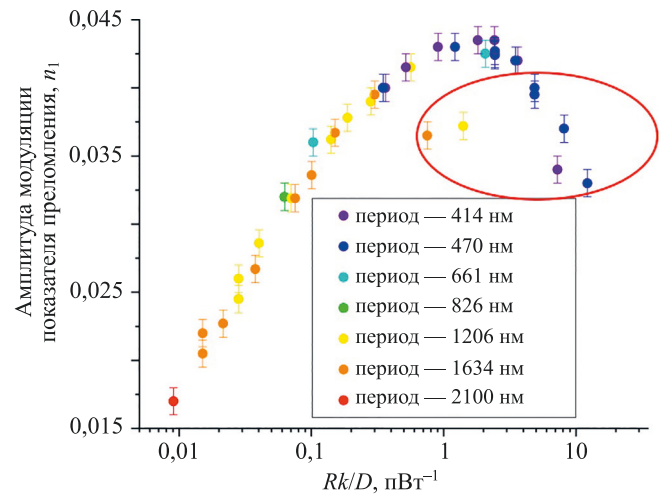


Рис. 8. Распределение n_1 голограмм от условий записи (частоты решетки и плотности мощности излучения), приток кислорода не ограничен

Fig. 8. The distribution of n_1 of the holograms on the recording conditions (grating frequency and recording power density), the oxygen flow is not limited

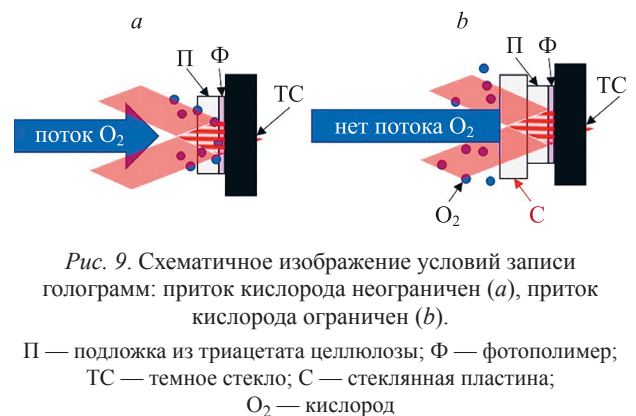


Рис. 9. Схематичное изображение условий записи голограмм: приток кислорода неограничен (а), приток кислорода ограничен (б).

П — подложка из триацетата целлюлозы; Ф — фотополимер; TC — темное стекло; C — стеклянная пластина; O₂ — кислород

Fig. 9. Schematic representation of the conditions for hologram recording: with oxygen flow (a), without oxygen flow (b).

П — cellulose triacetate substrate; Ф — photopolymer; TC — dark glass; C — glass plate; O₂ — oxygen

в случае ограничения притока кислорода достигается раньше, уже при $Rk/D \approx 0,5$ пВт⁻¹. Заметим, что в случае записи голограмм с ограничением притока кислорода (рис. 10, а), в отличие от случая записи голограмм без ограничения (рис. 10, б), наблюдается выход на насыщение $n_1 \approx 0,043$. Также отметим, что при ограничении притока кислорода наблюдается нестабильность характеристик получаемых голограмм (некоторые голограммы имеют очень малые значения n_1 , выбивающиеся из зависимости), что, предположительно, связано с неотработанным процессом иммерсирования стеклянной пластины, ограничивающей приток кислорода.

Экспериментально продемонстрировано влияние фотоиндуцированного разделения фаз на нарушение закона взаимозаместимости для Bayfol HX 200. Благодаря анализу нарушения закона взаимозаместимости для исследуемого фотополимера выявлены условия записи, при которых достигается наибольшая

Таблица 4. Условия записи образцов 11–14

Table 4. Samples 11–14 recording conditions

Плотность мощности, мВт/см ²	Время записи, с	Плотность мощности, мВт/см ²	Время записи, с
Образец 11: $d = 1206$ нм, $E = 20$ мДж/см ²		Образец 12: $d = 470$ нм, $E = 20$ мДж/см ²	
10	1037,0	10	1365,0
25	415,0	15	910,0
50	207,0	25	546,0
75	138,0	35	390,0
100	104,0	50	273,0
200	52,0	25	546,0
350	30,0	—	—
500	20,5	—	—
Образец 13: $d = 1634$ нм, $E = 20$ мДж/см ²		Образец 14: $d = 414$ нм, $E = 20$ мДж/см ²	
10	1020,0	25	630,0
25	408,0	75	210,0
50	204,0	200	79,0

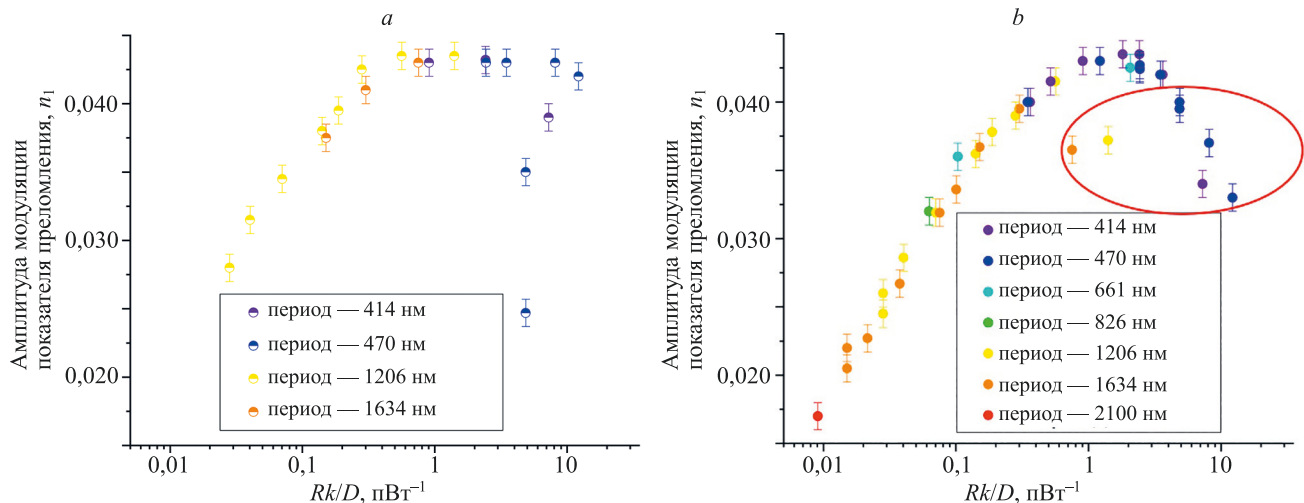


Рис. 10. Распределение n_1 голограмм от условий записи (частоты решетки и плотности мощности излучения), приток кислорода ограничен (а), приток кислорода не ограничен (для сравнения приведен рис. 8) (b)

Fig. 10. The distribution of n_1 of the holograms on the recording conditions (grating frequency and recording power density), the oxygen flow is limited (a), the oxygen flow is not limited (b)

модуляция первой гармоники показателя преломления. Показано, что в случае голографической записи без ограничения притока кислорода время записи должно быть менее 408 с, а величина Rk/D должна выбираться больше 0,8 пВт⁻¹. В случае ограничения притока кислорода величина Rk/D должна выбираться больше 0,5 пВт⁻¹.

Заключение

В работе исследован потенциал использования голографического фотополимера Bayfol HX 200 в диапазоне длин волн восстановления до 2100 нм. Показано, что модуляция показателя преломления в исследуемом фотополимере в ближнем инфракрасном диапазоне по сравнению с длинноволновой частью видимого диапа-

зона спектра отличается на величину, не превышающую точности измерений.

Впервые проведен численный анализ нарушения закона взаимозаместимости для фотополимера Bayfol HX 200. С практической точки зрения анализ нарушения закона взаимозаместимости в настоящей работе позволил выявить оптимальные условия для записи голограмм с наиболее сильным первым порядком дифракции для широкого диапазона периодов (414–1634 нм). Оптимальные условия записи найдены для двух случаев: голографическая запись происходит с постоянным притоком кислорода в фотополимер и без притока. Так, для случая, когда кислород поступает в фотополимер беспрепятственно, найдено ограничение по времени записи. Для случая записи с ограничением притока кислорода — оптимальные условия определяются соотно-

шением пространственной частоты решетки и средней интенсивности записывающего излучения.

Отметим, что дифракционная эффективность голограмм, записанных в Bayfol HX 200 толщиной 16 мкм, составила 97 % на длине волны восстановления 1550 нм, что делает его привлекательным материалом для изготовления дифракционной оптики, в том числе

для телекоммуникационных применений. Также дифракционная эффективность записанных голограмм достигала 75 % на длине волны восстановления 2099 нм, что также позволяет эффективно применять Bayfol HX 200 для создания дифракционной оптики для спектрального диапазона до 2100 нм.

Литература

1. Quintana J.A., Boj P.G., Crespo J., Pardo M., Satorre M.A. Line-focusing holographic mirrors for solar ultraviolet energy concentration // *Applied Optics*. 1997. V. 36. N 16. P. 3689–3693. <https://doi.org/10.1364/AO.36.003689>
2. Glebov L.B., Smirnov V., Rotari E., Cohanoschi I., Glebova L., Smolski O.V., Lumeau J., Lantigua C., Glebov A. Volume-chirped Bragg gratings: monolithic components for stretching and compression of ultrashort laser pulses // *Optical Engineering*. 2014. V. 53. N 5. P. 051514. <https://doi.org/10.1117/1.OE.53.5.051514>
3. Berneth H., Bruder F.-K., Fäcke T., Hagen R., Hönel D., Jürbergs D., Rölle T., Weiser M.-S. Holographic recording aspects of high-resolution Bayfol HX photopolymer // *Proceedings of SPIE*. 2011. V. 7957. P. 79570H. <https://doi.org/10.1117/12.876202>
4. Bruder F.-K., Hansen S., Kleinschmidt T., Künzel R., Manecke C., Orselli E., Rewitz C., Rölle T. Integration of volume holographic optical elements (vHOE) made with Bayfol® HX into plastic optical parts // *Proceedings of SPIE*. 2019. V. 10944. P. 1094402. <https://doi.org/10.1117/12.2510109>
5. Vázquez-Martín I., Gómez-Climente M., Marín-Sáez J., Collados M.V., Atencia J. True colour Denisyuk-type hologram recording in Bayfol HX self-developing photopolymer // *Proceedings of SPIE*. 2017. V. 10233. P. 102331U. <https://doi.org/10.1117/12.2265802>
6. Marín-Sáez J., Atencia J., Chemisana D., Collados M.V. Full modeling and experimental validation of cylindrical holographic lenses recorded in Bayfol HX photopolymer and partly operating in the transition regime for solar concentration // *Optics Express*. 2018. V. 26. N 10. P. A398–A412. <https://doi.org/10.1364/OE.26.00A398>
7. Bruder F.-K., Fäcke T., Grote F., Hagen R., Hönel D., Koch E., Rewitz C., Walze G., Wewer B. Performance optimization in mass production of volume holographic optical elements (vHOEs) using Bayfol HX photopolymer film // *Proceedings of SPIE*. 2017. V. 10233. P. 102330G. <https://doi.org/10.1117/12.2265022>
8. Berneth H., Bruder F.-K., Fäcke T., Jürbergs D., Hagen R., Hönel D., Rölle T., Walze G. Bayfol HX photopolymer for full-color transmission volume Bragg gratings // *Proceedings of SPIE*. 2014. V. 9006. P. 900602. <https://doi.org/10.1117/12.2038399>
9. Bruder F.-K., Fäcke T., Hagen R., Hönel D., Orselli E., Rewitz C., Rölle T., Walze G. Diffractive optics with high Bragg selectivity: volume holographic optical elements in Bayfol® HX photopolymer film // *Proceedings of SPIE*. 2015. V. 9626. P. 96260T. <https://doi.org/10.1117/12.2191587>
10. Bruder F.-K., Bang H., Fäcke T., Hagen R., Hönel D., Orselli E., Rewitz C., Rölle T., Vukicevic D., Walze G. Precision holographic optical elements in Bayfol HX photopolymer // *Proceedings of SPIE*. 2016. V. 9771. P. 977103. <https://doi.org/10.1117/12.2209636>
11. Bruder F.-K., Fäcke T., Hagen R., Hönel D., Kleinschmidt T.P., Orselli E., Rewitz C., Rölle T., Walze G. Diffractive optics in large sizes: computer-generated holograms (CGH) based on Bayfol HX photopolymer // *Proceedings of SPIE*. 2015. V. 9385. P. 93850C. <https://doi.org/10.1117/12.2077139>
12. Marín-Sáez J., Atencia J., Chemisana D., Collados M.V. Characterization of volume holographic optical elements recorded in Bayfol HX photopolymer for solar photovoltaic applications // *Optics Express*. 2016. V. 24. N 6. P. A720–A730. <https://doi.org/10.1364/OE.24.00A720>
13. Bruder F.-K., Frank J., Hansen S., Lorenz A., Manecke C., Meisenheimer R., Mills J., Pitzer L., Pochorovski I., Rölle T. Expanding the property profile of Bayfol HX films towards NIR recording and ultra-high index modulation // *Proceedings of SPIE*. 2021. V. 11765. P. 117650J. <https://doi.org/10.1117/12.2579235>

References

1. Quintana J.A., Boj P.G., Crespo J., Pardo M., Satorre M.A. Line-focusing holographic mirrors for solar ultraviolet energy concentration. *Applied Optics*, 1997, vol. 36, no. 16, pp. 3689–3693. <https://doi.org/10.1364/AO.36.003689>
2. Glebov L.B., Smirnov V., Rotari E., Cohanoschi I., Glebova L., Smolski O.V., Lumeau J., Lantigua C., Glebov A. Volume-chirped Bragg gratings: monolithic components for stretching and compression of ultrashort laser pulses. *Optical Engineering*, 2014, vol. 53, no. 5, pp. 051514. <https://doi.org/10.1117/1.OE.53.5.051514>
3. Berneth H., Bruder F.-K., Fäcke T., Hagen R., Hönel D., Jürbergs D., Rölle T., Weiser M.-S. Holographic recording aspects of high-resolution Bayfol HX photopolymer. *Proceedings of SPIE*, 2011, vol. 7957, pp. 79570H. <https://doi.org/10.1117/12.876202>
4. Bruder F.-K., Hansen S., Kleinschmidt T., Künzel R., Manecke C., Orselli E., Rewitz C., Rölle T. Integration of volume holographic optical elements (vHOE) made with Bayfol® HX into plastic optical parts. *Proceedings of SPIE*, 2019, vol. 10944, pp. 1094402. <https://doi.org/10.1117/12.2510109>
5. Vázquez-Martín I., Gómez-Climente M., Marín-Sáez J., Collados M.V., Atencia J. True colour Denisyuk-type hologram recording in Bayfol HX self-developing photopolymer. *Proceedings of SPIE*, 2017, vol. 10233, pp. 102331U. <https://doi.org/10.1117/12.2265802>
6. Marín-Sáez J., Atencia J., Chemisana D., Collados M.V. Full modeling and experimental validation of cylindrical holographic lenses recorded in Bayfol HX photopolymer and partly operating in the transition regime for solar concentration. *Optics Express*, 2018, vol. 26, no. 10, pp. A398–A412. <https://doi.org/10.1364/OE.26.00A398>
7. Bruder F.-K., Fäcke T., Grote F., Hagen R., Hönel D., Koch E., Rewitz C., Walze G., Wewer B. Performance optimization in mass production of volume holographic optical elements (vHOEs) using Bayfol HX photopolymer film. *Proceedings of SPIE*, 2017, vol. 10233, pp. 102330G. <https://doi.org/10.1117/12.2265022>
8. Berneth H., Bruder F.-K., Fäcke T., Jürbergs D., Hagen R., Hönel D., Rölle T., Walze G. Bayfol HX photopolymer for full-color transmission volume Bragg gratings. *Proceedings of SPIE*, 2014, vol. 9006, pp. 900602. <https://doi.org/10.1117/12.2038399>
9. Bruder F.-K., Fäcke T., Hagen R., Hönel D., Orselli E., Rewitz C., Rölle T., Walze G. Diffractive optics with high Bragg selectivity: volume holographic optical elements in Bayfol® HX photopolymer film. *Proceedings of SPIE*, 2015, vol. 9626, pp. 96260T. <https://doi.org/10.1117/12.2191587>
10. Bruder F.-K., Bang H., Fäcke T., Hagen R., Hönel D., Orselli E., Rewitz C., Rölle T., Vukicevic D., Walze G. Precision holographic optical elements in Bayfol HX photopolymer. *Proceedings of SPIE*, 2016, vol. 9771, pp. 977103. <https://doi.org/10.1117/12.2209636>
11. Bruder F.-K., Fäcke T., Hagen R., Hönel D., Kleinschmidt T.P., Orselli E., Rewitz C., Rölle T., Walze G. Diffractive optics in large sizes: computer-generated holograms (CGH) based on Bayfol HX photopolymer. *Proceedings of SPIE*, 2015, vol. 9385, pp. 93850C. <https://doi.org/10.1117/12.2077139>
12. Marín-Sáez J., Atencia J., Chemisana D., Collados M.V. Characterization of volume holographic optical elements recorded in Bayfol HX photopolymer for solar photovoltaic applications. *Optics Express*, 2016, vol. 24, no. 6, pp. A720–A730. <https://doi.org/10.1364/OE.24.00A720>
13. Bruder F.-K., Frank J., Hansen S., Lorenz A., Manecke C., Meisenheimer R., Mills J., Pitzer L., Pochorovski I., Rölle T. Expanding the property profile of Bayfol HX films towards NIR recording and ultra-high index modulation. *Proceedings of SPIE*, 2021, vol. 11765, pp. 117650J. <https://doi.org/10.1117/12.2579235>

14. Peng H., Nair D.P., Kowalski B.A., Xi W., Gong T., Wang C., Cole M., Cramer N.B., Xie X., McLeod R.R., Bowman C.N. High performance graded rainbow holograms via two-stage sequential orthogonal thiol-click chemistry // *Macromolecules*. 2014. V. 47. N 7. P. 2306–2315. <https://doi.org/10.1021/ma500167x>
15. Monte F.D., Martínez O., Rodrigo J. A., Calvo M. L., Cheben P. A volume holographic sol-gel material with large enhancement of dynamic range by incorporation of high refractive index species // *Advanced Materials*. 2006. V. 18. N 15. P. 2014–2017. <https://doi.org/10.1002/adma.200502675>
16. Bruder F.-K., Fäcke T., Rölle T. The chemistry and physics of Bayfol® HX film holographic photopolymer // *Polymers*. 2017. V. 9. N 10. P. 472. <https://doi.org/10.3390/polym9100472>
17. Kargaran A., Ebrahimi M., Riaz M., Hosseiny A., Jafari G.R. Quartic balance theory: Global minimum with imbalanced triangles // *Physical Review*. 2020. V. 102. N 1. P. 012310. <https://doi.org/10.1103/PhysRevE.102.012310>
18. Spiegler K.S., Kedem O. Thermodynamics of hyperfiltration (reverse osmosis): criteria for efficient membranes // *Desalination*. 1966. V. 1. N 4. P. 311–326. [https://doi.org/10.1016/S0011-9164\(00\)80018-1](https://doi.org/10.1016/S0011-9164(00)80018-1)
19. Lucarini V., Saarinen J.J., Peiponen K.-E., Vartiainen E.M. *Kramers-Kronig Relations in Optical Materials Research*. Springer Science & Business Media, 2005. 162 p. (Springer Series in Optical Sciences. V. 110). <https://doi.org/10.1007/b138913>
20. Kopyeva M.S., Filatova S.A., Kamynin V.A., Trikshev A.I., Kozlikina E.I., Astashov V.V., Loschenov V.B., Tsvetkov V.B. Ex-vivo exposure on biological tissues in the 2-μm spectral range with an all-fiber continuous-wave holmium laser // *Photonics*. 2021. V. 9. N 1. P. 20. <https://doi.org/10.3390/photonics9010020>
21. Kogelnik H. Coupled wave theory for thick hologram gratings // *Bell System Technical Journal*. 1969. V. 48. N 9. P. 2909–2947. <https://doi.org/10.1002/j.1538-7305.1969.tb01198.x>
22. Borisov V.N., Angervaks A.E., Ryskin A.I., Veniaminov A.V. Two-model spectral study of volume holograms in materials with diffusion-based mechanisms // *Optical Engineering*. 2019. V. 58. N 2. P. 024102. <https://doi.org/10.1117/1.OE.58.2.024102>
23. Mees E.K., James T.H. *The Theory of the Photographic Process*. New York: Macmillan, 1967. Chap. 7. P. 132.
24. Pottier P., Strain M.J., Packirisamy M. Integrated microspectrometer with elliptical Bragg mirror enhanced diffraction grating on silicon on insulator // *ACS Photonics*. 2014. V. 1. N 5. P. 430–436. <https://doi.org/10.1021/ph400165j>
25. Liu P., Zhao Y., Li Z., Sun X. Improvement of ultrafast holographic performance in silver nanoprisms dispersed photopolymer // *Optics Express*. 2018. V. 26. N 6. P. 6993–7004. <https://doi.org/10.1364/OE.26.006993>
26. Sheridan J.T., Lawrence J.R. Nonlocal-response diffusion model of holographic recording in photopolymer // *Journal of the Optical Society of America A*. 2000. V. 17. N 6. P. 1108–1114. <https://doi.org/10.1364/JOSAA.17.001108>
27. Zhao G., Mouroulis P. Diffusion model of hologram formation in dry photopolymer materials // *Journal of Modern Optics*. 1994. V. 41. N 10. P. 1929–1939. <https://doi.org/10.1080/09500349414551831>
28. Kelly J.V., O'Neill F.T., Sheridan J.T., Neipp C., Gallego S., Ortuno M. Holographic photopolymer materials with nonlocal and nonlinear response // *Proceedings of SPIE*. 2003. V. 5216. P. 127–138. <https://doi.org/10.1117/12.509138>
29. Bruder F.-K., Deuber F., Fäcke T., Hagen R., Hönel D., Jurbergs D., Rölle T., Weiser M.-S. Reaction-diffusion model applied to high resolution Bayfol HX photopolymer // *Proceedings of SPIE*. 2010. V. 7619. P. 76190I. <https://doi.org/10.1117/12.841956>
14. Peng H., Nair D.P., Kowalski B.A., Xi W., Gong T., Wang C., Cole M., Cramer N.B., Xie X., McLeod R.R., Bowman C.N. High performance graded rainbow holograms via two-stage sequential orthogonal thiol-click chemistry. *Macromolecules*, 2014, vol. 47, no. 7, pp. 2306–2315. <https://doi.org/10.1021/ma500167x>
15. Monte F.D., Martínez O., Rodrigo J. A., Calvo M. L., Cheben P. A volume holographic sol-gel material with large enhancement of dynamic range by incorporation of high refractive index species. *Advanced Materials*, 2006, vol. 18, no. 15, pp. 2014–2017. <https://doi.org/10.1002/adma.200502675>
16. Bruder F.-K., Fäcke T., Rölle T. The chemistry and physics of Bayfol® HX film holographic photopolymer. *Polymers*, 2017, vol. 9, no. 10, pp. 472. <https://doi.org/10.3390/polym9100472>
17. Kargaran A., Ebrahimi M., Riaz M., Hosseiny A., Jafari G.R. Quartic balance theory: Global minimum with imbalanced triangles. *Physical Review*, 2020, vol. 102, no. 1, pp. 012310. <https://doi.org/10.1103/PhysRevE.102.012310>
18. Spiegler K.S., Kedem O. Thermodynamics of hyperfiltration (reverse osmosis): criteria for efficient membranes. *Desalination*, 1966, vol. 1, no. 4, pp. 311–326. [https://doi.org/10.1016/S0011-9164\(00\)80018-1](https://doi.org/10.1016/S0011-9164(00)80018-1)
19. Lucarini V., Saarinen J.J., Peiponen K.-E., Vartiainen E.M. *Kramers-Kronig Relations in Optical Materials Research*. Springer Science & Business Media, 2005. 162 p. Springer Series in Optical Sciences, vol. 110. <https://doi.org/10.1007/b138913>
20. Kopyeva M.S., Filatova S.A., Kamynin V.A., Trikshev A.I., Kozlikina E.I., Astashov V.V., Loschenov V.B., Tsvetkov V.B. Ex-vivo exposure on biological tissues in the 2-μm spectral range with an all-fiber continuous-wave holmium laser. *Photonics*, 2021, vol. 9, no. 1, pp. 20. <https://doi.org/10.3390/photonics9010020>
21. Kogelnik H. Coupled wave theory for thick hologram gratings. *Bell System Technical Journal*, 1969, vol. 48, no. 9, pp. 2909–2947. <https://doi.org/10.1002/j.1538-7305.1969.tb01198.x>
22. Borisov V.N., Angervaks A.E., Ryskin A.I., Veniaminov A.V. Two-model spectral study of volume holograms in materials with diffusion-based mechanisms. *Optical Engineering*, 2019, vol. 58, no. 2, pp. 024102. <https://doi.org/10.1117/1.OE.58.2.024102>
23. Mees E.K., James T.H. *The Theory of the Photographic Process*. New York, Macmillan, 1967, chap. 7, pp. 132.
24. Pottier P., Strain M.J., Packirisamy M. Integrated microspectrometer with elliptical Bragg mirror enhanced diffraction grating on silicon on insulator. *ACS Photonics*, 2014, vol. 1, no. 5, pp. 430–436. <https://doi.org/10.1021/ph400165j>
25. Liu P., Zhao Y., Li Z., Sun X. Improvement of ultrafast holographic performance in silver nanoprisms dispersed photopolymer. *Optics Express*, 2018, vol. 26, no. 6, pp. 6993–7004. <https://doi.org/10.1364/OE.26.006993>
26. Sheridan J.T., Lawrence J.R. Nonlocal-response diffusion model of holographic recording in photopolymer. *Journal of the Optical Society of America A*, 2000, vol. 17, no. 6, pp. 1108–1114. <https://doi.org/10.1364/JOSAA.17.001108>
27. Zhao G., Mouroulis P. Diffusion model of hologram formation in dry photopolymer materials. *Journal of Modern Optics*, 1994, vol. 41, no. 10, pp. 1929–1939. <https://doi.org/10.1080/09500349414551831>
28. Kelly J.V., O'Neill F.T., Sheridan J.T., Neipp C., Gallego S., Ortuno M. Holographic photopolymer materials with nonlocal and nonlinear response. *Proceedings of SPIE*, 2003, vol. 5216, pp. 127–138. <https://doi.org/10.1117/12.509138>
29. Bruder F.-K., Deuber F., Fäcke T., Hagen R., Hönel D., Jurbergs D., Rölle T., Weiser M.-S. Reaction-diffusion model applied to high resolution Bayfol HX photopolymer. *Proceedings of SPIE*, 2010, vol. 7619, pp. 76190I. <https://doi.org/10.1117/12.841956>

Авторы

Борисов Владимир Николаевич — кандидат физико-математических наук, независимый исследователь, Санкт-Петербург, Российская Федерация, orcid.org/0000-0003-3525-5698, pvlc2013@gmail.com

Зверев Андрей Дмитриевич — и.о. младший научный сотрудник, Институт общей физики им. А.М. Прохорова Российской академии наук, Москва, 119991, Российская Федерация, <https://orcid.org/0000-0002-3383-906X>, izverevad@gmail.com

Authors

Vladimir N. Borisov — PhD (Physics & Mathematics), Independent Researcher, Saint Petersburg, Russian Federation, orcid.org/0000-0003-3525-5698, pvlc2013@gmail.com

Andrey D. Zverev — Acting Junior Researcher, Prokhorov General Physics Institute of the Russian Academy of Sciences, Moscow, 119991, Russian Federation, <https://orcid.org/0000-0002-3383-906X>, izverevad@gmail.com

Камынин Владимир Александрович — кандидат физико-математических наук, старший научный сотрудник, Институт общей физики им. А.М. Прохорова Российской академии наук, Москва, 119991, Российская Федерация, <https://orcid.org/0000-0002-4265-3478>, kamyninva@gmail.com

Копьева Мария Сергеевна — и.о. младший научный сотрудник, Институт общей физики им. А.М. Прохорова Российской академии наук, Москва, 119991, Российская Федерация; аспирант, Российский университет дружбы народов, Москва, 117198, Российская Федерация, <https://orcid.org/0000-0002-2113-6607>, mashutka_koryova@mail.ru

Окунь Роман Александрович — и.о. младший научный сотрудник, Институт общей физики им. А.М. Прохорова Российской академии наук, Москва, 119991, Российская Федерация, <https://orcid.org/0000-0003-2461-9978>, romaokun@rambler.ru

Цветков Владимир Борисович — доктор физико-математических наук, заместитель директора, Институт общей физики им. А.М. Прохорова Российской академии наук, Москва, 119991, Российская Федерация, [sc 57218466660, https://orcid.org/0000-0002-1483-3308](https://orcid.org/0000-0002-1483-3308), tsvetkov@lsk.gpi.ru

Vladimir A. Kamynin — PhD (Physics & Mathematics), Senior Researcher, Prokhorov General Physics Institute of the Russian Academy of Sciences, Moscow, 119991, Russian Federation, <https://orcid.org/0000-0002-4265-3478>, kamyninva@gmail.com

Maria S. Kopyeva — Acting Junior Researcher, Prokhorov General Physics Institute of the Russian Academy of Sciences, Moscow, 119991, Russian Federation; PhD Student, RUDN University, Moscow, 117198, Russian Federation, <https://orcid.org/0000-0002-2113-6607>, mashutka_kopyova@mail.ru

Roman A. Okun — Acting Junior Researcher, Prokhorov General Physics Institute of the Russian Academy of Sciences, Moscow, 119991, Russian Federation, <https://orcid.org/0000-0003-2461-9978>, romaokun@rambler.ru

Vladimir B. Tsvetkov — D.Sc., Deputy Director, Prokhorov General Physics Institute of the Russian Academy of Sciences, Moscow, 119991, Russian Federation, [sc 57218466660, https://orcid.org/0000-0002-1483-3308](https://orcid.org/0000-0002-1483-3308), tsvetkov@lsk.gpi.ru

*Статья поступила в редакцию 15.07.2022
Одобрена после рецензирования 10.11.2022
Принята к печати 29.01.2023*

*Received 15.07.2022
Approved after reviewing 10.11.2022
Accepted 29.01.2023*



Работа доступна по лицензии
Creative Commons
«Attribution-NonCommercial»

doi: 10.17586/2226-1494-2023-23-1-14-20

УДК 535.8

Исследование реакции кровеносных сосудов на локальный нагрев методом визуализирующей фотоплетизмографии

Анжелика Вадимовна Белавенцева¹✉, Наталья Петровна Подолян²,
Максим Александрович Волинский³, Валерий Викторович Зайцев⁴,
Анастасия Владимировна Саковская⁵, Олег Викторович Мамонтов⁶,
Роман Владимирович Ромашко⁷, Алексей Александрович Камшилин⁸

^{1,2,3,4,5,6,7,8} Институт автоматики и процессов управления Дальневосточного отделения Российской академии наук, Владивосток, 690041, Российская Федерация

³ Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация

^{4,6} Национальный медицинский исследовательский центр им. В.А. Алмазова Минздрава России, Санкт-Петербург, 197341, Российская Федерация

⁵ Тихоокеанский государственный медицинский университет, Владивосток, 690002, Российская Федерация

⁶ Первый Санкт-Петербургский государственный медицинский университет имени Павлова, Санкт-Петербург, 197022, Российская Федерация

¹ ange202@mail.ru✉, <https://orcid.org/0000-0001-8218-6766>

² podolian@iacp.dvo.ru, <https://orcid.org/0000-0002-2108-4376>

³ maxim.volynsky@gmail.com, <https://orcid.org/0000-0003-1679-4450>

⁴ zaytsevphoto@gmail.com, <https://orcid.org/0000-0003-3366-8376>

⁵ sakovska86@mail.ru, <https://orcid.org/0000-0002-4077-9739>

⁶ mamontoffoleg@gmail.com, <https://orcid.org/0000-0003-3063-439X>

⁷ romashko@dvo.ru, <https://orcid.org/0000-0003-0869-0993>

⁸ alexei.kamshilin@yandex.ru, <https://orcid.org/0000-0003-2506-9736>

Аннотация

Предмет исследования. Исследована возможность применения нового бесконтактного метода визуализирующей фотоплетизмографии для исследования терморегуляторной вазодилатации сосудов. **Метод.** Реакция перфузии в области внешнего предплечья в ответ на локальный нагрев до 41 ± 1 °C была оценена у шести добровольцев в возрасте 39–52 лет с помощью видеозаписи исследуемой области, синхронизированной с электрокардиограммой, и последующей корреляционной обработкой полученных данных. **Основные результаты.** Показано, что изменение перфузии при локальном нагреве носит двухфазный характер и обусловлено откликом нервной системы, опосредованным аксон-рефлексом в первой фазе вазодилатации, и синтезом оксида азота в клетках эндотелия во второй фазе. Установлено, что многократный рост перфузии в первой фазе нагрева зависит как от начальной температуры кожи, так и от разницы температуры ее нагрева. Обнаружено, что для существенного развития сосудистой реакции на гипертермию, связанную с активацией эндотелиальной функции, необходим нагрев тканей длительностью более 15 мин. **Практическая значимость.** Подтверждено, что метод визуализирующей фотоплетизмографии достоверно отражает работу механизмов регуляции периферического сосудистого сопротивления, что имеет большое прогностическое значение для выявления первичных признаков сердечно-сосудистых заболеваний.

Ключевые слова

визуализирующая фотоплетизмография, микроциркуляция, реактивность сосудов, перфузия, терморегуляция

Благодарности

Исследование реализовано при поддержке гранта № 21-15-00265 Российского научного фонда в части разработки измерительной системы, включая программное обеспечение, и проведения экспериментов. Обработка и хранение экспериментальных данных осуществлялись с использованием оборудования Центра коллективного пользования «Дальневосточный вычислительный ресурс» Института автоматики и процессов управления Дальневосточного отделения Российской академии наук.

© Белавенцева А.В., Подолян Н.П., Волинский М.А., Зайцев В.В., Саковская А.В., Мамонтов О.В., Ромашко Р.В., Камшилин А.А., 2023

Ссылка для цитирования: Белавенцева А.В., Подолян Н.П., Волынский М.А., Зайцев В.В., Саковская А.В., Мамонтов О.В., Ромашко Р.В., Камшилин А.А. Исследование реакции кровеносных сосудов на локальный нагрев методом визуализирующей фотоплетизмографии // Научно-технический вестник информационных технологий, механики и оптики. 2023. Т. 23, № 1. С. 14–20. doi: 10.17586/2226-1494-2023-23-1-14-20

Study of blood vessels reaction to local heating by imaging photoplethysmography

Anzhelika V. Belaventseva^{1✉}, Natalia P. Podolyan², Maxim A. Volynsky³,
Valery V. Zaytsev⁴, Anastasiia V. Sakovskaia⁵, Oleg V. Mamontov⁶,
Roman V. Romashko⁷, Alexei A. Kamshilin⁸

^{1,2,3,4,5,6,7,8} Institute of Automation and Control Processes, Far Eastern Branch of the Russian Academy of Sciences, Vladivostok, 690041, Russian Federation

³ ITMO University, Saint Petersburg, 197101, Russian Federation

^{4,6} Almazov National Medical Research Centre, Saint Petersburg, 197341, Russian Federation

⁵ Pacific State Medical University, Vladivostok, 690002, Russian Federation

⁶ Pavlov First Saint Petersburg State Medical University, Saint Petersburg, 197022, Russian Federation

¹ ange202@mail.ru✉, <https://orcid.org/0000-0001-8218-6766>

² podolian@iacp.dvo.ru, <https://orcid.org/0000-0002-2108-4376>

³ maxim.volynsky@gmail.com, <https://orcid.org/0000-0003-1679-4450>

⁴ zaytsevphoto@gmail.com, <https://orcid.org/0000-0003-3366-8376>

⁵ sakovska86@mail.ru, <https://orcid.org/0000-0002-4077-9739>

⁶ mamontoffoleg@gmail.com, <https://orcid.org/0000-0003-3063-439X>

⁷ romashko@dvo.ru, <https://orcid.org/0000-0003-0869-0993>

⁸ alexei.kamshilin@yandex.ru, <https://orcid.org/0000-0003-2506-9736>

Abstract

The possibility of using a new contactless method of imaging photoplethysmography to assess thermoregulatory vasodilatation of blood vessels was studied. Perfusion reaction in a region of the outer forearm in response to local heating up to 41 ± 1 °C was monitored in six volunteers aged 39–52 years using a video recording of the study area, synchronized with an electrocardiogram, and subsequent correlation processing of the data obtained. It was shown that the change in perfusion during local heating has a biphasic type and is due to the response of the nervous system mediated by the axon reflex in the first phase of vasodilation and the synthesis of nitric oxide in endothelial cells in the second phase of vasodilation. It was revealed that the multiple increase in perfusion in the first phase of heating depends both on the initial temperature of the skin and on the difference in its heating temperature. It was found that for a significant development of a vascular response to hyperthermia associated with the activation of endothelial function, heating of tissues for more than 15 minutes is necessary. It was shown that the method of imaging photoplethysmography reliably reflects the work of the mechanisms of regulation of peripheral vascular resistance which is of great prognostic value for the detection of primary signs of cardiovascular diseases.

Keywords

imaging photoplethysmography, microcirculation, vascular reactivity, blood perfusion, thermoregulation

Acknowledgements

The research was supported by the Russian Science Foundation (Grant 21-15-00265) in terms of the development of a measuring system, including software, and carrying out experiments. Processing and storage of experimental data were carried out using the equipment of Shared Resource Center “Far Eastern Computing Resource” of the Institute of Automation and Control Processes, Far Eastern Branch of the Russian Academy of Sciences.

For citation: Belaventseva A.V., Podolyan N.P., Volynsky M.A., Zaytsev V.V., Sakovskaia A.V., Mamontov O.V., Romashko R.V., Kamshilin A.A. Study of blood vessels reaction to local heating by imaging photoplethysmography. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 2023, vol. 23, no. 1, pp. 14–20 (in Russian). doi: 10.17586/2226-1494-2023-23-1-14-20

Введение

Микроциркуляторная система — часть системы кровообращения, которая обеспечивает ткани питательными веществами, удаление из них продуктов метаболизма и доставку кислорода. Работу системы микроциркуляции обеспечивает целый ряд функциональных процессов, регулируемых гормональной и вегетативной нервными системами. Эндотелий представляет собой важный эндокринный орган, играющий огромную роль в регуляции тонуса в микроциркуляторном русле и в поддержании сосудистого гомеостаза. Среди эндогенных регуляторов сердечно-сосудистой

системы, синтезируемых в клетках эндотелия, различают вазодилататоры (простациклин, эндотелиальные гиперполяризующие факторы и оксид азота (NO)), и вазоконстрикторы [1]. Эндотелиальная дисфункция, обусловленная дисбалансом секретируемых эндотелием NO и других биологически активных веществ, предшествует развитию любых патологических изменений в сосудах и носит системный характер. Так, эндотелиальной дисфункции способствует артериальная гипертензия [2, 3], курение [4], сахарный диабет [5, 6], нарушение толерантности к глюкозе, ожирение и дислипидемия [7]. Таким образом, оценка вазодилатационных свойств, связанных с содержанием NO, может

предоставить информацию о целостности и функции эндотелия и имеет важное прогностическое значение при выявлении первичных сердечно-сосудистых патологий.

Участие кожной микроциркуляции в терморегуляции человека характеризуется чрезвычайной лабильностью кровотока в ответ на температурные воздействия [8]. Локальный нагрев кожи вызывает локальную гиперемию, характеризующуюся двухфазным усилением кожного кровотока. Быстрый начальный пик наблюдается в течение 2–5 мин после начала нагревания и обуславливается аксон-рефлексом местного чувствительного нерва [8, 9]. В работе [10] показано, что в аксон-рефлекс вносит важный вклад катион-селективный ионный канал ваниллоидного типа 1 (TRPV-1), преимущественно локализованный на чувствительных нервах. За первоначальным пиком следует короткий спад, а затем продолжительное плато в течение 20–30 минутного нагревания. Позднее плато зависит преимущественно от эндотелиальных факторов, среди которых на NO приходится примерно две трети вклада [11]. Подтверждение роли эндотелия получено в работе [12], где обнаружено, что эндотелиальный ингибитор NG-амино-1-аргинин (NOS) не оказывает существенного влияния на рефлекторный сосудорасширяющий ответ, но снижает более чем на 50 % прирост кожного кровотока при местном нагреве. Эти данные убедительно показывают важную роль NOS для кожной вазодилаторной реакции в ответ на локальное нагревание кожи. Очевидно, что тепловая гиперемия кожи может быть использована для определения признаков глобальной микрососудистой дисфункции.

На сегодняшний день для оценки нарушений микроциркуляции крови при различных сердечно-сосудистых патологиях широко применяется метод тепловых проб [9]. Большинство особенностей реакции микроциркуляции на локальное нагревание кожи получены с помощью метода лазерной доплеровской флуометрии (ЛДФ) [13, 14]. Среди недостатков данного метода выделяют низкую воспроизводимость и чувствительность результатов к условиям проведения исследования. Кроме того, датчик ЛДФ оценивает параметры кровотока в одной точке контакта с кожей, а механизм сканирования, необходимый для визуализации, ограничивает скорость исследования и усложняет систему. Заметим, что, несмотря на почти полувековую историю использования метода ЛДФ в исследовании микроциркуляции, он до сих пор не применяется в клинической практике.

В последнее время все чаще используется метод визуализирующей фотоплетизмографии (ВФПГ) для оценки перфузии тканей *in vivo*, позволяя количественно оценить пространственное распределение параметров кровотока [15, 16]. В настоящей работе для оценки изменений перфузии ткани кровью, обусловленных термическим воздействием на биологическую ткань, использован метод ВФПГ, который дает возможность бесконтактного исследования большой площади ткани, что значительно повышает достоверность полученных результатов [17].

Метод визуализирующей фотоплетизмографии

Схема измерительной системы, основанная на методе ВФПГ, представлена на рис. 1. На внешнюю область предплечья наложена стеклянная пластина размером $70 \times 20 \times 2$ мм с прозрачным токопроводящим слоем, обеспечивающим локальный нагрев исследуемого участка. Исследуемая область равномерно освещена зелеными светодиодами (длина волны $\lambda = 530 \pm 25$ нм, выходная мощность $P = 7,5$ Вт), а набор изображений в отраженном свете зарегистрирован видеокамерой. Давление, оказываемое нагревательным элементом на кожу, измерено для каждого испытуемого. Во всех случаях оно не превышает 15 мм рт. ст., что не препятствует нормальному кровотоку в области измерения, и обеспечивает стабильную теплопроводность для передачи тепла от нагревателя.

Известно [18], что именно зеленый свет после взаимодействия с биологической тканью, содержащей кровеносные сосуды, приобретает наибольшую модуляцию на частоте сердцебиений, несмотря на то, что глубина его проникновения в кожу не превышает 0,5 мм. В рамках альтернативной модели формирования фотоплетизмографического сигнала модуляция обусловлена механическим изменением плотности капилляров в дерме вследствие пульсаций близлежащих артерий и артериол [19]. Таким образом, капиллярное русло оказывается распределенным преобразователем пульсаций глуболежащих артерий в модуляцию света на частоте сердцебиений. Так как амплитуда артериальных пульсаций определяет перфузию в исследуемой области (чем больше амплитуда пульсаций, тем выше перфузия),

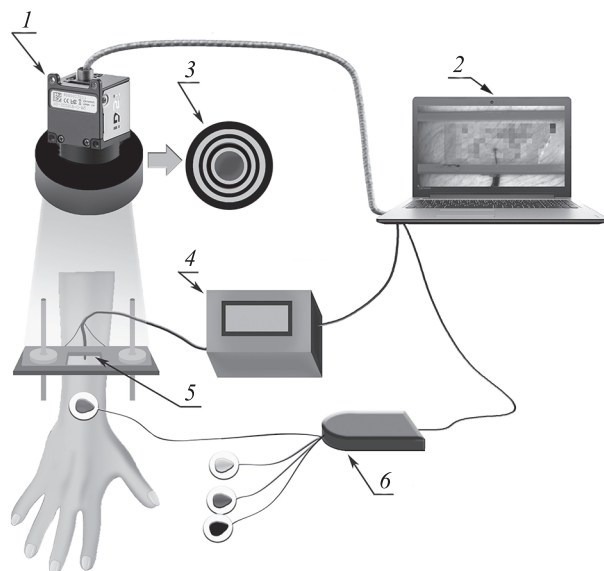


Рис. 1. Схема измерительной системы: 1 — цифровая видеокамера; 2 — персональный компьютер; 3 — светодиодный осветитель; 4 — блок контроля температуры; 5 — нагревательное устройство; 6 — электрокардиограф

Fig. 1. Schematic of the measuring system: 1 — digital video camera; 2 — computer; 3 — light-emitting diode illuminator; 4 — heating control unit; 5 — heating element; 6 — electrocardiograph

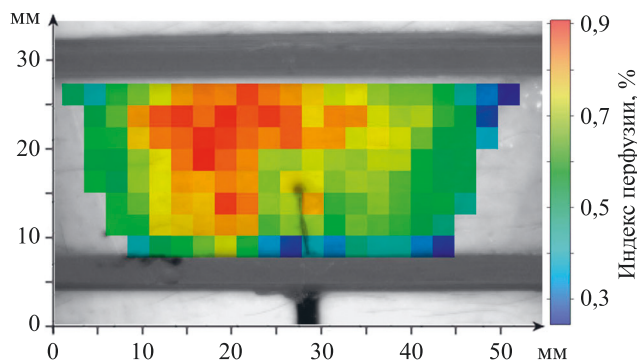


Рис. 2. Пример пространственного распределения индекса перфузии в области локального нагрева на 500 с после нагрева

Fig. 2. An example of spatial distribution of the perfusion index in the area of local heating at 500th s after the heating

то амплитуда модуляции света на частоте сердцебиений характеризует локальную перфузию исследуемого участка [20]. Соответственно, пространственные распределения амплитуды модуляции изображения на частоте сердцебиений отражают распределение перфузии (рис. 2).

Для регистрации отраженного света использована монохромная 10-разрядная видеокамера UI-3060CP-M-GL (Imaging Development Systems GmbH, Оберзульм, Германия) с объективом M1214-MP2 (Computar, Токио, Япония, $F = 12$ мм, угол обзора $89^\circ \times 67^\circ$). Камера закреплена в одном корпусе с осветителем, состоящем из 250 светодиодов, расположенных концентрическими кольцами вокруг линзы объектива (диаметр наибольшего кольца 9,8 см), что обеспечивает равномерное освещение области исследования на расстоянии около 30 см. Видеозапись осуществлена синхронно с записью электрокардиограммы (электрокардиограф «Кардиотехника-ЭКГ-8», производства «ИНКАРТ», Санкт-Петербург, Россия) для последующего получения надежного сигнала фотоплетизмографии. Видеокадры со сфокусированными изображениями освещенного участка кожи записаны с частотой 36 кадров в секунду на персональный компьютер через универсальную последовательную шину (USB) и обработаны с помощью специального программного обеспечения, реализованного на платформе MATLAB. Регистрируемый сигнал каждого пиксела изображения состоит из переменной составляющей (AC), модулированной на частоте сердцебиений, и медленно меняющейся (DC), которая относится к низкочастотным модуляциям, обусловленным процессами дыхания и работой нервной системы [21]. Для построения карт распределения перфузии использовано отношение AC/DC, которое уменьшает влияние неоднородности освещения и рельефа кожи на вычисляемый индекс перфузии [20]. Для уменьшения влияния бликов от верхнего слоя эпидермиса и поверхности стекла нагревательного элемента применен метод поляризационной фильтрации, реализованный посредством установки тонкопленочных поляризаторов с взаимно ортогональной ориентацией после излучателя и перед камерой.

В процессе обработки изображений для получения достоверного сигнала пульсовой волны в каждом пикселе начало каждого импульса нормализованного сигнала AC/DC было совмещено с соответствующим R-пиком электрокардиограммы. Затем пульсовые волны в 15 последовательных кардиоциклах усреднялись для вычисления индекса перфузии в каждом пикселе [18, 20]. Проведена оценка индекса перфузии на половине площади контакта, характеризующейся более высокой реакцией на нагревание. Расчет индекса перфузии выполнен как разность максимального и минимального значений усредненной формы импульса отношения AC/DC.

Локальная гипертермия кожных покровов до температуры $41 \pm 1^\circ\text{C}$ осуществлена с помощью прозрачной токопроводящей пластины и проконтролирована на персональном компьютере. Отметим, что токопроводящий слой не контактировал с кожей, а вазелин, нанесенный между стеклянной пластиной и кожей, обеспечил теплопроводность. Регистрация и контроль температуры осуществлены посредством термодатчика, расположенной между пластиной и кожей, синхронно с записываемыми видеокадрами.

Протокол исследования

Все измерения проведены в затемненной лаборатории в отсутствие посторонних источников излучения при температуре окружающей среды 23°C . В исследовании приняло участие 6 здоровых добровольцев мужского пола в возрасте 39–52 лет.

Перфузионная реакция на локальный нагрев оценена в три этапа. На первом этапе выполнена непрерывная видеозапись исследуемого участка синхронно с измерением сигнала ЭКГ и температуры кожи в течение 5 мин. На втором — осуществлен нагрев области предплечья до $41 \pm 1^\circ\text{C}$ со скоростью около $3,5^\circ\text{C}$ в минуту, чтобы избежать болезненных ощущений. Достигнутый уровень температуры кожи поддерживается в течение нескольких минут. Для каждого из добровольцев проведено по три исследования с различными длительностями нагрева в 7, 15 и 20 мин. На последнем этапе нагреватель выключался, и температура кожи релаксировала естественным образом. Проведено не более одного исследования каждого испытуемого в день. Исследования выполнены с разрешения междисциплинарного комитета по этике Федерального государственного бюджетного образовательного учреждения высшего образования «Тихоокеанский государственный медицинский университет» Минздрава России, протокол №10 от 21.06.2021.

Результаты

С помощью системы ВФПГ у всех испытуемых было выявлено многократное увеличение амплитуды пульсаций кожного кровотока после локального нагрева наружной части предплечья. Изменение индекса перфузии носит двухфазный характер. Во время фазы 1 вазодилатации (после включения нагревательной пластины) происходит быстрый рост перфузии, обусловленный

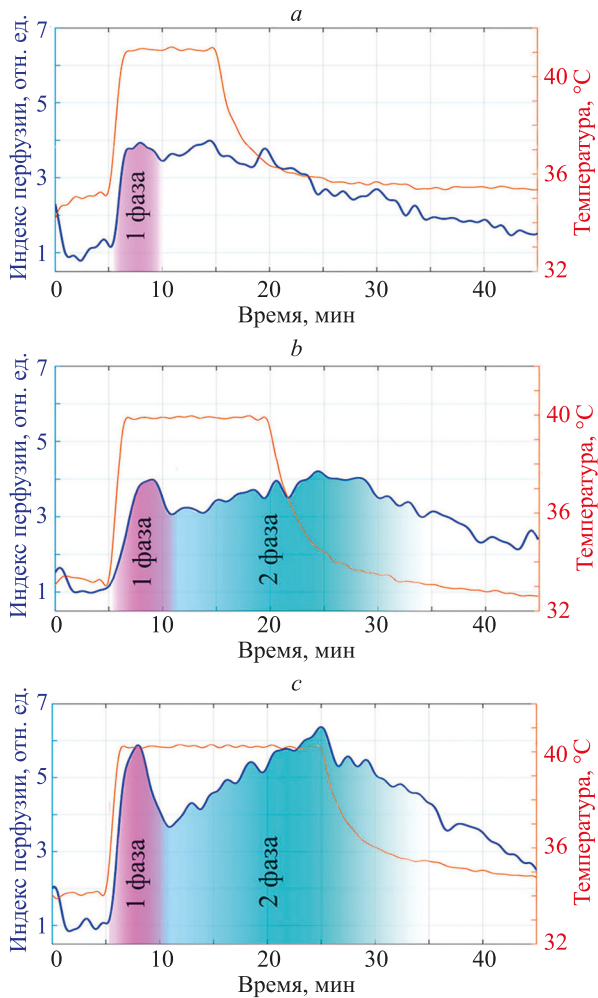


Рис. 3. Отклик перфузии на локальный нагрев при различной его длительности: 7 мин (a); 15 мин (b); 20 мин (c)

Fig. 3. The perfusion response to local heating at different durations — 7 minutes (a), 15 minutes (b), and 20 minutes (c)

аксон-рефлексом (рис. 3). Установлено, что величина перфузионного отклика на гипертермию в фазе 1 вазодилатации зависит от начальной температуры кожного покрова и разницы температур его нагрева. Проведенный корреляционный анализ (тест Пирсона) показал достоверную положительную корреляцию роста перфузии с разницей температур (коэффициент корреляции $r = 0,92$, уровень значимости $p < 0,001$) и отрицательную с начальной температурой ($r = -0,90$, $p < 0,001$).

Фаза 2 вазодилатации развивается при продолжительном нагреве, при котором, вслед за падением пер-

фузии и появлением локального минимума на кривой, происходит усиление интенсивности кровотока, что приводит к росту перфузии и появлению второго максимума (рис. 3, b, c). Такой рост перфузии связан с синтезом NO [11, 14]. Так, с увеличением продолжительности гиперемии, увеличивается перфузия в фазе 2, что выражается в статистически значимом увеличении интегрального отклика перфузии в группе с 15-минутным нагревом по сравнению с группой с 7-минутным нагревом, $p = 0,028$ (тест Уилкоксона), в то время как в перфузионных реакциях фазы 1 различий между группами не обнаружено. Таким образом, для существенного развития сосудистой реакции на гипертермию, опосредованную синтезом NO, необходим нагрев тканей длительностью более 15 мин. После отключения нагревательной пластины происходит снижение перфузии до ее первоначального значения.

Выводы

Выполненное исследование продемонстрировало возможность оценки реакции кровеносных сосудов на локальный нагрев методом визуализирующей фотоплетизмографии. Представленные зависимости отклика перфузии от локальной гипертермии являются характерными и согласуются с результатами, полученными при исследованиях вазодилатации с помощью ЛДФ [14]. Предложенный метод ВФПГ позволяет осуществлять мониторинг сосудистой реакции в значительной части области исследования в ответ на дозированную локальную гипертермию, чем принципиально отличается от других оптических методов диагностики. Стабильность локального нагрева кожи достигнута с использованием новой мультимодальной системы, в которой нагревательное стекло контактирует с кожей, в то время как перфузия оценена бесконтактным способом с помощью визуализирующей фотоплетизмографии.

Заключение

Таким образом, метод визуализирующей фотоплетизмографии достоверно отражает регуляцию сосудистого тонуса, что позволит в дальнейшем изучать роль дисфункции эндотелия в патогенезе сердечно-сосудистых патологии и исследовать различные физиологические нарушения, вызванные ей. Предлагаемая система обладает рядом преимуществ по сравнению с доступными методами оценки микроциркуляции: отличается низкой стоимостью, высокой информативностью и воспроизводимостью результатов.

Литература

1. Gimbrone M., Garcia-Cardena G. Endothelial cell dysfunction and the pathobiology of atherosclerosis // *Circulation Research*. 2016. V. 118. N 4. P. 620–636. <https://doi.org/10.1161/circresaha.115.306301>
2. Widlansky M.E., Gokce N., Keaney J.F., Vita J.A. The clinical implications of endothelial dysfunction // *Journal of the American College of Cardiology*. 2003. V. 42. N 7. P. 1149–1160. [https://doi.org/10.1016/s0735-1097\(03\)00994-x](https://doi.org/10.1016/s0735-1097(03)00994-x)

References

1. Gimbrone M., Garcia-Cardena G. Endothelial cell dysfunction and the pathobiology of atherosclerosis. *Circulation Research*, 2016, vol. 118, no. 4, pp. 620–636. <https://doi.org/10.1161/circresaha.115.306301>
2. Widlansky M.E., Gokce N., Keaney J.F., Vita J.A. The clinical implications of endothelial dysfunction. *Journal of the American College of Cardiology*, 2003, vol. 42, no. 7, pp. 1149–1160. [https://doi.org/10.1016/s0735-1097\(03\)00994-x](https://doi.org/10.1016/s0735-1097(03)00994-x)

3. Dharmashankar K., Widlansky M.E. Vascular endothelial function and hypertension: insights and directions // *Current Hypertension Reports*. 2010. V. 12. N 6. P. 448–455. <https://doi.org/10.1007/s11906-010-0150-2>
4. Celermajer D.S., Sorensen K.E., Georgakopoulos D., Bull C., Thomas O., Robinson J., Deanfield J.E. Cigarette smoking is associated with dose-related and potentially reversible impairment of endothelium-dependent dilation in healthy young adults // *Circulation*. 1993. V. 88. N 5. P. 2149–2155. <https://doi.org/10.1161/01.cir.88.5.2149>
5. Mäkimattila S., Virkamäki A., Groop P.-H., Cockcroft J., Utriainen T., Fagerudd J., Yki-Järvinen H. Chronic hyperglycemia impairs endothelial function and insulin sensitivity via different mechanisms in insulin-dependent diabetes mellitus // *Circulation*. 1996. V. 94. N 6. P. 1276–1282. <https://doi.org/10.1161/01.cir.94.6.1276>
6. Williams S.B., Cusco J.A., Roddy M.A., Johnstone M.T., Creager M.A. Impaired nitric oxide-mediated vasodilation in patients with non-insulin-dependent diabetes mellitus // *Journal of the American College of Cardiology*. 1996. V. 27. N 3. P. 567–574. [https://doi.org/10.1016/0735-1097\(95\)00522-6](https://doi.org/10.1016/0735-1097(95)00522-6)
7. Spieker L.E., Sudano I., Hürliemann D., Lerch P.G., Lang M.G., Binggeli C., Corti R., Ruschitzka F., Lüscher T.F., Noll G. High-density lipoprotein restores endothelial function in hypercholesterolemic men // *Circulation*. 2002. V. 105. N 12. P. 1399–1402. <https://doi.org/10.1161/01.cir.0000013424.28206.8f>
8. Minson C.T. Thermal provocation to evaluate microvascular reactivity in human skin // *Journal of Applied Physiology*. 2010. V. 109. N 9. P. 1239–1246. <https://doi.org/10.1152/jappphysiol.00414.2010>
9. Minson C.T., Berry L.T., Joyner M.J. Nitric oxide and neurally mediated regulation of skin blood flow during local heating // *Journal of Applied Physiology*. 2001. V. 91. N 4. P. 1619–1626. <https://doi.org/10.1152/jappl.2001.91.4.1619>
10. Wong B.J., Fieger S.M. Transient receptor potential vanilloid type-1 (TRPV-1) channels contribute to cutaneous thermal hyperaemia in humans // *The Journal of Physiology*. 2010. V. 588. N 21. P. 4317–4326. <https://doi.org/10.1113/jphysiol.2010.195511>
11. Brunt V.E., Minson C.T. Cutaneous thermal hyperemia: more than skin deep // *Journal of Applied Physiology*. 2011. V. 111. N 1. P. 5–7. <https://doi.org/10.1152/jappphysiol.00544.2011>
12. Kellogg D.L., Zhao J.L., Wu Y. Endothelial nitric oxide synthase control mechanisms in the cutaneous vasculature of humans in vivo // *American Journal of Physiology — Heart and Circulatory Physiology*. 2008. V. 295. N 1. P. H123–H129. <https://doi.org/10.1152/ajpheart.00082.2008>
13. Debbabi H., Bonnin Ph., Ducluzeau P.H., Leftheriotis G., Levy B.I. Noninvasive assessment of endothelial function in the skin microcirculation // *American Journal of Hypertension*. 2010. V. 23. N 5. P. 541–546. <https://doi.org/10.1038/ajh.2010.10>
14. Cracowski J.-L., Minson C.T., Salvat-Melis M., Halliwill J. Methodological issues in the assessment of skin microvascular endothelial function in humans // *Trends in Pharmacological Sciences*. 2006. V. 27. N 9. P. 503–508. <https://doi.org/10.1016/j.tips.2006.07.008>
15. Mamontov O.V., Shcherbinin A.V., Romashko R.V., Kamshilin A.A. Intraoperative imaging of cortical blood flow by camera-based photoplethysmography at green light // *Applied Sciences*. 2020. V. 10. N 18. P. 6192. <https://doi.org/10.3390/app10186192>
16. Kukul I., Trumpp A., Plötze K., Rost A., Zaunseder S., Matschke K., Rasche S. Contact-free optical assessment of changes in the chest wall perfusion after coronary artery bypass grafting by imaging photoplethysmography // *Applied Sciences*. 2020. V. 10. N 18. P. 6537. <https://doi.org/10.3390/app10186537>
17. Volynsky M.A., Margaryants N.B., Mamontov O.V., Kamshilin A.A. Contactless monitoring of microcirculation reaction on local temperature changes // *Applied Sciences*. 2019. V. 9. N 22. P. 4947. <https://doi.org/10.3390/app9224947>
18. Cui W.J., Ostrander L.E., Lee B.Y. In vivo reflectance of blood and tissue as a function of light wavelength // *IEEE Transactions on Biomedical Engineering*. 1990. V. 37. N 6. P. 632–639. <https://doi.org/10.1109/10.55667>
19. Kamshilin A.A., Nippolainen E., Sidorov I.S., Vasilev P.V., Erofeev N.P., Podolian N.P., Romashko R.V. A new look at the essence of the imaging photoplethysmography // *Scientific Reports*. 2015. V. 5. N 1. P. 10494. <https://doi.org/10.1038/srep10494>
20. Lyubashina O.A., Mamontov O., Volynsky M., Zaytsev V.V., Kamshilin A.A. Contactless assessment of cerebral autoregulation by
3. Dharmashankar K., Widlansky M.E. Vascular endothelial function and hypertension: insights and directions. *Current Hypertension Reports*, 2010, vol. 12, no. 6, pp. 448–455. <https://doi.org/10.1007/s11906-010-0150-2>
4. Celermajer D.S., Sorensen K.E., Georgakopoulos D., Bull C., Thomas O., Robinson J., Deanfield J.E. Cigarette smoking is associated with dose-related and potentially reversible impairment of endothelium-dependent dilation in healthy young adults. *Circulation*, 1993, vol. 88, no. 5, pp. 2149–2155. <https://doi.org/10.1161/01.cir.88.5.2149>
5. Mäkimattila S., Virkamäki A., Groop P.-H., Cockcroft J., Utriainen T., Fagerudd J., Yki-Järvinen H. Chronic hyperglycemia impairs endothelial function and insulin sensitivity via different mechanisms in insulin-dependent diabetes mellitus. *Circulation*, 1996, vol. 94, no. 6, pp. 1276–1282. <https://doi.org/10.1161/01.cir.94.6.1276>
6. Williams S.B., Cusco J.A., Roddy M.A., Johnstone M.T., Creager M.A. Impaired nitric oxide-mediated vasodilation in patients with non-insulin-dependent diabetes mellitus. *Journal of the American College of Cardiology*, 1996, vol. 27, no. 3, pp. 567–574. [https://doi.org/10.1016/0735-1097\(95\)00522-6](https://doi.org/10.1016/0735-1097(95)00522-6)
7. Spieker L.E., Sudano I., Hürliemann D., Lerch P.G., Lang M.G., Binggeli C., Corti R., Ruschitzka F., Lüscher T.F., Noll G. High-density lipoprotein restores endothelial function in hypercholesterolemic men. *Circulation*, 2002, vol. 105, no. 12, pp. 1399–1402. <https://doi.org/10.1161/01.cir.0000013424.28206.8f>
8. Minson C.T. Thermal provocation to evaluate microvascular reactivity in human skin. *Journal of Applied Physiology*, 2010, vol. 109, no. 9, pp. 1239–1246. <https://doi.org/10.1152/jappphysiol.00414.2010>
9. Minson C.T., Berry L.T., Joyner M.J. Nitric oxide and neurally mediated regulation of skin blood flow during local heating. *Journal of Applied Physiology*, 2001, vol. 91, no. 4, pp. 1619–1626. <https://doi.org/10.1152/jappl.2001.91.4.1619>
10. Wong B.J., Fieger S.M. Transient receptor potential vanilloid type-1 (TRPV-1) channels contribute to cutaneous thermal hyperaemia in humans. *The Journal of Physiology*, 2010, vol. 588, no. 21, pp. 4317–4326. <https://doi.org/10.1113/jphysiol.2010.195511>
11. Brunt V.E., Minson C.T. Cutaneous thermal hyperemia: more than skin deep. *Journal of Applied Physiology*, 2011, vol. 111, no. 1, pp. 5–7. <https://doi.org/10.1152/jappphysiol.00544.2011>
12. Kellogg D.L., Zhao J.L., Wu Y. Endothelial nitric oxide synthase control mechanisms in the cutaneous vasculature of humans in vivo. *American Journal of Physiology — Heart and Circulatory Physiology*, 2008, vol. 295, no. 1, pp. H123–H129. <https://doi.org/10.1152/ajpheart.00082.2008>
13. Debbabi H., Bonnin Ph., Ducluzeau P.H., Leftheriotis G., Levy B.I. Noninvasive assessment of endothelial function in the skin microcirculation. *American Journal of Hypertension*, 2010, vol. 23, no. 5, pp. 541–546. <https://doi.org/10.1038/ajh.2010.10>
14. Cracowski J.-L., Minson C.T., Salvat-Melis M., Halliwill J. Methodological issues in the assessment of skin microvascular endothelial function in humans. *Trends in Pharmacological Sciences*, 2006, vol. 27, no. 9, pp. 503–508. <https://doi.org/10.1016/j.tips.2006.07.008>
15. Mamontov O.V., Shcherbinin A.V., Romashko R.V., Kamshilin A.A. Intraoperative imaging of cortical blood flow by camera-based photoplethysmography at green light. *Applied Sciences*, 2020, vol. 10, no. 18, pp. 6192. <https://doi.org/10.3390/app10186192>
16. Kukul I., Trumpp A., Plötze K., Rost A., Zaunseder S., Matschke K., Rasche S. Contact-free optical assessment of changes in the chest wall perfusion after coronary artery bypass grafting by imaging photoplethysmography. *Applied Sciences*, 2020, vol. 10, no. 18, pp. 6537. <https://doi.org/10.3390/app10186537>
17. Volynsky M.A., Margaryants N.B., Mamontov O.V., Kamshilin A.A. Contactless monitoring of microcirculation reaction on local temperature changes. *Applied Sciences*, 2019, vol. 9, no. 22, pp. 4947. <https://doi.org/10.3390/app9224947>
18. Cui W.J., Ostrander L.E., Lee B.Y. In vivo reflectance of blood and tissue as a function of light wavelength. *IEEE Transactions on Biomedical Engineering*, 1990, vol. 37, no. 6, pp. 632–639. <https://doi.org/10.1109/10.55667>
19. Kamshilin A.A., Nippolainen E., Sidorov I.S., Vasilev P.V., Erofeev N.P., Podolian N.P., Romashko R.V. A new look at the essence of the imaging photoplethysmography. *Scientific Reports*, 2015, vol. 5, no. 1, pp. 10494. <https://doi.org/10.1038/srep10494>
20. Lyubashina O.A., Mamontov O., Volynsky M., Zaytsev V.V., Kamshilin A.A. Contactless assessment of cerebral autoregulation by

- photoplethysmographic imaging at green illumination // *Frontiers in Neuroscience*, 2019. V. 13. P. 1235. <https://doi.org/10.3389/fnins.2019.01235>
21. Allen J. Photoplethysmography and its application in clinical physiological measurement // *Physiological Measurement*. 2007. V. 28. N 3. P. R1–R39. <https://doi.org/10.1088/0967-3334/28/3/R01>

- photoplethysmographic imaging at green illumination. *Frontiers in Neuroscience*, 2019, vol. 13, pp. 1235. <https://doi.org/10.3389/fnins.2019.01235>
21. Allen J. Photoplethysmography and its application in clinical physiological measurement. *Physiological Measurement*, 2007, vol. 28, no. 3, pp. R1–R39. <https://doi.org/10.1088/0967-3334/28/3/R01>

Авторы

Белавенцева Анжелика Вадимовна — младший научный сотрудник, Институт автоматизации и процессов управления Дальневосточного отделения Российской академии наук, Владивосток, 690041, Российская Федерация, [sc 57192421923](https://orcid.org/0000-0001-8218-6766), <https://orcid.org/0000-0001-8218-6766>, ange202@mail.ru

Подольня Наталья Петровна — кандидат физико-математических наук, младший научный сотрудник, Институт автоматизации и процессов управления Дальневосточного отделения Российской академии наук, Владивосток, 690041, Российская Федерация, [sc 56707708600](https://orcid.org/0000-0002-2108-4376), <https://orcid.org/0000-0002-2108-4376>, podolian@iacp.dvo.ru

Волюнский Максим Александрович — кандидат технических наук, доцент, ведущий научный сотрудник, Институт автоматизации и процессов управления Дальневосточного отделения Российской академии наук, Владивосток, 690041, Российская Федерация; доцент, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, [sc 23006901100](https://orcid.org/0000-0003-1679-4450), <https://orcid.org/0000-0003-1679-4450>, maxim.volynsky@gmail.com

Зайцев Валерий Викторович — младший научный сотрудник, Институт автоматизации и процессов управления Дальневосточного отделения Российской академии наук, Владивосток, 690041, Российская Федерация; младший научный сотрудник, Национальный медицинский исследовательский центр им. В.А. Алмазова Минздрава России, Санкт-Петербург, 197341, Российская Федерация, [sc 57193801729](https://orcid.org/0000-0003-3366-8376), <https://orcid.org/0000-0003-3366-8376>, zaytsevphoto@gmail.com

Саковская Анастасия Владимировна — кандидат медицинских наук, младший научный сотрудник, Институт автоматизации и процессов управления Дальневосточного отделения Российской академии наук, Владивосток, 690041, Российская Федерация; доцент, Тихоокеанский государственный медицинский университет, Владивосток, 690002, Российская Федерация, [sc 57203402443](https://orcid.org/0000-0002-4077-9739), <https://orcid.org/0000-0002-4077-9739>, sakovska86@mail.ru

Мамонтов Олег Викторович — доктор медицинских наук, ведущий научный сотрудник, Институт автоматизации и процессов управления Дальневосточного отделения Российской академии наук, Владивосток, 690041, Российская Федерация; старший научный сотрудник, Национальный медицинский исследовательский центр им. В.А. Алмазова Минздрава России, Санкт-Петербург, 197341, Российская Федерация; ассистент, Первый Санкт-Петербургский государственный медицинский университет имени Павлова, Санкт-Петербург, 197022, Российская Федерация, [sc 6505846605](https://orcid.org/0000-0003-3063-439X), <https://orcid.org/0000-0003-3063-439X>, mamontoffoleg@gmail.com

Ромашко Роман Владимирович — доктор физико-математических наук, член-корреспондент РАН, руководитель лаборатории, Институт автоматизации и процессов управления Дальневосточного отделения Российской академии наук, Владивосток, 690041, Российская Федерация, [sc 6603647454](https://orcid.org/0000-0003-0869-0993), <https://orcid.org/0000-0003-0869-0993>, romashko@dvo.ru

Камшилин Алексей Александрович — доктор физико-математических наук, старший научный сотрудник, главный научный сотрудник Институт автоматизации и процессов управления Дальневосточного отделения Российской академии наук, Владивосток, 690041, Российская Федерация, [sc 7007138092](https://orcid.org/0000-0003-2506-9736), <https://orcid.org/0000-0003-2506-9736>, alexei.kamshilin@yandex.ru

Authors

Anzhelika V. Belaventseva — Junior Researcher, Institute of Automation and Control Processes, Far Eastern Branch of the Russian Academy of Sciences, Vladivostok, 690041, Russian Federation, [sc 57192421923](https://orcid.org/0000-0001-8218-6766), <https://orcid.org/0000-0001-8218-6766>, ange202@mail.ru

Natalia P. Podolyan — PhD (Physics & Mathematics), Junior Researcher, Institute of Automation and Control Processes, Far Eastern Branch of the Russian Academy of Sciences, Vladivostok, 690041, Russian Federation, [sc 56707708600](https://orcid.org/0000-0002-2108-4376), <https://orcid.org/0000-0002-2108-4376>, podolian@iacp.dvo.ru

Maxim A. Volynsky — PhD (Technical science), Associate Professor, Leading Researcher, Institute of Automation and Control Processes, Far Eastern Branch of the Russian Academy of Sciences, Vladivostok, 690041, Russian Federation; Associate Professor, ITMO University, Saint Petersburg, 197101, Russian Federation, [sc 23006901100](https://orcid.org/0000-0003-1679-4450), <https://orcid.org/0000-0003-1679-4450>, maxim.volynsky@gmail.com

Valery V. Zaytsev — Junior Researcher, Institute of Automation and Control Processes, Far Eastern Branch of the Russian Academy of Sciences, Vladivostok, 690041; Junior Researcher, Almazov National Medical Research Centre, Saint Petersburg, 197341, Russian Federation, [sc 57193801729](https://orcid.org/0000-0003-3366-8376), <https://orcid.org/0000-0003-3366-8376>, zaytsevphoto@gmail.com

Anastasiia V. Sakovskaia — PhD (Medicine), Junior Researcher, Institute of Automation and Control Processes, Far Eastern Branch of the Russian Academy of Sciences, Vladivostok, 690002, Russian Federation; Associate Professor, Pacific State Medical University, Vladivostok, 690002, Russian Federation, [sc 57203402443](https://orcid.org/0000-0002-4077-9739), <https://orcid.org/0000-0002-4077-9739>, sakovska86@mail.ru

Oleg V. Mamontov — D.Sc. (Medicine), Leading Researcher, Institute of Automation and Control Processes, Far Eastern Branch of the Russian Academy of Sciences, Vladivostok, 690041, Russian Federation; Senior Researcher, Almazov National Medical Research Centre, Saint Petersburg, 197341, Russian Federation; Assistant, Pavlov First Saint Petersburg State Medical University, Saint Petersburg, 197022, Russian Federation, [sc 6505846605](https://orcid.org/0000-0003-3063-439X), <https://orcid.org/0000-0003-3063-439X>, mamontoffoleg@gmail.com

Roman V. Romashko — D.Sc. (Physics & Mathematics), Corresponding Member of the Russian Academy of Sciences, Head of Laboratory, Institute of Automation and Control Processes, Far Eastern Branch of the Russian Academy of Sciences, Vladivostok, 690041, Russian Federation, [sc 6603647454](https://orcid.org/0000-0003-0869-0993), <https://orcid.org/0000-0003-0869-0993>, romashko@dvo.ru

Alexei A. Kamshilin — D.Sc. (Physics & Mathematics), Senior Researcher, Chief Researcher, Institute of Automation and Control Processes, Far Eastern Branch of the Russian Academy of Sciences, Vladivostok, 690041, Russian Federation, [sc 7007138092](https://orcid.org/0000-0003-2506-9736), <https://orcid.org/0000-0003-2506-9736>, alexei.kamshilin@yandex.ru

Статья поступила в редакцию 24.10.2022
Одобрена после рецензирования 29.11.2022
Принята к печати 29.01.2023

Received 24.10.2022
Approved after reviewing 29.11.2022
Accepted 29.01.2023



Работа доступна по лицензии
Creative Commons
«Attribution-NonCommercial»

doi: 10.17586/2226-1494-2023-23-1-21-27

УДК 535.421; 535.417

Передача 3D-голографической информации по радиоканалу методом, близким к SSB

Артём Леонович Пазоев¹, Сергей Александрович Шойдин²✉

^{1,2} Сибирский государственный университет геосистем и технологий, Новосибирск, 630108, Российская Федерация

¹ pazoev-al2018@sgugit.ru, <https://orcid.org/0000-0003-0302-4860>

² shoydin@ssga.ru✉, <https://orcid.org/0000-0002-2186-7928>

Аннотация

Предмет исследования. Приведены результаты исследования возможности передачи голографической информации по радиоканалу Wi-Fi 40 МГц. Показано, что использование для этого двух основных модальностей 3D-изображения — карты глубин голографируемого объекта и текстуры его поверхности достаточно для синтеза изображения на приемном конце канала связи полноценной голограммы, восстанавливающей объект голографирования с непрерывным вертикальным и горизонтальным параллаксами. **Метод.** Метод передачи 3D-голографической информации подобен известному в радиотехнике методу передачи информации на одной боковой полосе (Single-sideband modulation, SSB). Существенное отличие предложенного метода состоит в том, что пространственные частоты, образующие голограмму, являются результатом одновременной амплитудной и фазовой модуляции опорного сигнала. Это затрудняет их теоретический анализ. Экспериментальное подтверждение возможности такой передачи выполнено с использованием свободного FTP-клиента с открытым исходным кодом FileZilla. Применен протокол связи для передачи информации по беспроводному каналу Wi-Fi.

Основные результаты. Показано, что передаваемый информационный поток достаточен для синтеза на приемном конце канала связи голограммы, восстанавливающей 3D-изображения. При этом голографическое изображение динамически меняющегося объекта с телевизионной частотой кадровой развертки имеет непрерывный горизонтальный и вертикальный параллакс, а пространственное разрешение восстановленного изображения не хуже телевизионного изображения стандарта высокой четкости Full HD. **Практическая значимость.** Экспериментально подтверждена возможность передачи по радиоканалу всей необходимой информации для воспроизведения на приемном конце канала голографического 3D-видеопотока с разрешением не ниже, чем в стандартах телевидения высокой четкости с непрерывным параллаксом.

Ключевые слова

голография, голографическая информация, метод SSB

Ссылка для цитирования: Пазоев А.Л., Шойдин С.А. Передача 3D-голографической информации по радиоканалу методом, близким к SSB // Научно-технический вестник информационных технологий, механики и оптики. 2023. Т. 23, № 1. С. 21–27. doi: 10.17586/2226-1494-2023-23-1-21-27

Transmission of 3D holographic information over a radio channel by a method close to SSB

Artem L. Pazoev¹, Sergey A. Shoydin²✉

^{1,2} Siberian State University of Geosystems and Technologies, Novosibirsk, 630108, Russian Federation

¹ pazoev-al2018@sgugit.ru, <https://orcid.org/0000-0003-0302-4860>

² shoydin@ssga.ru✉, <https://orcid.org/0000-0002-2186-7928>

Abstract

The results of the research on the possibility of transmitting holographic information over the Wi-Fi 40 MHz radio channel are presented. It is shown that the use of two main 3D image modalities for this, — a depth map of the

holographic object and the texture of its surface, is sufficient to synthesize a full-fledged hologram at the receiving end of the communication channel, restoring the holographic object with continuous vertical and horizontal parallax. The method of transmitting 3D holographic information is similar to the well-known in radio engineering method of transmitting information on one sideband (Single-sideband modulation, SSB). The essential difference of the proposed method is that the spatial frequencies forming the hologram are the result of simultaneous amplitude and phase modulation of the reference signal. This complicates their theoretical analysis. Experimental confirmation of the possibility of such a transfer was performed using a free FTP client with open source FileZilla. A communication protocol has been applied to transmit information over a wireless Wi-Fi channel. It is shown that the transmitted information stream is sufficient to synthesize a hologram reconstructing 3D images at the receiving end of the communication channel. At the same time, the holographic image of a dynamically changing object with a television frame rate has continuous horizontal and vertical parallax, and the spatial resolution of the restored image was no worse than a high-definition television image of Full HD. The possibility of transmitting all the necessary information over the radio channel to reproduce a holographic 3D video stream at the receiving end of the channel with a resolution not lower than in high-definition television standards with continuous parallax has been experimentally confirmed.

Keywords

holography, holographic information, SSB method

For citation: Pazoev A.L., Shoydin S.A. Transmission of 3D holographic information over a radio channel by a method close to SSB. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 2023, vol. 23, no. 1, pp. 21–27 (in Russian). doi: 10.17586/2226-1494-2023-23-1-21-27

Введение

Большие объемы голографической информации являются большим преимуществом в различных областях [1–5]. В то же время в области голографического телевидения (ТВ) и задачах дополненной реальности большие объемы информации — одно из существенных препятствий передачи голографической информации по каналам связи [6, 7]. Ю.Н. Денисюк [7] сформулировал данную проблему в заголовке своей работы: «Достаточно ли известны фундаментальные принципы голографии для создания новых типов объемного кинематографа и искусственного интеллекта?». Прямыми решениями проблемы являются: либо существенное уменьшение размера голографического изображения, что плохо соответствует задачам дополненной реальности; либо сжатие голографической информации таким образом, чтобы ее объем не превышал разумных пределов, т. е. возможностей современных каналов связи 3G и 4G.

Оценка количества информации в кадре голографического телевидения

Оценку количества информации в классической аналоговой голограмме, записанной в видимом диапазоне длин волн форматом A4 можно провести по формуле Шеннона:

$$N = SW_x W_y \log(1 + m), \quad (1)$$

где S — площадь голограммы; W_x, W_y — x, y компоненты пространственных частот голограммы сфокусированных изображений, обратно пропорциональные длине волны λ , записывающего голограмму излучения; m — отношение сигнала и шума, или число дискрет яркости пикселей (в терминологии обработки фотографий — глубина цвета изображения).

При количестве дискретных значений каждого пикселя (глубине цвета) равном стандартному в ТВ-системах $m = 256$, получим результат оценки количества информации (1) в одном голографическом кадре — $1,7 \cdot 10^{11}$ Б.

Передача таких массивов с частотой кадровой развертки 25 Гц требует пропускной способности канала по Шеннону [8], оцениваемой по порядку величины, как $C = 4 \cdot 10^{12}$ Б/с.

Ширина полосы пропускания оценивается порядка 500 ГГц, что выводит такую передачу:

$$W = \frac{C}{\log_2(256)}$$

в терагерцовый диапазон [9], тем самым практически закрывая окно возможностей прямой передачи полной голографической информации по ТВ-каналам связи. Другими словами, прямая ТВ-трансляция 3D-кино Комара [10] невозможна.

Следовательно, требуется поиск путей сжатия голографической информации и желательно без существенных потерь.

Постановка задачи

Согласно патенту [11] основная информация о 3D движущемся живом объекте голографирования содержится в потоке двух модальностей 3D-изображения — карте глубин поверхности голографируемого объекта и текстуре его поверхности, что качественно имеет сходство с передачей сообщения на одной боковой полосе, без несущей (Single-sideband modulation, SSB), но отличается наличием амплитудной и фазовой модуляции сигнала. Анализ спектров этих мод и созданной по ним голограммы показывает существенный разрыв, обеспечивающий окно возможностей для эффективного сжатия передаваемой информации. Однако прямые математические методы энтропийного сжатия оказываются неэффективными, на что указывают работы [12–17]. Среди них находится и недавно созданная для кодирования объемных изображений группой «Объединенные фотографы» (Joint Photographic Experts Group, JPEG) группа P10 [18], которая сумела продвинуться в коррекции фокусировки при не резкой съемке, но не получила значимых результатов в кодировании голограмм 3D-объектов.

Следовательно, требуется поиск путей, ограничивающих требуемую для передачи полосу пропускания 3D-голографической информации, эквивалентных сжатию информации без потерь. Один из способов изложен в патенте РФ № 2707582 [11].

Суть способа заключается в отказе от передачи по радиоканалу всей 3D-голографической информации, а передаче только двух 2D-видеопотоков с текстурой поверхности голографируемого объекта и картой глубин. Синтез цифровой голограммы при этом производится только на приемном конце канала связи. Основные проблемы создания таких изображений описаны в [19, 20], где теоретически доказана возможность сжатия 3D-голографической информации методом, близким по смыслу к SSB. В [19] приведен численный эксперимент, показывающий, что восстановленное такой голограммой на приемном конце канала связи изображение дает на порядок лучшее разрешение, чем при спектральном сжатии голограмм. В [20] показано распространение метода на инфракрасный, ультрафиолетовый и радиодиапазон электромагнитного спектра. Таким образом, для завершения первого этапа работ сжатия голографической информации требуется экспериментальное подтверждение возможности передачи представленной 3D-голографической видеoinформации о 3D-объекте голографирования и качественного восстановления 3D-изображения живого объекта голографирования с ТВ-частотой кадровой развертки.

Методика эксперимента по передаче 3D-голографической информации, синтеза голограммы и восстановления ее 3D-изображения с непрерывным параллаксом

Для задач дополненной реальности подходит широко распространенная и активно развивающаяся сегодня технология беспроводной передачи данных Wi-Fi. В настоящей работе исследована возможность обеспечения передачи видеопотока 3D-голографической информации по беспроводному каналу связи (Wi-Fi), с кадровой частотой больше 25 кадров/с.

С помощью протокола File Transfer Protocol (FTP) выполнен эксперимент по передаче голографической

информации по Wi-Fi с помощью двух основных модальностей 3D-изображения — текстуры и карты глубин 3D-сцены (рис. 1), которые служат основой 3D-модели объекта при синтезе голограмм на приемном конце канала связи [11].

Для передачи 3D-видеопотока, позволяющего сформировать на приемном конце канала связи голографические изображения, одновременно переданы пакеты по 291 двойному кадру. Измеренное программой «FileZilla» время передачи пакетов кадров при воспроизведении в реальном времени показало, что передача полной голографической информации о 3D-объекте в реальном времени с кадровой частотой большей, чем 25 кадров/с вполне осуществима.

В тестовом эксперименте использованы два устройства: персональный компьютер (ПК), с подключенным к нему через сетевой кабель (Ethernet) роутером (модель ELTEX NTE-RG-1421G-Wac rev. A1), и мобильное устройство на базе операционной системы (ОС) Android (модель Xiaomi Mi A1) со встроенным модулем Wi-Fi стандарта IEEE 802.11n.

Программное соединение между двумя устройствами установлено через протокол передачи файлов FTP. При таком типе соединения первое устройство является FTP-сервером, а второе — клиентом (подключается к серверу), при этом передача файлов может осуществляться в любом направлении.

На ПК использована программа «FileZilla», работающая с протоколом FTP, через нее будут отправляться команды по передаче данных. После передачи данных программа возвращает общее время отправки файла, на основе которого можно определить скорость передачи данных и кадровую частоту передачи.

На мобильном устройстве используется приложение ES-проводник (менеджер файлов — проводник), одной из функций которого является создание и работа с локальным FTP-сервером (рис. 2).

В качестве голографической информации, т. е. предварительных 3D-кадров для «голографического видео» использована серия из 291 монохромного кадра с 3D-портретом человека, время съемки зависит от выбранной аппаратуры и поэтому не учитывалось. Кадры текстуры и карты глубин поверхности сформированы

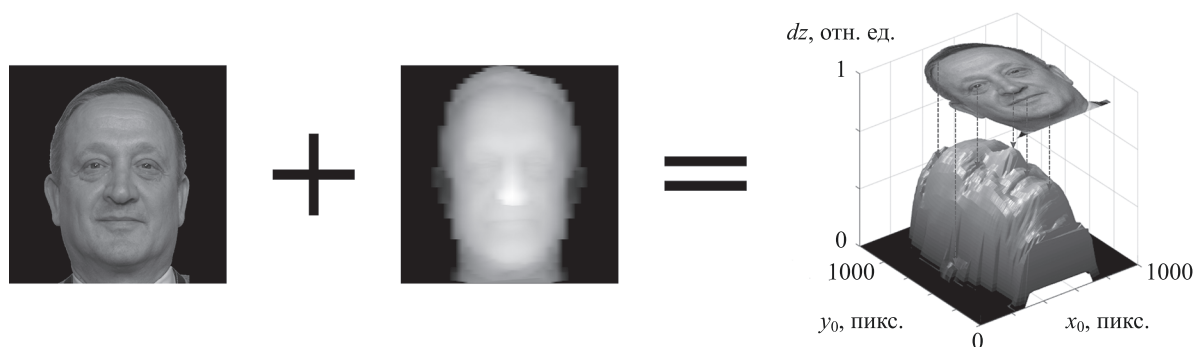


Рис. 1. Представление 3D-объекта в виде двух основных модальностей 3D-изображения — текстуры поверхности и карты глубин: x_0 и y_0 — ширина и высота поля основания 3D-изображения (в пикселях); dz — глубина 3D-изображения (на рисунке показана в относительных единицах)

Fig. 1. Representation of a 3D object in the form of two main 3D image modalities — surface textures and depth maps: x_0 and y_0 are the width and height of the 3D image base field (in pixels); dz — depth of the 3D image (shown in relative units in the figure)

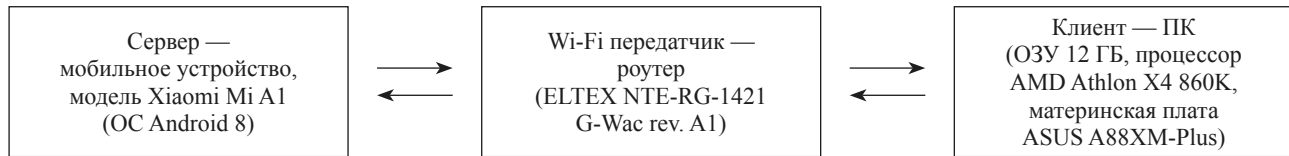


Рис. 2. Блок-схема устройства передачи информации

Fig. 2. Flowchart of the information transmission device

рованы с применением структурированного излучения аналогично с методикой [20], но от динамически меняющегося объекта. Фотографии некоторых кадров, из передаваемой серии кадров текстуры поверхности объекта и 3D-карты поверхности, приведены на рис. 3. Разрешение текстуры от 500×500 до 2000×2000 пикселей, и разные комбинации значений информационной емкости 3D-карты поверхности.

Результаты экспериментальных измерений

Результаты измерений времени передачи пакетов обработаны и собраны в таблице, в третьем столбце которой приведены форматы передаваемых кадров. Ввиду специфики применяемых программ преобразования Френеля, использованного при синтезе голограмм, связанной с аппаратом быстрого преобразования Фурье, на этапе эксперимента использовано представление кадров в виде квадратных матриц. Это привело к завышенным требованиям к пропускной способности канала, что

позволяет рассчитывать на передачу больших размеров кадров в дальнейшем. Видно, что разрешение передаваемых пар кадров текстуры и карты глубин изменяется от 500×500 пикселей, что близко к устаревшим ТВ-стандартам PAL и SECAM (720×576 пикселей), до 2000×2000 пикселей, что превышает разрешение в стандартах высокой четкости HD (1280×720 пикселей) и Full HD (1920×1080 пикселей). Выбранные форматы кадров соответствуют наиболее распространенным сегодня форматам с разной степенью сжатия. От метода синтеза голограммы количество передаваемой по каналу связи информации не зависит, и выбранная модель голограмм Лейта–Упатниекса не ограничивает общность полученных результатов. Таким образом, в эксперименте исследованы разные объемы пакетов передаваемых данных, что показало возможности метода для разных режимов передачи.

В таблице приведены: форматы файлов, в которые упаковываются передаваемые данные; общее количество передаваемой информации; указано измеренное

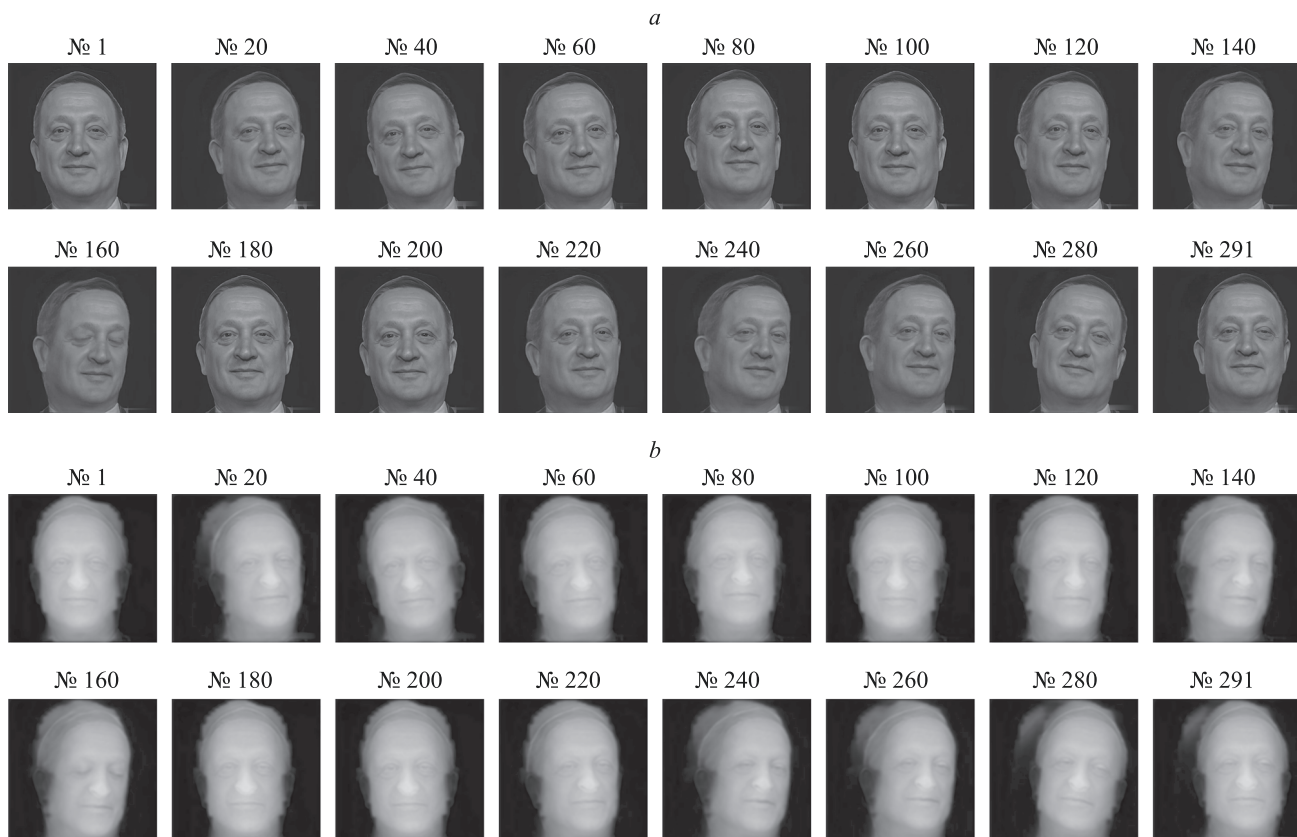


Рис. 3. Фотографии некоторых кадров из передаваемой серии, состоящей из 291 кадра 3D-видео: динамически изменяющейся текстуры поверхности объекта (a) и соответствующей кадрам текстуры карты поверхности (b)

Fig. 3. Photos of some frames from the transmitted series consisting of 291 frames of 3D video: dynamically changing texture of the surface of the object (a) and corresponding to the frames of the texture of the surface map (b)

Таблица. Время передачи пакетов и частота кадровой развертки

Table. Packet transmission time and frame rate

Номер эксперимента	Разрешение, пиксели (столбцы × строки)	Формат кадров	Объем пакета, МБ	Время передачи пакета при точности измерений 1 с, с	Вычисленная скорость передачи пакета, МБ/с	Кадровая частота при вычисленной скорости передачи, кадр/с	Кадровая частота при средней скорости передачи 11 МБ/с, кадр/с
1	500 × 500	BMP	139,4	13	10,72	22,39	22,96
2		PNG	28,4	3	9,46	97,00	112,71
3		JPG	6,7	1	6,70	291,00	477,76
4	1000 × 1000	BMP	555,6	51	10,89	5,71	5,76
5		PNG	66,3	5	13,23	58,20	48,28
6		JPG	18,6	1	18,60	291,00	172,09
7	2000 × 2000	BMP	2220,8	—	—	—	1,45
8		PNG	146,2	12	12,18	24,25	21,90
9		JPG	53,3	5	10,66	58,20	60,06

время передачи соответствующих пар кадров (прямые данные из программы «FileZilla»); скорость передачи данных каждого пакета; кадровая частота, вычисленная по скорости передачи каждого из экспериментов.

Видно, что в экспериментах № 2–3, 5–6 частота кадровой развертки оказалась больше 25 кадров/с. Это означает, что, например, для стандарта VGA и HD можно передавать 3D-голографическую информацию в формате PNG и JPEG без каких-либо затруднений. Для стандартов Full HD, 2K и выше (эксперименты № 8–9) — передача в формате PNG с частотой больше 25 кадров/с может быть затруднена, однако, в формате JPEG стандартного коэффициента качества 70 %, кадровая частота передачи по-прежнему остается в допустимых пределах.

Обсуждение результатов эксперимента

При использовании некоторых дополнительных мер скорость передачи может быть значительно увеличена, что, как следствие, повысит кадровую частоту передачи. Например, это возможно при передаче не квадратного кадра размером 2000 × 2000 пикселей, а вписанного в него и полностью соответствующего стандарту Full HD-кадра размером 1920 × 1080 пикселей. Это вполне возможно, и будет сделано при разработке прототипа установки голографического ТВ. Все варианты, приведенные в таблице, могут быть применены в устройствах 3D-дополненной реальности, как например: в телемедицине, системах дистанционного управления сложными объектами, включая системы безопасности, 3D-фототелеграфе и др.

По переданным по каналу связи 3D-голографическим изображениям выполнен синтез цифровых голограмм аналогично [20–23], по виртуально моделируемому излучению, повторяющему распространение в схеме Лейта–Упатниекса. Этими голограммами в дальнейшем были восстановлены виртуальные 3D-голографические изображения живого динамически меняющегося

объекта, обладающие как вертикальным, так и горизонтальным параллаксом (рис. 4). Их материальные аналоги были реализованы ранее в технологии Dot Matrix [24, 25].

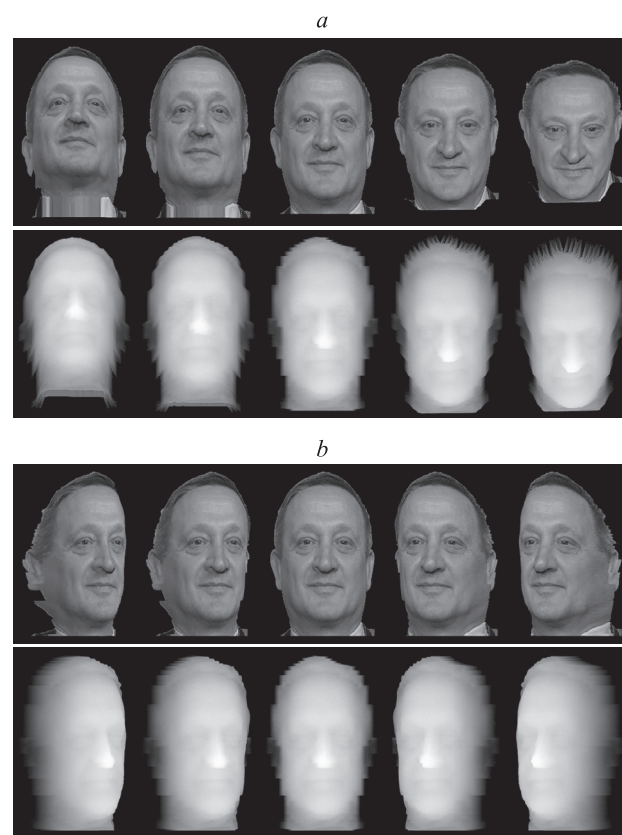


Рис. 4. Виртуальные 3D-голографические изображения с непрерывным параллаксом: с вертикальным (a) и горизонтальным (b)

Fig. 4. Virtual 3D holographic images with continuous parallax: vertical (a) and horizontal (b)

Заключение

С помощью протокола FTP проведен эксперимент передачи по беспроводному каналу связи Wi-Fi информации в виде двух 2D-кадров для синтеза голограмм, восстанавливающих с непрерывным параллаксом 3D-изображения с телевизионной частотой кадровой развертки, имитирующих 3D-видео с пространственным разрешением не хуже, чем в телевизионных стан-

дартах высокой четкости, как например Full HD. С учетом серии предыдущих работ [11, 19–25], представленный в настоящей работе их заключительный этап, в котором экспериментально доказана возможность передачи динамической 3D-голографической информации, показывает один из возможных путей реализации задачи создания голографического телевидения и 3D-дополненной реальности.

Литература

References

1. Денисюк Ю.Н. Об отображении оптических свойств объекта в волновом поле рассеянного им излучения // Оптика и спектроскопия. 1963. Т. 15. № 4. С. 522–532.
2. Leith E.N., Upatnieks J. Reconstructed wavefronts and communication theory // Journal of the Optical Society of America. 1962. V. 52. N 10. P. 1123–1130. <https://doi.org/10.1364/JOSA.52.001123>
3. Leith E.N., Upatnieks J. Wavefront reconstruction with diffused illumination and three-dimensional objects // Journal of the Optical Society of America. 1964. V. 54. N 11. P. 1295–1301. <https://doi.org/10.1364/JOSA.54.001295>
4. Van Heerden P.J. A new optical method of storing and retrieving information // Applied Optics. 1963. V. 2. N 4. P. 387–392. <https://doi.org/10.1364/AO.2.000387>
5. Hill B. Some aspects of a large capacity holographic memory // Applied Optics. 1972. V. 11. N 1. P. 182–191. <https://doi.org/10.1364/AO.11.000182>
6. Титарь В.П., Богданова Т.В. Проблемы создания голографической телевизионной системы // Радиоэлектроника и информатика. 1999. № 2(7). С. 38–42.
7. Денисюк Ю.Н. Достаточны ли известные фундаментальные принципы голографии для создания новых типов объемного кинематографа и искусственного интеллекта? // Журнал технической физики. 1991. Т. 61. № 8. С. 149–161.
8. Shannon C.E. Communication in the presence of noise // Proceedings of the IRE. 1949. V. 37. N 1. P. 10–21. <https://doi.org/10.1109/jrproc.1949.232969>
9. ГОСТ 24375-80. Радиосвязь. Термины и определения [Текст]: Дата введения 01.01.1980.
10. Комар В.Г. Информационная оценка качества изображения кинематографических систем // Техника кино и телевидения. 1971. № 10. С. 9–22.
11. Шойдин С.А. Способ дистанционного формирования голографической записи. Патент RU2707582C1. Бюл. 2019. № 34.
12. Naughton T.J., McDonald J.B., Javidi B. Efficient compression of Fresnel fields for internet transmission of three-dimensional images // Applied Optics. 2003. V. 42. N 23. P. 4758–4764. <https://doi.org/10.1364/AO.42.004758>
13. Yamaguchi I., Zhang T. Phase-shifting digital holography // Optics Letters. 1997. V. 22. N 16. P. 1268–1270. <https://doi.org/10.1364/OL.22.001268>
14. Blinder D., Ahar A., Bettens S., Birnbaum T., Symeonidou A., Ottevaere H., Schretter C., Schelkens P. Signal processing challenges for digital holographic video display systems // Signal Processing: Image Communication. 2019. V. 70. P. 114–130. <https://doi.org/10.1016/j.image.2018.09.014>
15. Bernardo M.V., Fernandes P., Arrifano A., Antonini M., Fonseca E., Fiadeiro P.T., Pinheiro A.M.G., Pereira M. Holographic representation: Hologram plane vs. object plane // Signal Processing: Image Communication. 2018. V. 68. P. 193–206. <https://doi.org/10.1016/j.image.2018.08.006>
16. Seo Y.-H., Choi H.-J., Kim D.-W. 3D scanning-based compression technique for digital hologram video // Signal Processing: Image Communication. 2007. V. 22. N 2. P. 144–156. <https://doi.org/10.1016/j.image.2006.11.007>
17. Naughton Th.J., McDonald J.B., Javidi B. Efficient compression of Fresnel fields for internet transmission of three-dimensional images // Applied Optics. 2003. V. 42. N 23. P. 4758–4764. <https://doi.org/10.1364/AO.42.004758>
18. Muhamad R.K., Birnbaum T., Gilles A., Mahmoudpour S., Oh K.-J., Pereira M., Perra C., Pinheiro A., Schelkens P. JPEG Pleno
1. Denisjuk Ju.N. Displaying the optical properties of an object in the wave field of radiation scattered by it. *Optika i spektroskopija*, 1963, vol. 15, no. 4, pp. 522–532. (in Russian)
2. Leith E.N., Upatnieks J. Reconstructed wavefronts and communication theory. *Journal of the Optical Society of America*, 1962, vol. 52, no. 10, pp. 1123–1130. <https://doi.org/10.1364/JOSA.52.001123>
3. Leith E.N., Upatnieks J. Wavefront reconstruction with diffused illumination and three-dimensional objects. *Journal of the Optical Society of America*, 1964, vol. 54, no. 11, pp. 1295–1301. <https://doi.org/10.1364/JOSA.54.001295>
4. Van Heerden P.J. A new optical method of storing and retrieving information. *Applied Optics*, 1963, vol. 2, no. 4, pp. 387–392. <https://doi.org/10.1364/AO.2.000387>
5. Hill B. Some aspects of a large capacity holographic memory. *Applied Optics*, 1972, vol. 11, no. 1, pp. 182–191. <https://doi.org/10.1364/AO.11.000182>
6. Titar V.P., Bogdanova T.V. Problems of creating a holographic TV system. *Radioelektronika i informatika*, 1999, no. 2(7), pp. 38–42. (in Russian)
7. Denisjuk Ju.N. Are the known fundamental principles of holography sufficient to create new types of 3D cinema and artificial intelligence? *Zhurnal tehniczeskoj fiziki*, 1991, vol. 61, no. 8, pp. 149–161. (in Russian)
8. Shannon C.E. Communication in the presence of noise. *Proceedings of the IRE*, 1949, vol. 37, no. 1, pp. 10–21. <https://doi.org/10.1109/jrproc.1949.232969>
9. GOST 24375-80. *Radio communication. Terms and Definitions*. 01.01.1980. (in Russian)
10. Komar V.G. Informational assessment of the image quality of cinematographic systems. *Tehnika kino i televidenija*, 1971, no. 10, pp. 9–22. (in Russian)
11. Shoidin S.A. Method of holographic recording remote formation. *Patent RU2707582C1*. 2019. (in Russian)
12. Naughton T.J., McDonald J.B., Javidi B. Efficient compression of Fresnel fields for internet transmission of three-dimensional images. *Applied Optics*, 2003, vol. 42, no. 23, pp. 4758–4764. <https://doi.org/10.1364/AO.42.004758>
13. Yamaguchi I., Zhang T. Phase-shifting digital holography. *Optics Letters*, 1997, vol. 22, no. 16, pp. 1268–1270. <https://doi.org/10.1364/OL.22.001268>
14. Blinder D., Ahar A., Bettens S., Birnbaum T., Symeonidou A., Ottevaere H., Schretter C., Schelkens P. Signal processing challenges for digital holographic video display systems. *Signal Processing: Image Communication*, 2019, vol. 70, pp. 114–130. <https://doi.org/10.1016/j.image.2018.09.014>
15. Bernardo M.V., Fernandes P., Arrifano A., Antonini M., Fonseca E., Fiadeiro P.T., Pinheiro A.M.G., Pereira M. Holographic representation: Hologram plane vs. object plane. *Signal Processing: Image Communication*, 2018, vol. 68, pp. 193–206. <https://doi.org/10.1016/j.image.2018.08.006>
16. Seo Y.-H., Choi H.-J., Kim D.-W. 3D scanning-based compression technique for digital hologram video. *Signal Processing: Image Communication*, 2007, vol. 22, no. 2, pp. 144–156. <https://doi.org/10.1016/j.image.2006.11.007>
17. Naughton Th.J., McDonald J.B., Javidi B. Efficient compression of Fresnel fields for internet transmission of three-dimensional images. *Applied Optics*, 2003, vol. 42, no. 23, pp. 4758–4764. <https://doi.org/10.1364/AO.42.004758>
18. Muhamad R.K., Birnbaum T., Gilles A., Mahmoudpour S., Oh K.-J., Pereira M., Perra C., Pinheiro A., Schelkens P. JPEG Pleno

- holography: scope and technology validation procedures // *Applied Optics*, 2021. V. 60. N 3. P. 641–651. <https://doi.org/10.1364/AO.404305>
19. Shoydin S.A., Pazoev A.L. Transmission of 3D holographic information via conventional communication channels and the possibility of multiplexing in the implementation of 3D hyperspectral images // *Photonics*, 2021. V. 8. N 10. P. 448–473. <https://doi.org/10.3390/photonics8100448>
 20. Shoydin S.A., Pazoev A.L. Remote formation of holographic record // *Optoelectronics, Instrumentation and Data Processing*, 2021. V. 57. N 1. P. 80–88. <https://doi.org/10.3103/S8756699021010118>
 21. Пазоев А.Л., Шойдин С.А. Передача голографической информации на одной боковой полосе // Интерэкспо ГЕО-Сибирь. 2021. Т. 8. С. 109–117. <https://doi.org/10.33764/2618-981X-2021-8-109-117>
 22. Шойдин С.А., Пазоев А.Л. Особенности синтеза SSB голограмм (SSBH) // *HoloExpo 2021: тезисы докладов XVII международной конференции по голографии и прикладным оптическим технологиям*. М.: МГТУ им. Н.Э. Баумана, 2021. С. 207–218.
 23. Шойдин С.А., Пазоев А.Л. Сжатие 3D голографической информации аналогично передаче информации на одной боковой полосе // *Оптический журнал*, 2022. Т. 89. № 3. С. 79–88. <https://doi.org/10.17586/1023-5086-2022-89-03-79-88>
 24. Шойдин С.А., Пазоев А.Л., Смык А.Ф., Шурыгин А.В. Синтезированные на приёмном конце канала связи голограммы 3D-объекта в технологии Dot Matrix // *Компьютерная оптика*, 2022. Т. 46. № 2. С. 204–213. <https://doi.org/10.18287/2412-6179-CO-1037>
 25. Шойдин С.А., Пазоев А.Л., Цыганов И.К., Дроздова Е.А. Запись голограммы, переданной по каналу связи на одной боковой полосе // *HoloExpo 2021: тезисы докладов XVII международной конференции по голографии и прикладным оптическим технологиям*. М.: МГТУ им. Н. Э. Баумана, 2021. С. 109–117.
 - holography: scope and technology validation procedures. *Applied Optics*, 2021, vol. 60, no. 3, pp. 641–651. <https://doi.org/10.1364/AO.404305>
 19. Shoydin S.A., Pazoev A.L. Transmission of 3D holographic information via conventional communication channels and the possibility of multiplexing in the implementation of 3D hyperspectral images. *Photonics*, 2021, vol. 8, no. 10, pp. 448–473. <https://doi.org/10.3390/photonics8100448>
 20. Shoydin S.A., Pazoev A.L. Remote formation of holographic record. *Optoelectronics, Instrumentation and Data Processing*, 2021, vol. 57, no. 1, pp. 80–88. <https://doi.org/10.3103/S8756699021010118>
 21. Pazoev A.L., Shoydin S.A. Transmission of holographic information on a single sideband. *Interexpo GEO-Siberia*, 2021, vol. 8, pp. 109–117. (in Russian). <https://doi.org/10.33764/2618-981X-2021-8-109-117>
 22. Shoydin S.A., Pazoev A.L. Features of the SSB holograms synthesis (SSBH). *HoloExpo 2021: Abstracts of the XVII International Conference on Holography*. Moscow, Bauman Moscow State Technical University, 2021, pp. 207–218. (in Russian)
 23. Shoydin S.A., Pazoev A.L. Compressing 3D holographic information similar to data transmission via a single sideband. *Journal of Optical Technology*, 2022, vol. 89, no. 3, pp. 176–182. <https://doi.org/10.1364/JOT.89.000176>
 24. Shoydin S.A., Pazoev A.L., Smyk A.F., Shurygin A.V. 3D object holograms synthesized in Dot Matrix technology at the receiving end of the communication channel. *Computer Optics*, 2022, vol. 46, no. 2, pp. 204–213. (in Russian). <https://doi.org/10.18287/2412-6179-CO-1037>
 25. Shoydin S.A., Pazoev A.L., Tsyganov I.K., Drozdova E.A. Recording of a hologram transmitted over a single sideband communication channel. *HoloExpo 2021: Abstracts of the XVII International Conference on Holography*. Moscow, Bauman Moscow State Technical University, 2021, pp. 109–117. (in Russian)

Авторы

Пазоев Артём Леонович — преподаватель, Сибирский государственный университет геосистем и технологий, Новосибирск, 630108, Российская Федерация, [sc 57213689704](https://orcid.org/0000-0003-0302-4860), <https://orcid.org/0000-0003-0302-4860>, pazoev-al2018@sgugit.ru

Шойдин Сергей Александрович — кандидат физико-математических наук, доцент, доцент, Сибирский государственный университет геосистем и технологий, Новосибирск, 630108, Российская Федерация, [sc 56022582500](https://orcid.org/0000-0002-2186-7928), <https://orcid.org/0000-0002-2186-7928>, shoydin@snga.ru

Статья поступила в редакцию 17.08.2022
Одобрена после рецензирования 10.12.2022
Принята к печати 30.01.2023

Authors

Artem L. Pazoev — Lecturer, Siberian State University of Geosystems and Technologies, Novosibirsk, 630108, Russian Federation, [sc 57213689704](https://orcid.org/0000-0003-0302-4860), <https://orcid.org/0000-0003-0302-4860>, pazoev-al2018@sgugit.ru

Sergey A. Shoydin — PhD (Physics & Mathematics), Associate Professor, Associate Professor, Siberian State University of Geosystems and Technologies, Novosibirsk, 630108, Russian Federation, [sc 56022582500](https://orcid.org/0000-0002-2186-7928), <https://orcid.org/0000-0002-2186-7928>, shoydin@snga.ru

Received 17.08.2022
Approved after reviewing 10.12.2022
Accepted 30.01.2023



Работа доступна по лицензии
Creative Commons
«Attribution-NonCommercial»

НАНОСТРУКТУРЫ. СВОЙСТВА И ПРИМЕНЕНИЯ NANOSTRUCTURES. PROPERTIES AND APPLICATIONS

doi: 10.17586/2226-1494-2023-23-1-28-34

УДК 535. 016

Влияние параметров анодирования на оптические свойства нанопористого оксида алюминия, сформированного поверх островковой пленки серебра

Игорь Юрьевич Никитин¹✉, Резида Даниловна Набиуллина²,
Алексей Викторович Нашекин³, Антон Андреевич Старовойтов⁴,
Игорь Аркадьевич Гладских⁵

^{1,2,4,5} Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация

³ Физико-технический институт им. А.Ф. Иоффе РАН, Санкт-Петербург, 194021, Российская Федерация

¹ nikitin0igor512@gmail.com ✉, <https://orcid.org/0000-0002-7911-9092>

² rezida2105@mail.ru, <https://orcid.org/0000-0002-1891-3057>

³ nashchekin@mail.ioffe.ru, <https://orcid.org/0000-0002-2542-7364>

⁴ Anton.starovoytov@gmail.com, <https://orcid.org/0000-0002-2160-1281>

⁵ 138020@mail.ru, <https://orcid.org/0000-0001-6389-1611>

Аннотация

Предмет исследования. Исследованы оптические свойства гибридной плазмонной тонкопленочной наноструктуры. Структура представляет собой пористую матрицу анодного оксида алюминия на поверхности островковой пленки серебра, сформированной на кварцевой подложке. **Метод.** Серебряная пленка из наночастиц, лежащая в основе структуры, создана методом термического напыления. Островковая структура серебра с размером наночастиц порядка 100 нм сформирована в процессе отжига на воздухе. На поверхность структуры методом электронно-лучевого испарения в вакууме нанесена пленка алюминия. Выполнено одноступенчатое прямое анодирование пленки алюминия, в результате которого образована нанопористая пленка оксида алюминия. Полученные структуры исследованы методами спектрометрии и электронной микроскопии. **Основные результаты.** Рассмотрены спектры отражения и оптической плотности наноструктур, полученные при различных продолжительности и токе анодирования. Выполнено сравнение полученных спектров наноструктур для оксида алюминия, серебряных наночастиц и оксида алюминия на поверхности серебряных наночастиц. Показано, что с увеличением времени анодирования, коэффициент отражения структуры приближается к величине показателя для анодного оксида алюминия, вследствие окисления алюминиевой пленки. Увеличение тока анодирования приводит к сдвигу минимума коэффициента отражения в длинноволновую область спектра. Это явление подтверждается ранее выполненными исследованиями и объясняется увеличением расстояния между порами, образующимися в результате анодирования. Численное моделирование спектров сечения экстинкции с использованием Mie калькулятора показало, что результаты моделирования оптических свойств структуры с размерами наночастицы порядка 100 нм сопоставимы со спектрами оптической плотности, полученными в результате эксперимента. При моделировании использовано приближение сферической частицы. Для более точного определения толщины оксидной пленки и приближения расчетных спектров оптической плотности к экспериментальным результатам требуется учет форм-фактора. **Практическая значимость.** Результаты работы могут быть использованы при проектировании и изготовлении сенсоров, оптических покрытий и источников фотонов для применения в экранах, оптических схемах и устройствах плазмоники.

Ключевые слова

локализованный плазмонный резонанс, нанопористый оксид алюминия, тонкие пленки, наночастицы, анодирование

Благодарности

Исследование выполнено за счет средств гранта Российского научного фонда № 21-72-10098, а также средств гранта поддержки научно-исследовательских работ бакалавров, магистров и аспирантов (НИР МиА) Университета ИТМО.

Ссылка для цитирования: Никитин И.Ю., Набиуллина Р.Д., Нащекин А.В., Старовойтов А.А., Гладских И.А. Влияние параметров анодирования на оптические свойства нанопористого оксида алюминия, сформированного поверх островковой пленки серебра // Научно-технический вестник информационных технологий, механики и оптики. 2023. Т. 23, № 1. С. 28–34. doi: 10.17586/2226-1494-2023-23-1-28-34

Anodization parameters influence on anodic aluminum oxide formed above the silver island film

Igor Yu. Nikitin^{1✉}, Rezida D. Nabiullina², Alexey V. Nashyokin³,
Anton A. Starovoytov⁴, Igor A. Gladskikh⁵

^{1,2,4,5} ITMO University Saint Petersburg, 197101, Russian Federation

³ Ioffe Institute Saint Petersburg, 194021, Russian Federation

¹ nikitin0igor512@gmail.com✉, <https://orcid.org/0000-0002-7911-9092>

² rezida2105@mail.ru, <https://orcid.org/0000-0002-1891-3057>

³ nashchekin@mail.ioffe.ru, <https://orcid.org/0000-0002-2542-7364>

⁴ Anton.starovoytov@gmail.com, <https://orcid.org/0000-0002-2160-1281>

⁵ 138020@mail.ru, <https://orcid.org/0000-0001-6389-1611>

Abstract

Optical properties of hybrid plasmonic thin film structure in the form of anodic alumina porous matrix above silver island film on the quartz substrate have been investigated. Silver nanoparticle film in the bottom of the structure has been obtained by physical vapor deposition in a vacuum chamber. The silver island film with the island of average diameter of 100 nm has been formed after annealing in the air atmosphere. Above the silver nanoparticle film an aluminum film has been deposited by the E-beam evaporation. As a result of one-step straight anodization, a nanoporous alumina thin film has been formed. The obtained structures were investigated, using spectroscopy and electron microscopy methods. The structure reflectance and optical density spectra have been obtained and analyzed for different anodization times and currents. To compare the results, the reflectance and optical density spectra have been obtained for silver nanoparticles and anodic alumina. When anodization times are increased, structure reflection coefficient spectra become more like the same characteristic for anodic aluminum oxide, which can be explained by film oxidation. At the same time a red shift of reflectance spectrums is observed in the structures with bigger maximum anodization currents. This effect has been observed in other works and can be explained by the increasing distance between the pores. A numerical modeling of optical properties with the help of Mie calculator for the structure with the nanoparticle size of 100 nm has shown that the results of the modeling can be compared to the experimentally obtained optical density spectra. The modeling was performed using spherical approximation. To obtain more precise results for alumina film thickness and nanoparticle optical properties, a silver nanoparticle form factor has to be considered. The results of this work can be used in sensors, optical coatings and photon sources fabrication methods. These can be used in screens, optical schemes and many other plasmonic devices

Keywords

localized plasmon resonance, nanoporous alumina, thin films, nanoparticles, anodization

Acknowledgements

The research was financed by the Russian Science Foundation grant no. 21-72-10098 and by ITMO University science research bachelor, master and PhD student supporting grant.

For citation: Nikitin I.Yu., Nabiullina R.D., Nashyokin F.V., Starovoytov A.A., Gladskikh I.A. Anodization parameters influence on anodic aluminum oxide formed above the silver island film. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 2023, vol. 23, no. 1, pp. 28–34 (in Russian). doi: 10.17586/2226-1494-2023-23-1-28-34

Введение

С совершенствованием устройств плазмоники исследование оптических свойств гибридных плазмонных наноструктур на основе пористых матриц и металлических наночастиц оказывается актуальным. Вследствие усиленного поглощения и рассеяния света [1], а также способности к увеличению интенсивности собственного излучения объектов, помещенных в ячейки матрицы [2,3], такие структуры могут применяться в качестве источников одиночных фотонов в квантовых коммуникациях и фотонных компьютерах [4], усилителей поглощения в солнечных батареях [5], а также различного рода сенсоров и датчиков.

В последнее время создано большое количество образцов гибридных плазмонных наноструктур, на основе металлических пленок и нанопористых ма-

триц, полученных в результате анодирования [6–8]. Под нанопористой матрицей в данном случае рассматривается массив из квазицилиндрических несквозных отверстий, диаметром 11–37 нм в пленке оксида алюминия, образовавшийся в результате процесса анодирования, принципы которого подробно описаны в работах [9–12]. Глубина и диаметр отверстий зависят от условий анодирования. В настоящей работе предложен новый подход к созданию гибридных плазмонных наноструктур на основе нанопористых матриц, при котором металлические наночастицы располагаются в основании пор анодированного алюминия. Данная компоновка имеет существенные достоинства. Так, например, ближние поля металлических наночастиц, образующиеся при возбуждении локализованного плазмонного резонанса, фокусируются вблизи барьерного слоя диэлектрических пор оксида, где обычно оседают

исследуемые наноразмерные объекты, помещаемые в матрицу. К таким объектам могут относиться квантовые точки и красители [13, 14].

В работе рассмотрен метод получения гибридной структуры на основе островковой пленки серебра и тонкой пленки нанопористой матрицы оксида алюминия, а также влияние параметров анодирования на оптические свойства конечной гибридной структуры. Проведено сравнение спектров отражения и оптической плотности для серебряных наночастиц, тонкой пленки оксида алюминия и гибридной наноструктуры. Выполнено численное моделирование с целью определения возможных размеров серебряных наночастиц и толщины оксидной пленки на их поверхности в сферическом приближении. Рассмотрены спектры отражения в зависимости от времени и максимального тока анодирования.

Материалы и методы

Получение образцов гибридной плазмонной структуры включает следующие этапы.

Подготовка кварцевых подложек — обезжиривание и очищение в изопропанол и деионизованной воде. После высыхания — размещение подложек в вакуумной камере PVD-75 (Kurt J. Lesker Ltd.), где методом осаждения паров из газовой фазы напыляется пленка серебра (99,99 %) толщиной 15 нм [15]. Проведение отжига образцов на воздухе при температуре 200 °C в течение 1 ч для формирования отдельных островков серебра, благодаря явлению поверхностной диффузии. Повторное размещение подложек в вакуумную камеру и напыление пленки алюминия (99,999 %) эквивалентной толщины 100 и 160 нм. Испарение серебра и алюминия осуществляется при давлении остаточных паров порядка 10^{-6} торр резистивным нагревом и пучком высокоэнергетических электронов соответственно. После напыления алюминия осуществляется одноступенчатое прямое анодирование образцов в 0,3 М растворе щавелевой кислоты при постоянном напряжении 40 В [10, 12]. Перед анодированием каждого образца на блоке питания наблюдались изменения значений тока. В качестве оптимальных параметров структуры с толщиной алюминия 160 нм выбраны параметры анодирования 45 с при 0,15 А максимального тока анодирования, которые использованы при сравнении гибрид-

ной структуры с пористой матрицей без серебряных наночастиц и островковой пленки серебра. Данные для экспериментов по варьированию параметров анодирования представлены в таблице.

Эквивалентная толщина характеризует количество материала, которое может содержаться в плоской тонкой пленке материала. В данной работе эквивалентная толщина измерена при помощи измерительного блока установки Sigma 242 (Sigma Instruments) резонансно-частотным методом.

После каждого анодирования образцы структуры вместе с контактами были несколько раз промыты деионизованной водой и высушены в вытяжном шкафу.

На каждом этапе эксперимента получены спектры оптической плотности образцов на спектрофотометре СФ-56 (ЛОМО) и спектры отражения на спектрометре РМА-12 (Hamamatsu) при угле падения 45°. Морфология образцов исследована с помощью сканирующего электронного микроскопа (СЭМ) (JEOL).

Результаты и обсуждение

В результате эксперимента получена гибридная тонкопленочная структура на основе пористой матрицы оксида алюминия поверх серебряной островковой пленки серебра (Ag). На рис. 1 показаны схематичные изображения серебряной островковой пленки из наночастиц и вид структуры, а также их СЭМ-изображения.

В основании пористой матрицы обнаружены металлические наночастицы диаметром более 100 нм, покрытые пленкой анодированного алюминия. Спектры оптической плотности и отражения серебряных наночастиц анодированного алюминия (эквивалентная толщина алюминия 160 нм, время анодирования — 45 с) без наночастиц и гибридной структуры (эквивалентные толщины серебра — 15 нм и алюминия — 160 нм, время анодирования — 45 с) представлены на рис. 2. В спектрах оптической плотности и отражения серебряной пленки наблюдается максимум на длине волны около 600 нм, связанный с возбуждением плазмонного резонанса серебряных наночастиц [16, 17]. Большая ширина спектра вызвана широким распределением частиц по формам и размерам в процессе самоорганизации при напылении и отжиге пленки.

Спектры оптической плотности и отражения анодированного алюминия монотонно возрастают с увеличе-

Таблица. Данные экспериментов по вариации параметров анодирования

Table. Anodization parameters variation experiments data

Эксперимент	Эквивалентная толщина, нм		Время анодирования, с	Максимальный ток анодирования, А
	серебра	алюминия		
Вариация времени анодирования	15,4 ± 0,2	100,4 ± 12,9	40	0,150
			60	0,150
			50	0,150
Вариация максимального тока анодирования			40	0,150
			40	0,234
			40	0,402

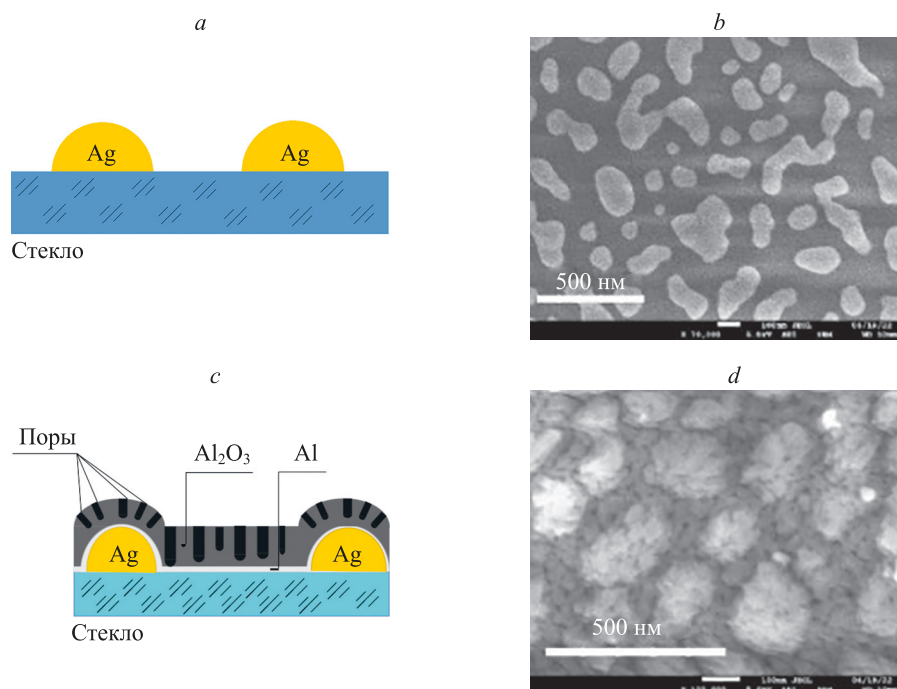


Рис. 1. Схематичное изображение (a) и СЭМ-изображение (b) серебряной островковой пленки; схема (c) и СЭМ-изображение (d) гибридной плазмонной оксид-алюминиево-серебряной структуры.

Цветовые обозначения: темно-серый — анодный оксид алюминия (Al₂O₃); черный — поры; светло-серый — алюминий (Al); желтый — серебряные наночастицы (Ag); синий — кварцевая подложка

Fig. 1. Suggested scheme for a silver island film (a). SEM Image of the silver island film (b). Suggested scheme of hybrid plasmonic alumina-silver structure. Dark-grey — anodic alumina, black — pores, light-grey — aluminum, yellow — silver nanoparticles, blue — quartz substrate (c). Hybrid structure SEM image (d)

нием длины волны, что может быть связано с наличием остаточной пленки алюминия в процессе анодирования. Большая величина оптической плотности может быть также связана с сильным рассеянием света на пористой структуре анодированного алюминия.

В результате видно, что форма спектров отражения и оптической плотности схожи для чистой пленки серебра и анодированного алюминия соответственно. Для гибридной структуры спектр оптической плотности больше похож на спектр серебряной пленки, со сме-

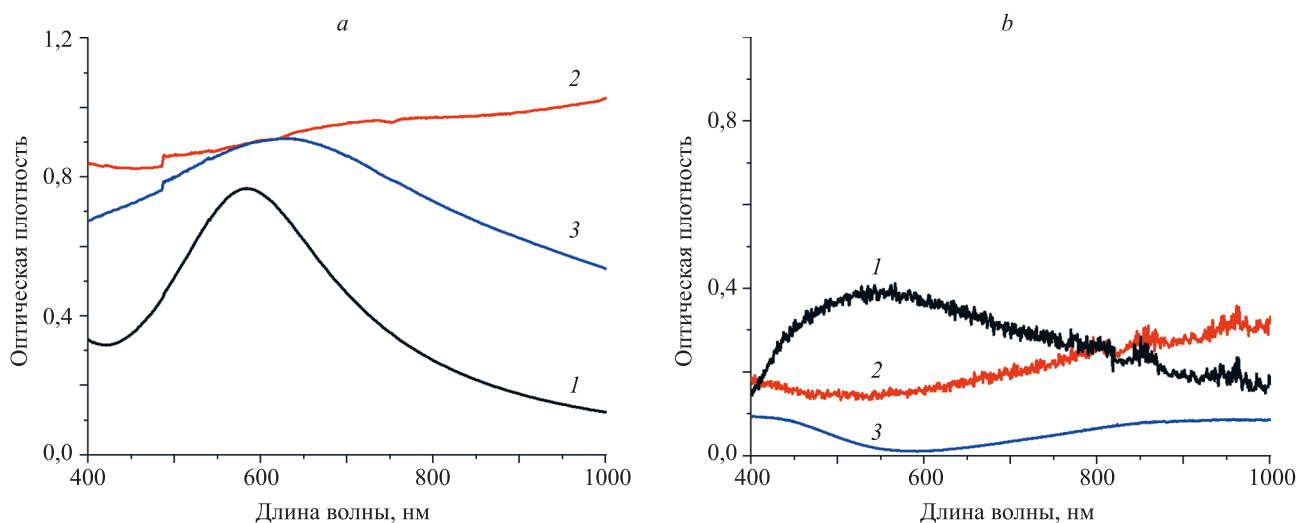


Рис. 2. Спектры оптической плотности (a) и отражения (b) серебряных наночастиц (кривая 1), анодного оксида алюминия (кривая 2) и гибридной наноструктуры (кривая 3)

Fig. 2. Optical density (a) and reflection (b) spectra of silver nanoparticles (curve 1), anodic alumina (curve 2), hybrid nanostructure (curve 3)

щенным в длинноволновую область спектра плазмонным резонансом. Одновременно в спектре отражения на длине волны плазмонного резонанса наблюдается минимум, а форма спектра близка к анодированному алюминию, с меньшим по величине отражением.

Данные результаты можно объяснить следующим образом. Известно, что основной вклад в экстинкцию крупных металлических наночастиц вносит рассеяние на них света, что демонстрируют спектры отражения. Для анодированного оксида алюминия отражение будет как от самого оксида алюминия, так и от поверхности подложки. Отметим, что большой вклад в отражение может вносить остаточная тонкая пленка чистого алюминия, а в спектр оптической плотности — рассеяние света на неоднородной пористой структуре анодированного алюминия.

Сдвиг максимума оптической плотности в длинноволновую область спектра может быть связан с наличием пленки оксида алюминия на поверхности серебряных наночастиц. Численное моделирование при помощи Мие калькулятора¹ показало, что на поверхности частиц может присутствовать пленка оксида алюминия толщиной до 10 нм. Радиус при моделировании равен 100 нм. Фактический радиус частицы может отличаться, так как форма частиц отлична от сферической. В качестве данных показателя преломления были взяты данные Ag_McPeak² и Al₂O₃ Query³. Сравнение спектров оптической плотности образца и спектра сечения экстинкции, полученного в результате численного моделирования, представлено на рис. 3.

На рис. 3 видно, что максимумы сечения экстинкции и оптической плотности находятся на длине волны 630 нм. В то время, как максимум оптической плотности серебряных наночастиц (рис. 2) — на длине волны 580 нм. Длинноволновый сдвиг вызывается оксидной пленкой на поверхности частиц.

Спектры отражения гибридных образцов в зависимости от времени анодирования показаны на рис. 4, а. Характер изменений отражения гибридной пленки при увеличении максимального тока анодирования может быть проанализирован, исходя из спектров на рис. 4, б.

Заметим, что с увеличением времени анодирования, отражение уменьшается, что соответствует уменьшению слоя чистого алюминия при анодировании в щавелевой кислоте. Электролит, в котором происходит анодирование, — агрессивная среда, которая влияет на размер пор. Чем больше и глубже поры, тем больше света через них проходит, и тем меньше света рассеивается и отражается назад. Видно, что отражение и рассеяние света существенно в длинноволновой области, где у алюминия повышенное отражение. Это

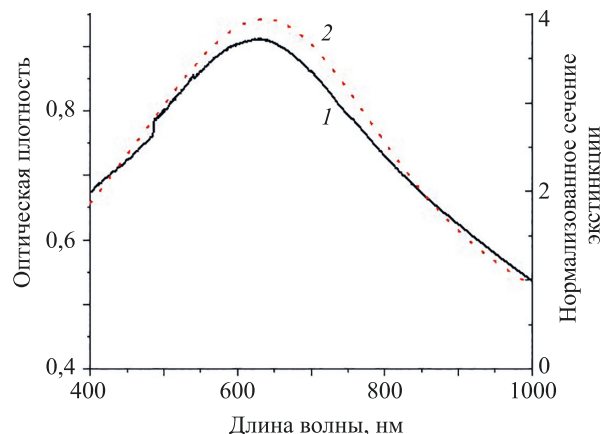


Рис. 3. Сравнение спектров оптической плотности гибридной пленки (кривая 1) и сечения экстинкции (кривая 2), полученного при численном моделировании

Fig. 3. Comparison of the optical density spectra of the hybrid film (curve 1) and the extinction cross section (curve 2) obtained by numerical simulation

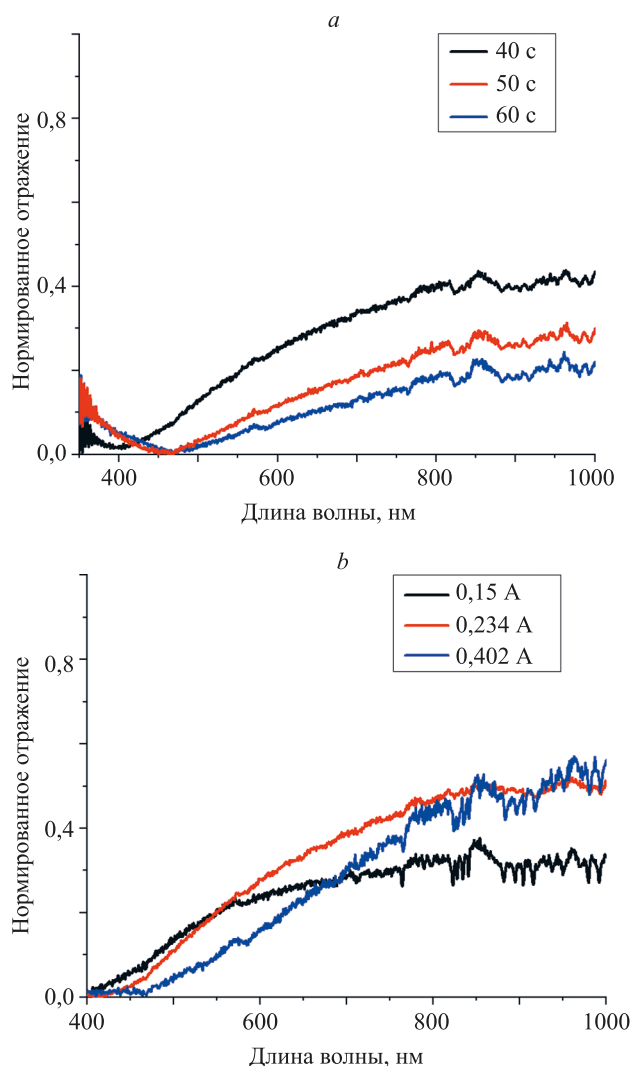


Рис. 4. Спектры отражения гибридных структур при различной длительности (а) и максимальном значении тока анодирования (б)

Fig. 4. Reflection spectra of hybrid structures at different duration (a) and maximum anodization current value (b)

¹ Мие калькулятор | Новый физтех. Университет ИТМО [Электронный ресурс]. URL: <https://physics.itmo.ru/ru/mie#/spectrum> (дата обращения: 26.07.2022).

² Refractive index of Ag (Silver) — McPeak [Электронный ресурс]. URL: <https://refractiveindex.info/?shelf=main&book=Ag&page=McPeak> (дата обращения: 10.12.2022).

³ Refractive index of Al₂O₃ (Aluminium sesquioxide, Sapphire, Alumina) — Query [Электронный ресурс]. URL: <https://refractiveindex.info/?shelf=main&book=Al2O3&page=Query> (дата обращения: 10.12.2022).

может указывать на присутствие в образцах остаточной алюминиевой пленки.

При изменении тока анодирования изменяется форма спектра. Минимум отражения начинает сдвигаться в длинноволновую область. Известно, что при увеличении потенциала, а также тока анодирования толстых пленок оксида алюминия, увеличивается дистанция между порами [11]. При увеличении потенциала или тока анодирования, некоторые поры перегреваются вследствие локального выделения джоулева тепла и разрушаются. В результате этого дистанция между уцелевшими порами увеличивается, что может влиять на длинноволновый сдвиг минимума поглощения [9].

Заключение

В работе созданы гибридные наноструктуры на основе пористой матрицы анодного оксида алюминия на поверхности островковой пленки серебра. Представляют интерес малые значения коэффициента отражения на длине волны максимума оптической плотности наночастиц. Этот феномен требует дальнейшего исследования, так как для обыкновенных серебряных наночастиц или чистой пленки оксида алюминия пониженного коэффициента отражения не наблюдается.

Длинноволновый сдвиг максимумов оптических плотностей гибридной структуры и серебряных наночастиц относительно друг друга может быть объяснен присутствием тонкой пленки оксида алюминия толщиной до 10 нм на поверхности серебряных наночастиц. Данное предположение подтверждено численным моделированием. Отметим, что при построении модели частица считалась сферической, поэтому необходимо учесть форм-фактор и факт присутствия подложки для последующих моделей. Также это связано с тем, что форма влияет на спектр оптической плотности частиц. В зависимости от тока и времени анодирования спектры отражения и оптической плотности наноструктур изменились. При увеличении тока анодирования произошел сдвиг минимума отражения в длинноволновую область, что может быть связано с изменением дистанции между порами. С увеличением времени анодирования значение коэффициента отражения уменьшалось, что может быть связано с увеличением пропускания материала из-за увеличения размеров и глубины пор анодного оксида алюминия. В дальнейшем следует произвести опыты с большим количеством образцов для увеличения статистической выборки. Данная структура имеет потенциал для использования в сенсорике, светодиодах и оптических покрытиях.

Литература

1. Maoz B.M., Chaikin Y., Tesler A.B., Bar Elli O., Fan Z., Govorov A.O., Markovich G. Amplification of chiroptical activity of chiral biomolecules by surface plasmons // *Nano Letters*. 2013. V. 13. N 3. P. 1203–1209. <https://doi.org/10.1021/nl304638a>
2. Xu S., Cao Y., Zhou J., Wang X., Wang X., Xu W. Plasmonic enhancement of fluorescence on silver nanoparticle films // *Nanotechnology*. 2011. V. 22. N 27. P. 275715. <https://doi.org/10.1088/0957-4484/22/27/275715>
3. Choudhari K.S., Kulkarni S.D., Santhosh C., George S.D. Photoluminescence enhancement and morphological properties of nanoporous anodic alumina prepared in oxalic acid with varying time and temperature // *Microporous and Mesoporous Materials*. 2018. V. 271. P. 138–145. <https://doi.org/10.1016/j.micromeso.2018.06.004>
4. Zhou Z.K., Lei D.Y., Liu J., Liu X., Xue J., Zhu Q., Chen H., Liu T., Li Y., Zhang H., Wang X. Shaping the emission spectral profile of quantum dots with periodic dielectric and metallic nanostructures // *Advanced Optical Materials*. 2014. V. 2. N 1. P. 56–64. <https://doi.org/10.1002/adom.201300354>
5. Ho W.J., Cheng P.Y., Hsiao K.Y. Plasmonic silicon solar cell based on silver nanoparticles using ultra-thin anodic aluminum oxide template // *Applied Surface Science*. 2015. V. 354. P. 25–30. <https://doi.org/10.1016/j.apsusc.2015.05.049>
6. Xu Q., Sun H.-Y., Yang Y.-H., Liu L.-H., Li Z.-Y. Optical properties and color generation mechanism of porous anodic alumina films // *Applied Surface Science*. 2011. V. 258. N 5. P. 1826–1830. <https://doi.org/10.1016/j.apsusc.2011.10.054>
7. Bae Y., Yu J., Jung Y., Lee D., Choi D. Cost-effective and high-throughput plasmonic interference coupled nanostructures by using quasi-uniform anodic aluminum oxide // *Coatings*. 2019. V. 9. N 7. P. 420. <https://doi.org/10.3390/coatings9070420>
8. Wang X.G., Wang J., Li J.-F., Tao D.-W., Zhou W.-M., Li Y., Wang C.-W. Silver loaded anodic aluminum oxide defective photonic crystals and their application for surface enhanced Raman scattering // *Optical Materials*. 2020. V. 105. P. 109982. <https://doi.org/10.1016/j.optmat.2020.109982>
9. Santos A. Nanoporous anodic alumina photonic crystals: Fundamentals, developments and perspectives // *Journal of Materials Chemistry C*. 2017. V. 5. N 23. P. 5581–5599. <https://doi.org/10.1039/c6tc05555a>

References

1. Maoz B.M., Chaikin Y., Tesler A.B., Bar Elli O., Fan Z., Govorov A.O., Markovich G. Amplification of chiroptical activity of chiral biomolecules by surface plasmons. *Nano Letters*, 2013, vol. 13, no. 3, pp. 1203–1209. <https://doi.org/10.1021/nl304638a>
2. Xu S., Cao Y., Zhou J., Wang X., Wang X., Xu W. Plasmonic enhancement of fluorescence on silver nanoparticle films. *Nanotechnology*, 2011, vol. 22, no. 27, pp. 275715. <https://doi.org/10.1088/0957-4484/22/27/275715>
3. Choudhari K.S., Kulkarni S.D., Santhosh C., George S.D. Photoluminescence enhancement and morphological properties of nanoporous anodic alumina prepared in oxalic acid with varying time and temperature. *Microporous and Mesoporous Materials*, 2018, vol. 271, pp. 138–145. <https://doi.org/10.1016/j.micromeso.2018.06.004>
4. Zhou Z.K., Lei D.Y., Liu J., Liu X., Xue J., Zhu Q., Chen H., Liu T., Li Y., Zhang H., Wang X. Shaping the emission spectral profile of quantum dots with periodic dielectric and metallic nanostructures. *Advanced Optical Materials*, 2014, vol. 2, no. 1, pp. 56–64. <https://doi.org/10.1002/adom.201300354>
5. Ho W.J., Cheng P.Y., Hsiao K.Y. Plasmonic silicon solar cell based on silver nanoparticles using ultra-thin anodic aluminum oxide template. *Applied Surface Science*, 2015, vol. 354, pp. 25–30. <https://doi.org/10.1016/j.apsusc.2015.05.049>
6. Xu Q., Sun H.-Y., Yang Y.-H., Liu L.-H., Li Z.-Y. Optical properties and color generation mechanism of porous anodic alumina films. *Applied Surface Science*, 2011, vol. 258, no. 5, pp. 1826–1830. <https://doi.org/10.1016/j.apsusc.2011.10.054>
7. Bae Y., Yu J., Jung Y., Lee D., Choi D. Cost-effective and high-throughput plasmonic interference coupled nanostructures by using quasi-uniform anodic aluminum oxide. *Coatings*, 2019, vol. 9, no. 7, pp. 420. <https://doi.org/10.3390/coatings9070420>
8. Wang X.G., Wang J., Li J.-F., Tao D.-W., Zhou W.-M., Li Y., Wang C.-W. Silver loaded anodic aluminum oxide defective photonic crystals and their application for surface enhanced Raman scattering. *Optical Materials*, 2020, vol. 105, pp. 109982. <https://doi.org/10.1016/j.optmat.2020.109982>
9. Santos A. Nanoporous anodic alumina photonic crystals: Fundamentals, developments and perspectives. *Journal of Materials Chemistry C*, 2017, vol. 5, no. 23, pp. 5581–5599. <https://doi.org/10.1039/c6tc05555a>

10. Liu S., Tian J., Zhang W. Fabrication and application of nanoporous anodic aluminum oxide: A review // *Nanotechnology*. 2021. V. 32. N 22. P. 222001. <https://doi.org/10.1088/1361-6528/abe25f>
11. Kumeria T., Santos A. Nanoporous alumina membranes for chromatography and molecular transporting // *Springer Series in Materials Science*. 2015. V. 219. P. 293–318. https://doi.org/10.1007/978-3-319-20334-8_10
12. Alekseeva N., Cema G., Podorozhkin D., Solovyev V., Trifonov S., Veisman V. Physical properties of self-assembled porous alumina structures filled with iodine // *Journal of Self-Assembly and Molecular Electronics (SAME)*. 2015. V. 2. N 1. P. 27–40. <https://doi.org/10.13052/jsame2245-4551.212>
13. Yin H., Li X., Que L. Fabrication and characterization of aluminum oxide thin film micropatterns on the glass substrate // *Microelectronic Engineering*. 2014. V. 128. P. 66–70. <https://doi.org/10.1016/j.mee.2014.05.020>
14. Nabiullina R.D., Nikitin I., Soloveva E., Gladskikh I., Starovoytov A.A. Optical properties of nanoporous aluminum oxide activated by molecular clusters of pseudocyanine dye // *Proceedings of SPIE*. 2022. V. 12131. P. 121310S. <https://doi.org/10.1117/12.2621343>
15. Toropov N.A., Gladskikh I.A., Parfenov P.S., Vartanyan T.A. Fabrication and laser-assisted modification of the Ag particles ensembles supporting quadrupole plasmon oscillations // *Optical and Quantum Electronics*. 2017. V. 49. N 4. P. 154. <https://doi.org/10.1007/s11082-017-0996-5>
16. Amendola V., Pilot R., Frascioni M., Maragò O.M., Iati M.A. Surface plasmon resonance in gold nanoparticles: A review // *Journal of Physics: Condensed Matter*. 2017. V. 29. N 20. P. 203002. <https://doi.org/10.1088/1361-648X/aa60f3>
17. Nikitin I.Y., Nabiullina R.D., Borodina L.N., Starovoytov A.A., Gladskikh I.A. Optical properties of a hybrid films of J-aggregates and aluminum oxide formed on an island Ag film // *Proc. of the International Conference Laser Optics (ICLO)*. 2022. <https://doi.org/10.1109/iclo54117.2022.9840201>
10. Liu S., Tian J., Zhang W. Fabrication and application of nanoporous anodic aluminum oxide: A review. *Nanotechnology*, 2021, vol. 32, no. 22, pp. 222001. <https://doi.org/10.1088/1361-6528/abe25f>
11. Kumeria T., Santos A. Nanoporous alumina membranes for chromatography and molecular transporting. *Springer Series in Materials Science*, 2015, vol. 219, pp. 293–318. https://doi.org/10.1007/978-3-319-20334-8_10
12. Alekseeva N., Cema G., Podorozhkin D., Solovyev V., Trifonov S., Veisman V. Physical properties of self-assembled porous alumina structures filled with iodine. *Journal of Self-Assembly and Molecular Electronics (SAME)*, 2015, vol. 2, no. 1, pp. 27–40. <https://doi.org/10.13052/jsame2245-4551.212>
13. Yin H., Li X., Que L. Fabrication and characterization of aluminum oxide thin film micropatterns on the glass substrate. *Microelectronic Engineering*, 2014, vol. 128, pp. 66–70. <https://doi.org/10.1016/j.mee.2014.05.020>
14. Nabiullina R.D., Nikitin I., Soloveva E., Gladskikh I., Starovoytov A.A. Optical properties of nanoporous aluminum oxide activated by molecular clusters of pseudocyanine dye. *Proceedings of SPIE*, 2022, vol. 12131, pp. 121310S. <https://doi.org/10.1117/12.2621343>
15. Toropov N.A., Gladskikh I.A., Parfenov P.S., Vartanyan T.A. Fabrication and laser-assisted modification of the Ag particles ensembles supporting quadrupole plasmon oscillations. *Optical and Quantum Electronics*, 2017, vol. 49, no. 4, pp. 154. <https://doi.org/10.1007/s11082-017-0996-5>
16. Amendola V., Pilot R., Frascioni M., Maragò O.M., Iati M.A. Surface plasmon resonance in gold nanoparticles: A review. *Journal of Physics: Condensed Matter*, 2017, vol. 29, no. 20, pp. 203002. <https://doi.org/10.1088/1361-648X/aa60f3>
17. Nikitin I.Y., Nabiullina R.D., Borodina L.N., Starovoytov A.A., Gladskikh I.A. Optical properties of a hybrid films of J-aggregates and aluminum oxide formed on an island Ag film. *Proc. of the International Conference Laser Optics (ICLO)*, 2022. <https://doi.org/10.1109/iclo54117.2022.9840201>

Авторы

Никитин Игорь Юрьевич — инженер, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, <https://orcid.org/0000-0002-7911-9092>, nikitinigor512@gmail.com

Нобиуллина Резида Даниловна — инженер, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, [sc 57190217593](https://orcid.org/0000-0002-1891-3057), [https://orcid.org/0000-0002-1891-3057, rezida2105@mail.ru](mailto:rezida2105@mail.ru)

Нашекин Алексей Викторович — кандидат физико-математических наук, старший научный сотрудник, Физико-технический институт им. А.Ф. Иоффе РАН, Санкт-Петербург, 194021, Российская Федерация, [sc 6603372975](https://orcid.org/0000-0002-2542-7364), <https://orcid.org/0000-0002-2542-7364>, nashchekin@mail.ioffe.ru

Старовойтов Антон Андреевич — кандидат физико-математических наук, доцент, доцент, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, [sc 23006700300](https://orcid.org/0000-0002-2160-1281), <https://orcid.org/0000-0002-2160-1281>, Anton.starovoytov@gmail.com

Гладских Игорь Аркадьевич — кандидат физико-математических наук, старший научный сотрудник, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, [sc 56020051200](https://orcid.org/0000-0001-6389-1611), <https://orcid.org/0000-0001-6389-1611>, 138020@mail.ru

Статья поступила в редакцию 18.08.2022
Одобрена после рецензирования 06.11.2022
Принята к печати 15.01.2023

Authors

Igor Yu. Nikitin — Engineer, ITMO University, Saint Petersburg, 197101, Russian Federation, <https://orcid.org/0000-0002-7911-9092>, nikitinigor512@gmail.com

Rezida D. Nabiullina — Engineer, ITMO University, Saint Petersburg, 197101, Russian Federation, [sc 57190217593](https://orcid.org/0000-0002-1891-3057), [https://orcid.org/0000-0002-1891-3057, rezida2105@mail.ru](mailto:rezida2105@mail.ru)

Alexey V. Nashyokin — PhD (Physics & Mathematics), Senior Researcher, Ioffe Institute, Saint Petersburg, 194021, Russian Federation, [sc 6603372975](https://orcid.org/0000-0002-2542-7364), <https://orcid.org/0000-0002-2542-7364>, nashchekin@mail.ioffe.ru

Anton A. Starovoytov — PhD (Physics & Mathematics), Associate Professor, Associate Professor, ITMO University, Saint Petersburg, 197101, Russian Federation, [sc 23006700300](https://orcid.org/0000-0002-2160-1281), <https://orcid.org/0000-0002-2160-1281>, Anton.starovoytov@gmail.com

Igor A. Gladskikh — PhD (Physics & Mathematics), Senior Researcher, ITMO University, Saint Petersburg, 197101, Russian Federation, [sc 56020051200](https://orcid.org/0000-0001-6389-1611), <https://orcid.org/0000-0001-6389-1611>, 138020@mail.ru

Received 18.08.2022
Approved after reviewing 06.11.2022
Accepted 15.01.2023



Работа доступна по лицензии
Creative Commons
«Attribution-NonCommercial»

АВТОМАТИЧЕСКОЕ УПРАВЛЕНИЕ И РОБОТОТЕХНИКА

AUTOMATIC CONTROL AND ROBOTICS

doi: 10.17586/2226-1494-2023-23-1-35-43

УДК 681.51

**Анализ точности оценивания состояний асинхронного электропривода
алгоритмами Люенбергера и Калмана**

**Виктор Григорьевич Букреев¹, Елена Борисовна Шандарова²,
Филипп Владимирович Перовщиков³**

^{1,2,3} Томский политехнический университет, Томск, 634050, Российская Федерация

¹ bukreev@tpu.ru, <https://orcid.org/0000-0001-9861-9765>

² shandarovaelena@mail.ru, <https://orcid.org/0000-0001-7473-2771>

³ fvp2@tpu.ru, <https://orcid.org/0000-0002-7932-2744>

Аннотация

Предмет исследования. В сложных электромеханических объектах, содержащих электроприводы с асинхронными двигателями, часто затруднена или невозможна установка датчиков выходных переменных величин. В этом случае для определения выходных координат двигателя в систему управления электропривода необходимо введение наблюдателей состояния. Основные проблемы создания наблюдателей — наличие шумов и помех в измерительных цепях системы управления, которые влияют на точность оценки неизмеряемых переменных состояния. В работе представлено сравнение точности оценок, полученных в результате работы алгоритмов наблюдателя на основе фильтра Калмана и наблюдателя Люенбергера в системе асинхронного электропривода с векторным регулированием при зашумленности каналов измерения токов в обмотках статора асинхронного двигателя. **Методы.** Для синтеза алгоритмов наблюдателей состояния использованы методы теории идентификации и квазилинеаризация нелинейных моделей рассматриваемого объекта управления. Имитационная модель асинхронного двигателя построена на классической векторной полеориентированной системе управления, где в качестве сигнала обратной связи использована оценка угловой скорости вала двигателя. Модель реализует следующие блоки: математическую модель асинхронного двигателя в двухфазной неподвижной системе координат α - β ; структуру алгоритма наблюдателей; процедуру преобразования базиса вектора тока и управляющего напряжения из неподвижного во вращающийся и наоборот; пропорционально-интегральные регуляторы тока, потокосцепления и скорости. S-образный задатчик интенсивности формирует кривую задания скорости. Входными сигналами для наблюдателей являются статорные напряжения и токи эталонной модели асинхронного двигателя. Коэффициенты адаптации для наблюдателя Люенбергера подобраны экспериментально из условия получения средней минимальной величины модуля разности оцениваемых величин. Ковариационные матрицы для наблюдателя на основе фильтра Калмана настроены на основании эксперимента, исходя из обеспечения минимума среднего значения абсолютной ошибки. **Основные результаты.** Получены временные зависимости переходных процессов угловой скорости вала, модуля векторов потокосцепления ротора и статорных токов. Выполнена оценка зависимостей при пуске асинхронного двигателя с номинальными значениями и значениями частоты и напряжения, составляющих 10 % от номинальных значений. Исследована работа алгоритмов оценивания при наличии шумовой составляющей, а также при изменении параметров схемы замещения асинхронного двигателя на ± 10 %. Получены результаты моделирования работы электропривода в пусковых режимах с механической нагрузкой, равной номинальному значению при частоте питающего напряжения 50 Гц и при 10 % от номинала для напряжения 1 Гц. Показано, что наибольшие относительные ошибки оценивания возникают в пусковом режиме работы электропривода, а максимальная точность достигается в случае применения нелинейного фильтра Калмана. **Практическая значимость.** Результаты работы могут быть использованы при разработке систем автоматического управления бездатчиковыми электроприводами и частотно-регулируемого электропривода центробежных насосных установок добычи нефти.

Ключевые слова

нелинейный фильтр Калмана, наблюдатель Люенбергера, полеориентированное управление, векторное управление, асинхронный двигатель

Ссылка для цитирования: Букреев В.Г., Шандарова Е.Б., Перевошчиков Ф.В. Анализ точности оценивания состояний асинхронного электропривода алгоритмами Люенбергера и Калмана // Научно-технический вестник информационных технологий, механики и оптики. 2023. Т. 23, № 1. С. 35–43. doi: 10.17586/2226-1494-2023-23-1-35-43

State estimation accuracy analysis of an induction electric drive by the algorithms of Luenberger and Kalman

Victor G. Bukreev^{1✉}, Elena B. Shandarova², Filipp V. Perevoshchikov³

^{1,2,3} Tomsk Polytechnic University, Tomsk, 634050, Russian Federation

¹ bukreev@tpu.ru✉, <https://orcid.org/0000-0001-9861-9765>

² shandarovaelena@mail.ru, <https://orcid.org/0000-0001-7473-2771>

³ fvp2@tpu.ru, <https://orcid.org/0000-0002-7932-2744>

Abstract

In complex electromechanical objects containing electric drives with induction motors, it is often difficult or impossible to install sensors of output variables. In this case, to determine the output coordinates of the motor, it is necessary to introduce state observers into the control system of the electric drive. The main problems of creating observers are the presence of noise and interference in the measuring circuits of the control system which affect the accuracy of the estimation of immeasurable state variables. The paper presents a comparison of the accuracy of estimates obtained as a result of the work of the observer algorithms based on the Kalman filter and the observer of the Luenberger in the induction electric drive system, with vector control at the noise level of the current measurement channels in the stator windings of the induction motor. To synthesize algorithms for state observers, methods of identification theory and quasi-linearization of nonlinear models of the control object under consideration were used. The simulation model of an induction motor is based on a classical vector field-oriented control system where an estimate of the angular speed of the motor shaft is used as a feedback signal. The model implements the following blocks: a mathematical model of an induction motor in a two-phase fixed coordinate system α - β ; the structure of the observer algorithm; the procedure for converting the basis of the current vector and the control voltage from stationary to rotating and vice versa; proportional-integral regulators of current, flux linkage and angular speed. The S-shaped intensity setter forms a speed setting curve. The input signals for observers are the stator voltages and currents of the reference model of an induction motor. The adaptation coefficients for the Luenberger observer were selected experimentally from the condition of obtaining the average minimum value of the difference modulus of the estimated values. The covariance matrices for the observer based on the Kalman filter are configured on the basis of the experiment, ensuring a minimum of the average value of the absolute error. The time dependences of the transients of the angular speed of the shaft, the modulus of the flux linkage vectors of the rotor and stator currents are obtained. The dependencies were evaluated when starting an induction motor with nominal values and values of frequency and voltage amounting to 10 % of the nominal values. The work of estimation algorithms in the presence of a noise component, as well as when changing the parameters of the induction motor replacement circuit by $\pm 10\%$, is investigated. The results of modeling the operation of the electric drive in starting modes with a mechanical load equal to the nominal value at a supply voltage frequency of 50 Hz and at 10 % of the nominal value for a voltage of 1 Hz are obtained. It is shown that the greatest relative estimation errors occur in the starting mode of the electric drive, and the maximum accuracy is achieved in the case of using a nonlinear Kalman filter. The results of the work can be used in the development of automatic control systems for sensorless electric drives and frequency-controlled electric drive of centrifugal pumping units for oil production.

Keywords

nonlinear Kalman filter, Luenberger observer, field-oriented control, vector control, induction motor

For citation: Bukreev V.G., Shandarova E.B., Perevoshchikov F.V. State estimation accuracy analysis of an induction electric drive by the algorithms of Luenberger and Kalman. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 2023, vol. 23, no. 1, pp. 35–43 (in Russian). doi: 10.17586/2226-1494-2023-23-1-35-43

Введение

Современный электропривод представляет собой совокупность электромеханического преобразователя энергии, силового преобразователя и устройства управления. В некоторых случаях исполнительный механизм технологического оборудования и силовая часть электропривода — единый конструктивный модуль, расположенный на значительном удалении от системы управления. Это обуславливает невозможность получения измерительной информации о переменных состояниях электропривода исполнительного механизма [1–5].

В настоящее время высокий уровень микропроцессорной техники создает предпосылки для реализации в режиме реального времени процедуры вычисления

неизмеряемых переменных состояния в различных цифровых системах с помощью математических алгоритмов, называемых наблюдателями. Этот факт предопределяет более качественное решение задачи управления многими сложными электромеханическими объектами [6, 7].

Все более актуально для промышленного применения — использование бездатчиковой системы управления автоматизированным электроприводом, являющегося нелинейным объектом и обладающим стохастическими свойствами, связанными со случайными внешними воздействиями, шумами в каналах измерения и сетевыми помехами [8–10]. Помимо этого, особый интерес представляет качество работы алгоритма оценки при отличающихся расчетных параметрах

схемы замещения от реальных, так как идентификация параметров электропривода с необходимой точностью весьма сложна. Отметим, что в процессе работы двигателя неизбежно возникает дрейф заданных значений, которые изменяют параметры математической модели объекта управления, а, соответственно, и наблюдателя состояния. Данные изменения могут привести к неадекватной оценке переменных состояния и необратимо ухудшить заданные показатели качества бездатчикового электропривода, который ориентируется на вычисленные переменные состояния, как сигналы обратной связи. В настоящей работе выполнен анализ различных вариаций характеристик наблюдателя состояния при введении в имитационную модель перечисленных возмущающих воздействий.

Способы решения задач оценивания неизмеряемых переменных состояния

Для решения задачи управления электропривода необходимо иметь полную информацию о его векторе состояния. Так как из-за конструктивных особенностей технологического оборудования использование датчика скорости невозможно, вектор переменных состояния имеет вид $\mathbf{X} = [\mathbf{Y} \ \mathbf{Z}]^T$, где \mathbf{Y} и \mathbf{Z} — вектора измеряемых и неизмеряемых переменных состояния. В таких системах рационально использовать наблюдающее устройство (наблюдатель), которое позволяет получить оценку неизмеряемой компоненты вектора состояния объекта.

На рис. 1 представлена классификация подходов наиболее часто используемых решений в задаче оценивания неизмеряемых переменных состояния объекта или системы [11].

В данной работе рассмотрены алгоритмы оценки неизмеряемых переменных состояния на основе фильтра Калмана и адаптивного наблюдателя Льюенбергера, которые используют уравнения электромагнитных процессов асинхронного двигателя [12–14].

Моделирование алгоритмов наблюдателей переменных состояния

Наблюдатель Льюенбергера для оценки частоты вращения асинхронного двигателя. Математическая модель электромагнитных процессов асинхронного двигателя в двухфазной неподвижной системе координат α – β может быть записана в векторно-матричной форме [15]:

$$\begin{cases} \frac{d\mathbf{X}}{dt} = \mathbf{A}(\omega)\mathbf{X} + \mathbf{B}\mathbf{U}, \\ \mathbf{Y} = \mathbf{C}\mathbf{X} \end{cases} \quad (1)$$

где вектора: $\mathbf{X} = [i_{1\alpha} \ i_{1\beta} \ \psi_{2\alpha} \ \psi_{2\beta}]^T$ — переменных состояния объекта, $\mathbf{Y} = [i_{1\alpha} \ i_{1\beta}]^T$ — измеряемых величин, $\mathbf{U} = [U_{1\alpha} \ U_{1\beta}]^T$ — управляющих воздействий; матрицы:

$$\mathbf{C} = [\mathbf{E} \ \mathbf{0}] \text{ — выхода, } \mathbf{E} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \text{ — единичная, } \mathbf{A}(\omega) = \begin{bmatrix} -\frac{R_e}{L_e} \mathbf{E} & \frac{A_r K_r}{L_e} \mathbf{E} - \frac{Z_p K_r \omega}{L_e} \mathbf{D} \\ K_r R_2 \mathbf{E} & -A_r \mathbf{E} + Z_p \omega \mathbf{D} \end{bmatrix} \text{ — параметров собствен-}$$

ной динамики объекта, $\mathbf{D} = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$ — оператора по-

ворота на 90° , $\mathbf{B} = \begin{bmatrix} \frac{1}{L_e} \mathbf{E} \\ \mathbf{0} \end{bmatrix}$ — параметров управления;

компоненты пространственных векторов: $i_{1\alpha}$, $i_{1\beta}$, $U_{1\alpha}$, $U_{1\beta}$ — тока и напряжения статора, $\psi_{2\alpha}$, $\psi_{2\beta}$ — потокосцепления ротора; $R_e = R_1 + R_2 K_r^2$, $L_e = L_1 - \frac{L_m^2}{L_1}$, $K_r = \frac{L_m}{L_2}$, $A_r = \frac{R_2}{L_2}$ — коэффициенты, учитывающие параметры асинхронного двигателя; R_1 и R_2 — активные сопротивления статора и ротора; L_1 и L_2 — полные индуктивности статора и ротора; L_m — индуктивность цепи намагничивания; Z_p — количество пар полюсов; ω — угловая скорость вала двигателя.

Запишем уравнения перенастраиваемой модели асинхронного двигателя на основе системы уравнений (1) в виде [16]:

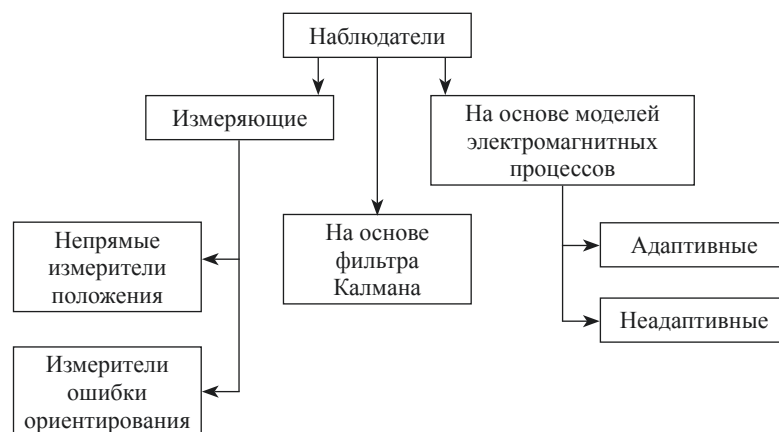


Рис. 1. Классификация наблюдателей состояния

Fig. 1. Classification of observers

$$\begin{cases} \frac{d\hat{\mathbf{X}}}{dt} = \hat{\mathbf{A}}(\hat{\omega})\hat{\mathbf{X}} + \mathbf{B}\mathbf{U} + \mathbf{L}(\hat{\mathbf{Y}} - \mathbf{Y}), \\ \hat{\mathbf{Y}} = \mathbf{C}\hat{\mathbf{X}} \end{cases}, \quad (2)$$

где $\hat{\mathbf{X}} = [\hat{i}_{1\alpha} \ \hat{i}_{1\beta} \ \hat{\psi}_{2\alpha} \ \hat{\psi}_{2\beta}]^T$ и $\hat{\mathbf{Y}} = [\hat{i}_{1\alpha} \ \hat{i}_{1\beta}]^T$ — оценки векторов переменных состояния и измеряемых величин;

$$\hat{\mathbf{A}}(\hat{\omega}) = \begin{bmatrix} -\frac{R_e}{L_e} \mathbf{E} & \frac{A_r K_r}{L_e} \mathbf{E} - \frac{Z_p K_r \hat{\omega}}{L_e} \mathbf{D} \\ K_r R_2 \mathbf{E} & -A_r \mathbf{E} + Z_p \hat{\omega} \mathbf{D} \end{bmatrix} — перенастраиваемая матрица параметров собственной динамики.$$

Для обеспечения глобальной устойчивости алгоритма адаптации, при выполнении отрицательной определенности производной функции Ляпунова, запишем закон изменения оценки частоты вращения асинхронного двигателя [16, 17]:

$$\dot{\hat{\omega}}(t) = \frac{1}{Z_p} [K_i (\hat{\mathbf{Y}} - \mathbf{Y})^T \mathbf{D} \hat{\Psi}_2 + K_p (\hat{\mathbf{Y}} - \mathbf{Y})^T \mathbf{D} \hat{\Psi}_2], \quad (3)$$

где $\hat{\Psi}_2 = [\hat{\psi}_{2\alpha} \ \hat{\psi}_{2\beta}]^T$ — оценка вектора потокоцепления; K_i и K_p — интегральный и пропорциональный коэффициенты закона адаптации. Явный вид стабилизирующей добавки \mathbf{L} в уравнении системы (2) имеет вид:

$$\mathbf{L} = \begin{bmatrix} -\hat{\omega} Z_p \frac{R_1}{R_2} \frac{L_2}{L_e} \mathbf{D} \\ \mathbf{0} \end{bmatrix}. \quad (4)$$

Структура коэффициентов добавки и ее влияние на процесс оценивания рассмотрена, например в работах [16, 17].

На основе уравнений (2)–(4) реализуем модель наблюдателя в среде математического моделирования MATLAB Simulink. Подберем экспериментально коэффициенты закона адаптации (3) из условия минимума модуля средней разности оцениваемых величин. Входными сигналами для наблюдателя примем статорные напряжения и токи статора эталонной модели асинхронного двигателя в двухфазной неподвижной системе координат. В результате моделирования получим временные зависимости переходных процессов скорости вала и модулей векторов потокоцепления

ротора и статорных токов. Проведем оценку переходных процессов при пуске асинхронного двигателя на номинальных значениях частоты и напряжения (рис. 2), а также при работе двигателя на частоте и напряжении, равных десятой части от номинальных значений. В результате получим графики переходных процессов при идеальном измерении составляющих токов статора (рис. 2, а) и при наложении аддитивной шумовой составляющей (рис. 2, б).

Алгоритм нелинейного фильтра Калмана для оценивания частоты вращения асинхронного двигателя

Классический вариант оценки неизмеряемых переменных состояния стохастической непрерывной системы — расширенный фильтр Калмана. Для рассматриваемого случая математическую модель электро-механической системы запишем следующим образом [8, 10, 18]:

$$\begin{cases} \frac{d\mathbf{X}}{dt} = \mathbf{A}(\omega)\mathbf{X} + \mathbf{B}\mathbf{U} + \mathbf{w} = \mathbf{f}(\mathbf{X}, \mathbf{U}), \\ \mathbf{Y} = \mathbf{C}\mathbf{X} + \mathbf{v} \end{cases}, \quad (5)$$

где $\mathbf{f}(\mathbf{X}, \mathbf{U})$ — дифференцируемая нелинейная функция динамики системы; $\mathbf{X} = [i_{1\alpha} \ i_{1\beta} \ \psi_{2\alpha} \ \psi_{2\beta} \ \omega]^T$ — расширенный вектор переменных состояния асинхронного двигателя с включением угловой скорости вращения вала двигателя; $\mathbf{A}(\omega) =$

$$= \begin{bmatrix} -\frac{R_e}{L_e} \mathbf{E} & \frac{A_r K_r}{L_e} \mathbf{E} - \frac{Z_p K_r \omega}{L_e} \mathbf{D} & \mathbf{0} \\ K_r R_2 \mathbf{E} & -A_r \mathbf{E} + Z_p \omega \mathbf{D} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix} — матрица параметров$$

собственной динамики объекта; \mathbf{w} и \mathbf{v} — нормальный случайный процесс и белый гауссовский шум измерений с нулевыми математическими ожиданиями $E[\mathbf{w}] = \mathbf{0}$ и $E[\mathbf{v}] = \mathbf{0}$ соответственно. Отметим, что нелинейная система (5) составлена в предположении о доминирующем значении электро-механической постоянной времени асинхронного двигателя по сравнению со значениями его электромагнитных постоянных. Это предопределяет равенство нулю производной угловой

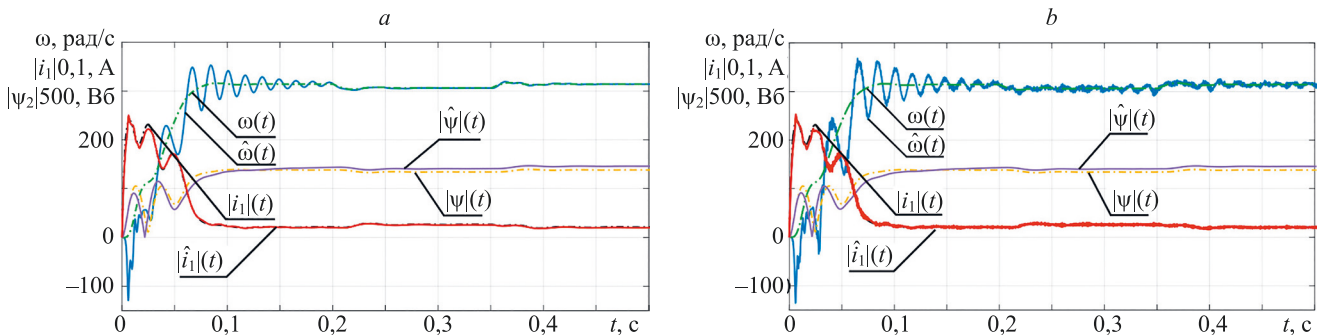


Рис. 2. Динамика изменения переменных состояния асинхронного двигателя при пуске на частоту 50 Гц при использовании наблюдателя Люенбергера без наложения шумовой компоненты в каналы измерений (а); с наложением шумовой составляющей (б)

Fig. 2. Dynamics of the change in the variables of the state of the induction motor when starting at a frequency of 50 Hz and using the Luenberger observer without superimposing a noise component into the measurement channels (a); with the superimposition of a noise component (b)

скорости вращения вала асинхронного двигателя на малом интервале дискретности расчетов и появлению нулевых значений параметров соответствующих компонент матрицы $\mathbf{A}(\omega)$.

Для дальнейшей организации вычислительного алгоритма оценки переменных состояния представим уравнения (5) в разностной форме [14, 15, 18]:

$$\begin{cases} \mathbf{X}_k = \mathbf{A}_k \mathbf{X}_{k-1} + \mathbf{B}_k \mathbf{U}_k + \mathbf{w}_k, \\ \mathbf{Y}_k = \mathbf{C}_k \mathbf{X}_k + \mathbf{v}_k \end{cases} \quad (6)$$

где $\mathbf{A}_k \approx \mathbf{A}h + \mathbf{I} =$

$$= \begin{bmatrix} \frac{-R_e h}{L_e} + 1 & 0 & \frac{A_r K_r h}{L_e} & \frac{\omega_k K_r Z_p h}{L_e} & 0 \\ 0 & \frac{-R_e h}{L_e} + 1 & \frac{-\omega_k K_r Z_p h}{L_e} & \frac{A_r K_r h}{L_e} & 0 \\ K_r R_2 h & 0 & -A_r h + 1 & \omega_k Z_p h & 0 \\ 0 & K_r R_2 h & \omega_k Z_p h & -A_r h + 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

и $\mathbf{B}_k = \mathbf{B}h$ — дискретные матрицы параметров собственной динамики и параметров управления; $\mathbf{C}_k = \mathbf{C}$ — матрица выхода; h — интервал интегрирования; k — номер интервала расчета.

Численная реализация алгоритма фильтрации (6) состоит из этапа экстраполяции и этапа коррекции [18]. Первоначально запишем модель для экстраполяции:

$$\begin{cases} \hat{\mathbf{X}}_{k|k-1} = \mathbf{A}_k \hat{\mathbf{X}}_{k-1|k-1} + \mathbf{B}_k \mathbf{U}_k, \\ \mathbf{P}_{k|k-1} = \mathbf{P}_k + \mathbf{A}_k \mathbf{P}_{k-1|k-1} \mathbf{A}_k^T + \mathbf{Q}_k \end{cases} \quad (7)$$

где $\hat{\mathbf{X}}_{k-1|k-1}$ — апостериорная оценка вектора переменных состояния; $\hat{\mathbf{X}}_{k|k-1}$ — априорная оценка вектора переменных состояния; $\mathbf{P}_{k-1|k-1} = \text{cov}(\mathbf{X}_{k-1} - \hat{\mathbf{X}}_{k-1|k-1})$ — апостериорная ковариационная матрица вектора переменных состояния, включающая в себя оценку дисперсий погрешности вычисленного состояния; $\mathbf{P}_{k|k-1} = \text{cov}(\mathbf{X}_k - \hat{\mathbf{X}}_{k|k-1})$ — априорная ковариационная матрица вектора переменных состояния; $\mathbf{P}_k \approx$

$$\approx \left(\frac{\partial}{\partial \mathbf{X}} (\mathbf{A}(\omega) \mathbf{X} + \mathbf{B} \mathbf{U}) \Big|_{\mathbf{X}=\hat{\mathbf{X}}_{k|k-1}} \right) h + \mathbf{I} =$$

$$= \begin{bmatrix} \frac{-R_e h}{L_e} + 1 & 0 & \frac{A_r K_r h}{L_e} & \frac{\omega_k K_r Z_p h}{L_e} & \frac{\Psi_{2\beta_k} Z_p h}{L_e} \\ 0 & \frac{-R_e h}{L_e} + 1 & \frac{-\omega_k K_r Z_p h}{L_e} & \frac{A_r K_r h}{L_e} & \frac{-\Psi_{2\alpha_k} Z_p h}{L_e} \\ K_r R_2 h & 0 & -A_r h + 1 & \omega_k Z_p h & -\Psi_{2\beta_k} Z_p h \\ 0 & K_r R_2 h & \omega_k Z_p h & -A_r h + 1 & \Psi_{2\alpha_k} Z_p h \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} —$$

дискретизированная матрица собственной динамики линеаризованной системы уравнений (5), где производная матрицы $\mathbf{A}(\omega)$ по векторному аргументу представляет собой матрицу Якоби; $\mathbf{Q}_k = \text{cov}(\mathbf{w}_k) = \mathbf{E}[\mathbf{w}_k \mathbf{w}_k^T]$ — ковариационная матрица случайного процесса, которая при некоррелируемых случайных величинах преобразуется в матрицу $\mathbf{Q}_k = \text{diag}(\sigma_{\mathbf{X}}^2)$, на главной диагонали которой расположены дисперсии компонент вектора

переменных состояния. После экстраполяции произведем коррекцию согласно системе уравнений [18]:

$$\begin{cases} \hat{\mathbf{X}}_{k|k} = \hat{\mathbf{X}}_{k|k-1} + \mathbf{K}_k \tilde{\mathbf{Y}}_k, \\ \mathbf{P}_{k|k} = (\mathbf{E} - \mathbf{K}_k \mathbf{H}_k) \mathbf{P}_{k|k-1} \end{cases} \quad (8)$$

где $\mathbf{H}_k = \frac{\partial}{\partial \mathbf{X}} (\mathbf{C} \mathbf{X}) \Big|_{\mathbf{X}=\hat{\mathbf{X}}_{k|k-1}}$ — линеаризованная матрица выхода; $\tilde{\mathbf{Y}}_k = \mathbf{Y}_k - \mathbf{H}_k \hat{\mathbf{X}}_{k|k-1}$ — отклонение вектора измеряемых величин и его оценки; $\mathbf{K}_k = \mathbf{P}_{k|k-1} \mathbf{H}_k^T \mathbf{S}_k^{-1}$ — оптимальная по Калману матрица усиления отклонения вектора измеряемых величин; $\mathbf{S}_k = \mathbf{H}_k \mathbf{P}_{k|k-1} \mathbf{H}_k^T + \mathbf{R}_k = \text{cov}(\tilde{\mathbf{Y}}_k)$ и $\mathbf{R}_k = \text{cov}(\mathbf{v}_k) = \mathbf{E}[\mathbf{v}_k \mathbf{v}_k^T]$ — ковариационные матрицы вектора отклонения и шума измерений.

В соответствии с уравнениями (5)–(8) создадим скрипт в MATLAB Simulink. Алгоритм оценки начнем с инициализации вектора $\hat{\mathbf{X}}_{k|k-1}$, матрицы $\mathbf{P}_{k|k-1}$ и вектора измерений \mathbf{Z}_k , элементы которых имеют первоначально нулевые значения. Далее вычислим значения компонент матрицы \mathbf{K}_k , которые в свою очередь позволят рассчитать скорректированное значение вектора $\hat{\mathbf{X}}_{k|k}$ переменных состояния и матрицу $\mathbf{P}_{k|k}$. Экстраполяция значений вектора $\hat{\mathbf{X}}_{k|k}$ переменных состояния производится по результатам вычисления \mathbf{K}_k до момента времени следующего наблюдения вектора \mathbf{Z}_k измеряемых переменных. Такую последовательность действий повторим в каждом интервале вычислений алгоритма оценивания. Проведем исследование работоспособности алгоритма оценки на тестовых массивах данных с шумовой составляющей и без нее. В результате установлена высокая степень сходимости вычислительного алгоритма при изменении значений матрицы $\mathbf{A}(\omega)$ параметров, зависящих от частоты вращения асинхронного двигателя. Ковариационные матрицы настроены экспериментальным образом на основе минимума среднего значения абсолютной ошибки оценки ω . На рис. 3 представлены графики переходных процессов частоты вращения, потокоцепления ротора, тока статора асинхронного двигателя и их оценки, полученные в результате моделирования. Временные зависимости, при идеальном измерении вектора \mathbf{Y} , показаны на рис. 3, а, а в условиях электромагнитных помех, воздействующих на датчики — на рис. 3, б.

Для численного определения качества оценок рассмотренных алгоритмов наблюдателей воспользуемся средним арифметическим значением относительных ошибок оценивания, которое вычислим по формуле:

$$\eta_x = \frac{1}{N} \sum_{j=1}^N \frac{|x_j(t_j) - \hat{x}_j(t_j)|}{|x_j(t_j)|}, \quad (9)$$

где $x_j(t_j)$ и $\hat{x}_j(t_j)$ — компонента и оценка компоненты векторов переменных состояния в момент времени t_j .

В табл. 1 приведены значения ошибок оценивания компонент вектора переменных состояния по исходным данным без и с шумами, рассчитанных по формуле (9).

В интервале времени 0,2–0,35 с электродвигателю подключена ступенчатая нагрузка, равная номинальной

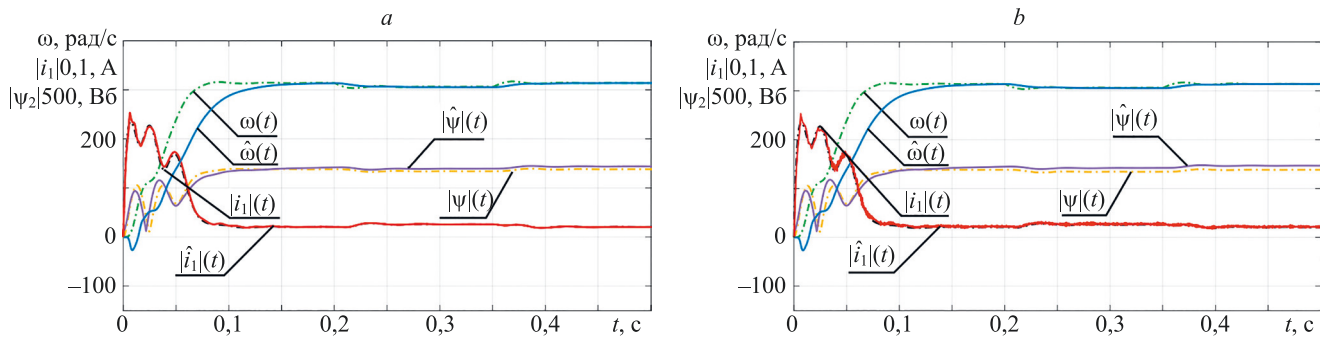


Рис. 3. Динамика изменения переменных состояния асинхронного двигателя при пуске на частоту 50 Гц при использовании фильтра Калмана без наложения шумовой компоненты в каналы измерений (а); с наложением шумовой составляющей (б)

Fig. 3. Dynamics of the change in the variables of the induction motor state when starting at a frequency of 50 Hz and using the Kalman filter without superimposing a noise component into the measurement channels (a); with the superimposition of a noise component (b)

нагрузке в случае работы на частоте 50 Гц и равная 10 % от номинальной нагрузки в случае запуска на частоте 1 Гц.

Во временном интервале 0,35–0,5 с произошел сброс приложенной ранее нагрузки.

Из анализа представленных данных видно, что во всех режимах работы и при отклонениях параметров схемы замещения электродвигателя в диапазоне $\pm 10\%$ от номинальных оценка вектора тока в разы лучше при использовании наблюдателя на основе фильтра Калмана. Аналогичный вывод можно сделать об оценке вектора потокоцепления за исключением моментов

пуска, где расхождение ошибок больше, чем в 10 раз. Похожая тенденция наблюдается при оценке угловой скорости. Точность вычисления при использовании фильтра Калмана выше, но периодически видны нарушения. Так, например, при пуске на номинальную скорость (в интервале 0,35–0,5 с) без наложения шумовой составляющей в каналы измерения с увеличенными параметрами схемы замещения, оценка на основе фильтра Калмана хуже, чем у наблюдателя Люенбергера в 2,58 раз, что не наблюдается в аналогичных модельных экспериментах при уменьшении параметров двигателя и наличии помех в сигналах тока и напряжения.

Таблица 1. Ошибки оценивания без и с введением шумовой компоненты в каналы измерения

Table 1. Estimation errors without superimposing a noise component into the measurement channels

Время, с	ω , %				$ i_1 $, %				$ \psi_2 $, %			
	наблюдатель Люенбергера		фильтр Калмана		наблюдатель Люенбергера		фильтр Калмана		наблюдатель Люенбергера		фильтр Калмана	
	$\omega \cdot 10$	$\omega/10$	$\omega \cdot 10$	$\omega/10$	$ i_1 \cdot 10$	$ i_1 /10$	$ i_1 \cdot 10$	$ i_1 /10$	$ \psi_2 \cdot 10$	$ \psi_2 /10$	$ \psi_2 \cdot 10$	$ \psi_2 /10$
Без введения шумовой компоненты в каналы измерения												
0–0,2 (Пуск 50 Гц)	110,4	81,08	31,98	32,94	2,60	5,58	2,54	0,57	11,99	22,41	30,82	14,65
0,2–0,35	0,18	0,93	0,49	0,52	2,20	4,34	0,17	0,18	4,50	6,08	1,25	0,75
0,35–0,5	0,19	0,34	0,48	0,26	3,80	4,02	0,24	0,13	5,28	5,62	0,56	0,60
0–0,2 (Пуск 1 Гц)	338,9	203,9	116,6	85,22	3,97	7,96	1,30	0,99	3,46	15,99	55,15	21,18
0,2–0,35	0,64	2,433	5,05	0,81	4,32	3,35	0,34	0,11	5,45	4,92	5,58	1,39
0,35–0,5	0,99	4,08	1,07	1,52	3,71	2,94	0,13	0,14	5,52	5,73	2,67	2,49
С введением шумовой компоненты в каналы измерения												
0–0,2 (Пуск 50 Гц)	147,5	167,9	31,89	32,47	5,98	7,02	2,59	1,80	12,65	21,08	30,88	14,93
0,2–0,35	1,21	1,14	0,49	0,52	6,39	6,91	0,35	1,44	4,56	5,94	1,23	0,77
0,35–0,5	1,10	1,59	0,48	0,26	6,58	6,93	0,49	1,50	5,34	5,56	0,56	0,62
0–0,2 (Пуск 1 Гц)	375,7	293,2	118,6	79,68	4,04	7,97	1,30	0,93	3,42	15,93	55,33	19,18
0,2–0,35	3,22	2,43	5,06	0,81	4,30	3,35	0,35	0,11	5,43	4,92	5,59	1,39
0,35–0,5	3,82	5,11	1,08	1,51	3,69	2,92	0,15	0,17	5,56	5,76	2,67	2,48

Моделирование бездатчиковой системы управления регулируемым электроприводом

Для исследования режимов работы бездатчикового электропривода создана имитационная модель его системы управления с использованием закона векторного регулирования (рис. 4).

Модель представляет классическую реализацию векторной полеориентированной системы управления, которая использует в качестве сигнала обратной связи оценку по угловой скорости. Блок «Induction motor» включает в себя нелинейные уравнения асинхронного двигателя (1); блок «Observer» реализует структуру алгоритма наблюдателей, описанных с помощью формул (2), (6), (7); блок «Regulators» состоит из ПИ-регуляторов тока, потокосцепления и скорости; блоки «Parke transform» и «Reverse Parke transform» проводят процедуру преобразования базиса вектора тока и управляющего напряжения из неподвижного во вращающийся и наоборот; блок «Intensity setter» формирует кривую задания скорости. В данном случае использован S-образный задатчик интенсивности.

Используя S-задатчик интенсивности задания частоты вращения асинхронного двигателя, выполним численное моделирование системы управления при отработке следующих режимов:

№ 1 – пуск двигателя до номинального значения ω_n частоты вращения;

№ 2 – работа на номинальной частоте вращения ω_n ;

№ 3 – уменьшение частоты вращения до $0,5\omega_n$;

№ 4 – работа с частотой вращения $0,5\omega_n$;

№ 5 – уменьшение частоты вращения до $0,1\omega_n$;

№ 6 – работа с частотой вращения $0,1\omega_n$;

№ 7 – торможение и остановка двигателя.

В результате моделирования получены зависимости переходных процессов всех компонент вектора переменных состояния. Приведем переходные процессы угловой скорости вращения вала двигателя $\omega(t)$, задающего значения скорости $\omega_s(t)$ и электромагнитного момента $M_e(t)$ при использовании наблюдателя Люенбергера (рис. 5, а) и фильтра Калмана (рис. 5, б).

Для вычисления относительной ошибки ξ_ω применим соотношение:

$$\xi_\omega = \frac{1}{M} \sum_{i=1}^M \frac{|\omega_s(t_i) - \omega(t_i)|}{|\omega(t_i)|}. \quad (10)$$

Табл. 2 показывает значения относительных ошибок регулирования при работе векторной системы управления, замкнутой по оценке угловой скорости. Значения получены при помощи формулы (10).

Исходя из полученных количественных данных табл. 2, видно, что незначительно лучше в номинальном режиме (№ 1) работы электропривода себя показывает наблюдатель Люенбергера, так как его относительная ошибка на 11,3 % меньше, чем ошибка в аналогичном режиме работы, полученная при вычислении по алгоритму фильтра Калмана. Несмотря на это, в остальных шести режимах работы двигателя существенный выигрыш в качестве оценки имеет наблюдатель на основе фильтра Калмана. Так, например, абсолютный максимум разности ошибок виден в статике на скорости $0,1\omega_n$, где ошибка ξ_ω фильтра Калмана в 10,74 раза меньше, чем у наблюдателя Люенбергера.

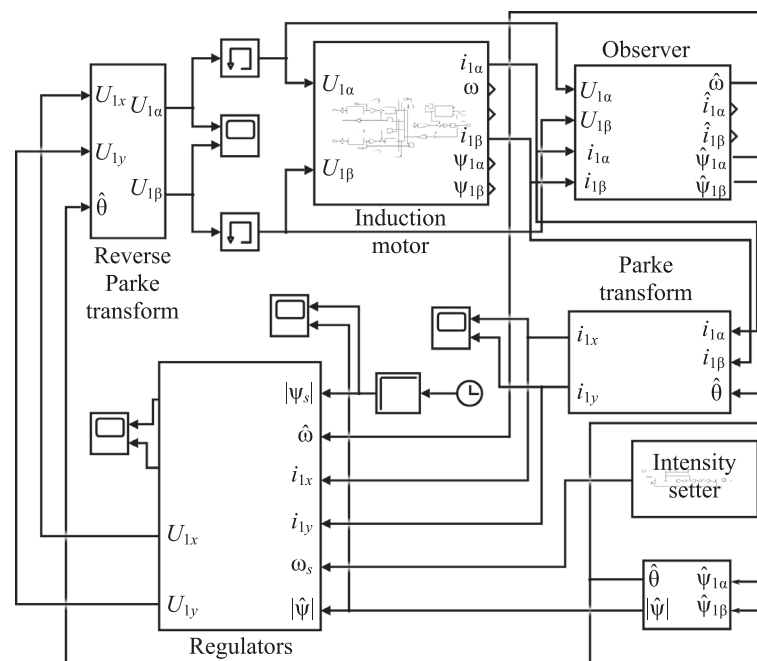


Рис. 4. Имитационная модель векторной бездатчиковой системы управления асинхронного двигателя, где U_{1x} , U_{1y} , i_{1x} , i_{1y} — проекции вектора напряжения и тока статора во вращающейся системе координат xy ; $|\psi_s|$, ω_s — задание на модуль потокосцепления и угловую скорость вращения ротора; $\hat{\theta}$ — оценка аргумента вектора потокосцепления

Fig. 4. Simulation model of a vector sensorless induction motor control system with the projections of the stator voltage and current vector in the rotating coordinate system xy ; $|\psi_s|$, ω_s is the task for the flux linkage module and the angular speed of the rotor; $\hat{\theta}$ is the evaluation of the flux vector argument

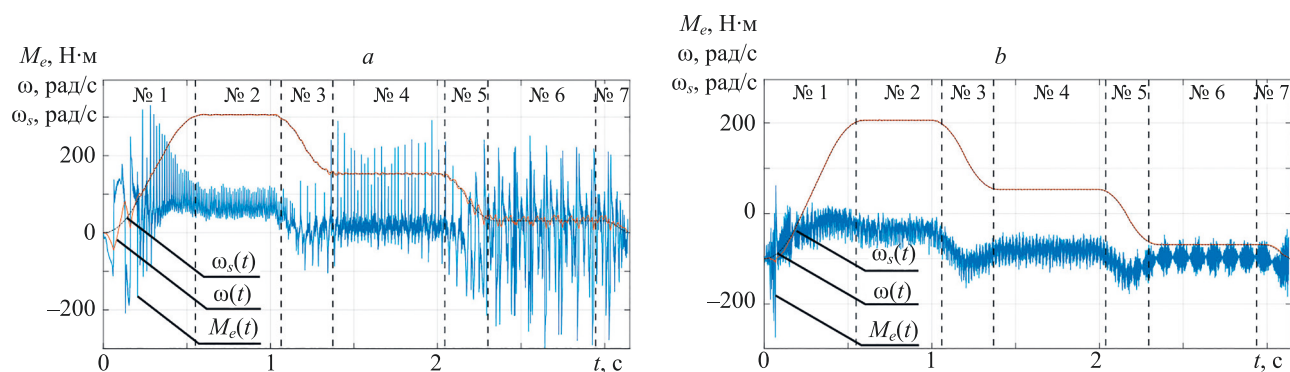


Рис. 5. Динамика изменения частоты вращения вала двигателя и электромагнитного момента при использовании: наблюдателя Люенбергера (а); фильтра Калмана (б)

Fig. 5. Dynamics of changes in the shaft speed and electromagnetic torque when using the Luenberger observer (a) and the Kalman filter (b)

Таблица 2. Ошибки регулирования при использовании бездатчиковой векторной системы управления, %

Table 2. Control errors when using a sensorless vector control system, percent

Номер режима	Наблюдатель Люенбергера	Фильтр Калмана
1	22,860	5,692
2	0,246	0,274
3	0,668	0,243
4	0,891	0,172
5	3,580	0,425
6	3,157	0,294
7	4,107	2,024

Это связано в первую очередь с тем, что закон адаптации угловой скорости наблюдателя Люенбергера имеет фиксированную структуру с постоянными параметрами пропорциональной и интегральной части в отличие от постоянно меняющейся матрицы \mathbf{K}_k , компоненты которой оптимальны и позволяют получить лучшую оценку в условиях зашумленности сигнала. Однако «платой» за лучшее качество служит обилие матричных операций и сложность подбора семи ковариаций в отличие от

наблюдателя Люенбергера, где нужно настроить лишь два параметра закона адаптации.

Заключение

В работе исследован асинхронный электропривод насосного агрегата. В программном комплексе MATLAB построена имитационная модель асинхронного двигателя, основанная на системе дифференциальных уравнений в неподвижной системе координат.

Для решения задачи бездатчикового векторного управления асинхронным электроприводом насосного агрегата реализованы алгоритмы оценки неизмеримых переменных, позволяющие провести оценку угловой скорости асинхронного двигателя, реализующие наблюдатель Люенбергера и нелинейный фильтр Калмана.

В процессе вычислительных экспериментов установлено, что наилучшую оценку при наличии шума в измерительных каналах во всех режимах работы электропривода позволяет получить фильтр Калмана, ошибка оценивания при работе которого в несколько раз меньше, чем при использовании наблюдателя Люенбергера. Однако следует отметить, что наблюдатель Люенбергера существенно быстрее реагирует на возмущения во всех режимах работы электропривода.

Литература

1. Виноградов А.Б., Сибирцев А., Колодин И. Адаптивно-векторная система управления бездатчикового асинхронного электропривода серии ЭПВ // Силовая электроника. 2006. № 9. С. 50–55.
2. Вдовин В.В., Панкратов В.В. Синтез адаптивного наблюдателя координат бездатчикового асинхронного электропривода // Известия Томского политехнического университета. 2012. Т. 320. № 4. С. 147–153.
3. Краснова С.А., Уткин В.А. Каскадный синтез наблюдателей состояния динамических систем. М.: Наука, 2006. 272 с.
4. Данг Б., Пыркин А.А., Бобцов А.А., Ведяков А.А. Синтез адаптивного наблюдателя для нестационарных нелинейных систем с неизвестными полиномиальными параметрами // Научно-технический вестник информационных технологий, механики и оптики. 2021. Т. 21. № 3. С. 374–379. <https://doi.org/10.17586/2226-1494-2021-21-3-374-379>

References

1. Vinogradov A.B., Sibirtcev A., Kolodin I. Adaptive-vector control system of a sensorless asynchronous explosion-proof electric drive (EPV series). *Silovaya Elektronika*, 2006, no. 9, pp. 50–55. (in Russian)
2. Vdovin V.V., Pankratov V.V. Synthesis of an adaptive coordinate observer of a sensorless asynchronous electric drive. *Bulletin of the Tomsk Polytechnic University*, 2012, vol. 320, no. 4, pp. 147–153. (in Russian)
3. Krasnova S.A., Utkin V.A. *Cascade Synthesis of the State Observers for the Dynamic Systems*. Moscow, Nauka Publ., 2006, 272 p. (in Russian)
4. Dang B., Pyrkin A.A., Bobtsov A.A., Vedyakov A.A. Adaptive observer design for time-varying nonlinear systems with unknown polynomial parameters. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 2021, vol. 21, no. 3, pp. 374–379. (in Russian). <https://doi.org/10.17586/2226-1494-2021-21-3-374-379>

5. Арановский С.В., Бобцов А.А., Никифоров В.О. Синтез наблюдателя для нелинейного объекта в условиях гармонического возмущения, приложенного к выходной переменной // Научно-технический вестник Санкт-Петербургского государственного университета информационных технологий, механики и оптики. 2010. № 3(67). С. 32–38.
6. Mouna B.H., Lassaad S. Speed sensorless indirect stator field oriented control of induction motor based on Luenberger observer // Proc. of the IEEE International Symposium on Industrial Electronics. V. 3. 2006. P. 2473–2478. <https://doi.org/10.1109/ISIE.2006.295961>
7. Khadar S., Kouzou A., Hafaifa A. Sensorless direct torque control of induction motor with an open-end stator winding using an adaptive Luenberger observer // Proc. of the 2018 15th International Multi-Conference on Systems, Signals & Devices (SSD). 2018. P. 1355–1362. <https://doi.org/10.1109/SSD.2018.8570691>
8. Пилаев С.Н., Еремин М.Ю., Аксенов И.И., Афоничева Д.Д. Моделирование наблюдателя состояния асинхронного электродвигателя на основе фильтра Калмана в программе Simintech // Наука в центральной России. 2021. № 6(54). С. 67–76. <https://doi.org/10.35887/2305-2538-2021-6-67-76>
9. Белов М.П., Фюонг Ч.Х., Носилов И.С. Управление электроприводом крупного радиотелескопа с линейно-квадратичным гауссовским регулятором // Известия СПбГЭТУ «ЛЭТИ». 2017. № 9. С. 52–58.
10. Гаргаев А.Н., Каширских В.Г. Применение фильтра Калмана для динамической идентификации двигателей постоянного тока // Вестник Кузбасского государственного технического университета. 2013. № 1. С. 128–130.
11. Калачев Ю.Н. Наблюдатели состояния в векторном электроприводе. М., 2015. 60 с.
12. Yin Z., Bai C., Du N., Du C., Liu J. Research on internal model control of induction motors based on Luenberger disturbance observer // IEEE Transactions on Power Electronics. 2021. V. 36. P. 8155–8170. <https://doi.org/10.1109/TPEL.2020.3048429>
13. Zerdali E. Adaptive extended Kalman filter for speed-sensorless control of induction motors // IEEE Transactions on Energy Conversion. 2019. V. 34. N 2. P. 789–800. <https://doi.org/10.1109/TEC.2018.2866383>
14. Zhao Q., Yang Z., Sun X., Ding Q. Speed-sensorless control system of a bearingless induction motor based on iterative central difference Kalman filter // International Journal of Electronics. 2020. V. 107. N 9. P. 1524–1542. <https://doi.org/10.1080/00207217.2020.1727026>
15. Nos O.V. Matrix transformations in mathematical models of an induction motor // Proc. of the 2008 9th International Conference on Actual Problems of Electronic Instrument Engineering. V. 1. 2008. P. 142–144. <https://doi.org/10.1109/APEIE.2008.4897073>
16. Вдовин В.В. Адаптивные алгоритмы оценивания координат бездатчиковых электроприводов переменного тока с расширенным диапазоном регулирования: диссертация на соискание ученой степени кандидата технических наук. Новосибирск: НГТУ, 2014. 244 с.
17. Панкратов В.В., Котин Д.А. Адаптивные алгоритмы бездатчикового векторного управления асинхронными электроприводами подъемно-транспортных механизмов. Новосибирск: НГТУ, 2012. 162 с.
18. Vas P. Sensorless Vector and Direct Torque Control. Oxford University Press, 1998. 760 p.
5. Aranovskiy S., Bobtsov A., Nikiforov V. Observer design of an unknown sinusoidal output disturbance for nonlinear plant. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 2010, no. 3(67), pp. 32–38. (in Russian)
6. Mouna B.H., Lassaad S. Speed sensorless indirect stator field oriented control of induction motor based on Luenberger observer. *Proc. of the IEEE International Symposium on Industrial Electronics. V. 3*, 2006, pp. 2473–2478. <https://doi.org/10.1109/ISIE.2006.295961>
7. Khadar S., Kouzou A., Hafaifa A. Sensorless direct torque control of induction motor with an open-end stator winding using an adaptive Luenberger observer. *Proc. of the 2018 15th International Multi-Conference on Systems, Signals & Devices (SSD)*, 2018, pp. 1355–1362. <https://doi.org/10.1109/SSD.2018.8570691>
8. Pilyaev S., Eremin M., Aksenov I., Afonicheva D. Simulation of an asynchronous state observer electric motor based on the kalman filter in the simintech program. *Science in Central Russia*, 2021, no. 6(54), pp. 67–76. (in Russian). <https://doi.org/10.35887/2305-2538-2021-6-67-76>
9. Belov M.P., Phuong T.H., Nosirov I.S. Large radio telescope control system with linear-quadratic gaussian controller. *Proceedings of Saint Petersburg Electrotechnical University Journal*, 2017, no. 9, pp. 52–58. (in Russian)
10. Gargaev A.N., Kashirskih V.G. Application of Kalman filter for dynamic identification of DC motors. *Bulletin of the Kuzbass State Technical University*, 2013, no. 1, pp. 128–130. (in Russian)
11. Kalachev Yu.N. *State Observers in Vector Electric Drives*. Moscow, 2015, 60 p. (in Russian)
12. Yin Z., Bai C., Du N., Du C., Liu J. Research on internal model control of induction motors based on Luenberger disturbance observer. *IEEE Transactions on Power Electronics*, 2021, vol. 36, pp. 8155–8170. <https://doi.org/10.1109/TPEL.2020.3048429>
13. Zerdali E. Adaptive extended Kalman filter for speed-sensorless control of induction motors. *IEEE Transactions on Energy Conversion*, 2019, vol. 34, no. 2, pp. 789–800. <https://doi.org/10.1109/TEC.2018.2866383>
14. Zhao Q., Yang Z., Sun X., Ding Q. Speed-sensorless control system of a bearingless induction motor based on iterative central difference Kalman filter. *International Journal of Electronics*, 2020, vol. 107, no. 9, pp. 1524–1542. <https://doi.org/10.1080/00207217.2020.1727026>
15. Nos O.V. Matrix transformations in mathematical models of an induction motor. *Proc. of the 2008 9th International Conference on Actual Problems of Electronic Instrument Engineering. V. 1*, 2008, pp. 142–144. <https://doi.org/10.1109/APEIE.2008.4897073>
16. Vdovin V.V. *Adaptive algorithms for coordinate estimation of the sensorless AC electric drives with an extended control range*. Dissertation for the degree of candidate of technical sciences. Novosibirsk, NSTU, 2014, 244 p. (in Russian)
17. Pankratov V.V., Kotin D.A. *Adaptive Algorithms For Sensorless Vector Control of Asynchronous Electric Drives of Carrying and Lifting Machines*. Novosibirsk, NSTU, 2012, 143 c. (in Russian)
18. Vas P. *Sensorless Vector and Direct Torque Control*. Oxford University Press, 1998, 760 p.

Авторы

Букреев Виктор Григорьевич — доктор технических наук, профессор, профессор, Томский политехнический университет, Томск, 634050, Российская Федерация, [sc 7005611043](https://orcid.org/0000-0001-9861-9765), <https://orcid.org/0000-0001-9861-9765>, bukreev@tpu.ru

Шандарова Елена Борисовна — кандидат технических наук, доцент, Томский политехнический университет, Томск, 634050, Российская Федерация, [sc 6506850084](https://orcid.org/0000-0001-7473-2771), <https://orcid.org/0000-0001-7473-2771>, shandarovaelena@mail.ru

Перовошиков Филипп Владимирович — студент, Томский политехнический университет, Томск, 634050, Российская Федерация, [sc 57216875510](https://orcid.org/0000-0002-7932-2744), <https://orcid.org/0000-0002-7932-2744>, fvp2@tpu.ru

Authors

Victor G. Bukreev — D.Sc., Full Professor, Tomsk Polytechnic University, Tomsk, 634050, Russian Federation, [sc 7005611043](https://orcid.org/0000-0001-9861-9765), <https://orcid.org/0000-0001-9861-9765>, bukreev@tpu.ru

Elena B. Shandarova — PhD, Associate Professor, Associate Professor, Tomsk Polytechnic University, Tomsk, 634050, Russian Federation, [sc 6506850084](https://orcid.org/0000-0001-7473-2771), <https://orcid.org/0000-0001-7473-2771>, shandarovaelena@mail.ru

Filipp V. Perevoshchikov — Student, Tomsk Polytechnic University, Tomsk, 634050, Russian Federation, [sc 57216875510](https://orcid.org/0000-0002-7932-2744), <https://orcid.org/0000-0002-7932-2744>, fvp2@tpu.ru

doi: 10.17586/2226-1494-2023-23-1-44-53

УДК 681.5

Метод оптимизации структуры в иерархических распределенных системах управления

Андрей Юрьевич Онуфрей¹, Александр Владимирович Разумов²,
Виталий Викторович Какаев³✉

^{1,2,3} Военно-космическая академия имени А.Ф. Можайского, Санкт-Петербург, 197198, Российская Федерация

¹ vka@mil.ru, <https://orcid.org/0000-0002-6189-5886>

² vka@mil.ru, <https://orcid.org/0000-0003-1208-2368>

³ vka@mil.ru ✉, <https://orcid.org/0000-0001-9898-2740>

Аннотация

Предмет исследования. Предложено описание системы управления в виде ориентированного графа и способ формализации задачи выбора варианта структуры системы управления. Представлены результаты исследования иерархических распределенных систем управления на основе аналитических и статистических моделей оценивания показателей эффективности и оптимизации их структуры. Разработаны модель и метод оптимизации систем управления с иерархической структурой, позволяющих на основе синтеза опорного варианта однородной иерархической структуры провести оптимизацию системы управления с произвольной структурой. По сравнению с известными методами, допущение об однородности структуры системы управления дает возможность использовать аналитическое решение при выборе опорного плана в соответствии с предложенным критерием в векторном пространстве «оперативность–стоимость». **Метод.** Предложенный метод оптимизации базируется на доказательстве возможности исследования однородных иерархических распределенных систем управления аналитическими методами и получения зависимости стоимости и времени решения задач в системе управления от параметров ее структуры. Метод оптимизации структуры состоит из двух основных этапов. На первом этапе, в рамках аналитической модели, решаются прямая задача оптимизации по минимуму времени обработки при ограничении на стоимость и обратная задача оптимизации по минимуму стоимости при ограничении на время обработки. Результатом является выбор лучшего опорного плана однородной структуры системы управления. На втором этапе, на основе имитационного моделирования, решается задача определения «критических» участков (пунктов управления), ограничивающих эффективность функционирования системы управления. Найденные «критические» участки подлежат улучшению за счет изменения их структуры и введения новых технических решений, обеспечивающих заданные показатели эффективности всей системы управления. **Основные результаты.** Создана модель иерархической структуры системы управления. Показана процедура выбора опорного варианта и алгоритм модификации структуры системы управления. Приведен пример оптимизации структуры системы управления по критерию требуемой пропускной способности. Пример показал, что применение предложенного метода позволяет выбрать вариант структуры, удовлетворяющий выбранному критерию и заданным ограничениям. **Практическая значимость.** Представленный метод может быть применен на ранних этапах проектирования распределенных информационных систем управления при выборе вариантов их построения и обосновании требований к техническим характеристикам элементов структуры.

Ключевые слова

иерархическая распределенная система управления, аналитическая модель, имитационная модель, структура, синтез, выбор варианта, разрез в графе, критический участок, пропускная способность, оптимизация, критерий выбора

Ссылка для цитирования: Онуфрей А.Ю., Разумов А.В., Какаев В.В. Метод оптимизации структуры в иерархических распределенных системах управления // Научно-технический вестник информационных технологий, механики и оптики. 2023. Т. 23, № 1. С. 44–53. doi: 10.17586/2226-1494-2023-23-1-44-53

A method of optimizing the structure of hierarchical distributed control systems

Andrey Yu. Onufrey¹, Alexander V. Razumov², Vitaliy V. Kakaev³✉

^{1,2,3} Mozhaisky Military Aerospace Academy, Saint Petersburg, 197198, Russian Federation

¹ vka@mil.ru, <https://orcid.org/0000-0002-6189-5886>

² vka@mil.ru, <https://orcid.org/0000-0003-1208-2368>

³ vka@mil.ru✉, <https://orcid.org/0000-0001-9898-2740>

Abstract

A description of the control system represented in the form of an oriented graph and a method for formalization of the problem of choosing a variant of the control system structure are proposed. The research results of hierarchical distributed control systems based on analytical and statistical models for evaluating performance indicators and optimizing their structure are presented. A model and a method for optimization control systems with a hierarchical structure have been developed which makes it possible to optimize a control system with an arbitrary structure based on the synthesis of a reference variant of a homogeneous hierarchical structure. In comparison with the known methods, the assumption of uniformity of the structure of the control system makes it possible to use an analytical solution when choosing a reference plan in accordance with the proposed criterion in the vector space “efficiency-cost”. The proposed optimization method is based on proving the possibility to analytically investigate homogeneous hierarchical distributed control systems and on establishing the dependency of the cost and time of task solving in the control system on the structure parameters. The method for optimizing the structure in hierarchical distributed control systems consists of two main stages. At the first stage, within the framework of the analytical model, the direct optimization problem is solved by minimizing the processing time with a cost constraint and the inverse optimization problem is solved by minimizing the cost with a processing time constraint. The result is the choice of the best reference plan for a homogeneous control system structure. At the second stage, based on simulation modeling, the problem of determining the “critical” areas (control points) that limit the effectiveness of the control system is solved. The found “critical” areas are subject to improvement by changing their structure and introducing new technical solutions that provide the specified performance indicators of the entire management system. A model of the hierarchical structure of the management system is given. The procedure for selecting the reference variant and the algorithm for modifying the structure of the control system are shown. An example of optimizing the structure of the control system is given according to the criterion of the required throughput. The example showed that the application of the proposed method allows choosing a variant of the structure of the control system that satisfies the selected criterion and the specified constraints. It is advisable to apply the proposed method at the early stages of designing distributed information control systems when choosing variants of their construction and substantiating requirements for the technical characteristics of structural elements.

Keywords

hierarchical distributed control system, analytical model, simulation model, structure, analytical synthesis, construction options, graph section, critical area, throughput, optimization, selection criteria

For citation: Onufrey A.Yu., Razumov A.V., Kakaev V.V. A method of optimizing the structure of hierarchical distributed control systems. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 2023, vol. 23, no. 1, pp. 44–53 (in Russian). doi: 10.17586/2226-1494-2023-23-1-44-53

Введение

Основными тенденциями при построении современных систем управления (СУ) являются распределенность, многосвязность, многофункциональность, использование искусственного интеллекта, новейших информационных и коммуникационных технологий. Принципы и способы построения таких СУ рассмотрены в работах [1–5]. Указанные тенденции способствовали повышению качественных характеристик СУ, но наряду с положительным эффектом внедрение новых решений значительно усложнило процесс их исследования и проектирования. Одна из важных задач при проектировании СУ — формирование возможных вариантов структуры и выбора наилучшего с точки зрения успешности решения поставленных задач. Данному вопросу посвящен ряд работ в России и за рубежом. Анализ их содержания показал, что большинство касаются вопросов моделирования процессов обработки информации и оценки эффективности функционирования [6–12], а также синтеза различных вариантов построения СУ [13–15]. Синтез структур СУ в настоящее время ведется в направлении применения методов аналитического и имитационного моделиро-

вания с использованием современных технологий и модельно-ориентированного проектирования, синтеза нелинейных СУ на основе квазилинейных моделей объектов. Работы [16–19] направлены на оптимизацию структур СУ на этапах выбора варианта из множества альтернатив. В [16, 17] рассмотрены модели и методы многокритериальной оптимизации СУ с иерархической структурой. В [18, 19] предложен метод выбора вариантов из однородных множеств и вероятностная модель задачи условной оптимизации.

Дальнейшее развитие распределенных иерархических СУ в настоящее время происходит в направлении совершенствования вычислительных средств (повышения быстродействия элементной базы и вычислительных устройств, надежности, снижения массогабаритных характеристик вычислительных средств), оптимизации протоколов обмена между пунктами управления и совершенствования алгоритмов обработки и передачи информации между потребителями. Данные исследования привели к возможности распределенной обработки данных на вычислительных средствах в иерархических СУ. Однако процессы обработки информации в условиях динамически изменяющейся информационной нагрузки и режимов эксплуатации,

с учетом важности и приоритетности решения задач в иерархических распределенных СУ, в настоящее время недостаточно исследованы. В связи с этим одной из важных задач остается оптимизация структуры СУ на основе формирования обобщенных моделей функционирования СУ и синтеза возможных вариантов ее структуры.

Постановка задачи исследования

Пусть задано множество W_S вариантов иерархических структур СУ:

$$W_S = (S, m_k : k = 1, S),$$

где S — число уровней управления в иерархии СУ; m_k — число пунктов на k -м уровне иерархии структуры.

Требуется выбрать вариант структуры из множества альтернатив, удовлетворяющих условиям выполнения целевой задачи СУ.

Решение проведем на основе реализации двухступенчатой процедуры оптимизации: выбора аналитическим способом опорного, упрощенного, варианта СУ и его дальнейшего направленного улучшения с использованием имитационного моделирования.

Выбор опорного варианта СУ

Представим структуру СУ в виде графа $G(X, Y)$, где X — множество вершин, соответствующих пунктам управления; Y — множество связей между пунктами. Для иерархических СУ структура графа представляет собой некоторую иерархию с несколькими уровнями управления. На рис. 1 показан граф СУ с k -уровнями управления, x_j^k — вершина графа, соответствующая j -му пункту управления k -го уровня иерархии СУ, $k = 1, S, j = 1, m_k$.

В опорном варианте иерархической структуры учтем следующие допущения:

— предположим, что иерархическая структура однородна, т. е. характеристики обработки задач пункта-

ми управления одного и того же уровня иерархии одинаковы и каждому пункту данного уровня функционально подчинено одинаковое число пунктов нижестоящего уровня иерархии;

— все потоки информации простейшие, времена поступления и обслуживания задач пунктами распределены по экспоненциальному закону, очереди задач для каждого пункта не ограничены, пункты работают безотказно.

Для расчета опорной однородной структуры решим двухпараметрическую задачу векторной оптимизации в факторном пространстве «время—стоимость» с взаимно противоположными критериями. В качестве показателя эффективности на первом этапе оптимизации выберем показатель T_{sum} суммарных затрат времени на решение задач пунктами управления иерархической структуры.

Тогда прямая задача векторной оптимизации сводится к определению иерархической структуры, при которой

$$\min_{S(m_k)} T_{sum}(S(m_k)),$$

при

$$C_{sum}(S(m_k)) \leq C_{доп}, \quad (1)$$

а обратная задача векторной оптимизации заключается в определении иерархической структуры, при которой

$$\min_{S(m_k)} C_{sum}(S(m_k)),$$

при

$$T_{sum}(S(m_k)) \leq T_{доп}, \quad (2)$$

где $S(m_k)$ — структура СУ; C_{sum} — суммарные затраты на создание и эксплуатацию пунктов управления иерархической структуры; $C_{доп}$, $T_{доп}$ — ограничения по стоимости и времени решения задач.

Выбранные показатели качества (T_{sum} , C_{sum}) связаны с параметрами иерархической структуры следующими соотношениями [14]:

$$T_{sum} = \sum_{k=1}^S (t_k n_k + t_{kk}) = \sum_{k=1}^S \left(\frac{\Lambda \prod_{j=1}^{k-1} g_j}{\mu_k (\mu_k m_k - \Lambda \prod_{j=1}^{k-1} g_j)} + t_{kk} \right),$$

$$n_k = \frac{\lambda_k}{\mu_k - \lambda_k}, \lambda_k = \frac{\Lambda}{m_k} \prod_{j=1}^{k-1} g_j, C_{sum} = \sum_{k=1}^S m_k C_k, \quad (3)$$

где $(t_k n_k)$, t_k , λ_k и n_k — средние затраты времени при решении всех задач, среднее время решения задачи, интенсивность потока поступающих задач и среднее число задач в пунктах управления k -го уровня; Λ — интенсивность суммарного потока задач управления; g_k — коэффициент интеграции, показывающий, какая часть задачи, поступившей на k -й уровень управления, будет передана на вышестоящий уровень; $\mu_k = 1/t_k$ и C_k — интенсивность решения задач и затраты на производство и эксплуатацию пунктов управления k -го уровня иерархии структуры; t_{kk} — среднее время координации действий пунктов управления k -го уровня иерархии со стороны вышестоящих.

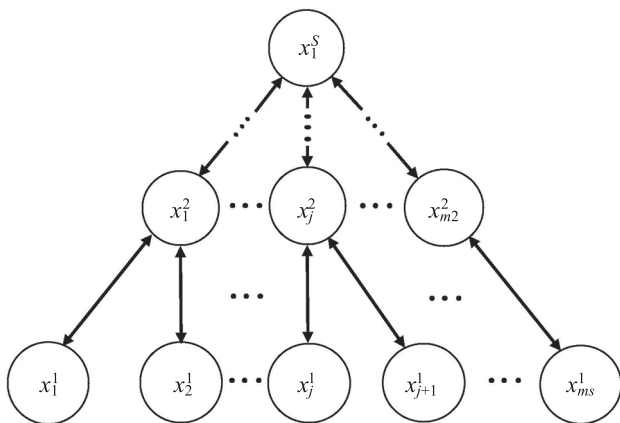


Рис. 1. Иерархическая структура системы управления, представленная в виде графа

Fig. 1. The hierarchical structure of the control system represented as a graph

Прямую и обратную задачи решим аналитически, методом неопределенных множителей Лагранжа. Функция Лагранжа для задачи (1) имеет вид:

$$L(m_k, \beta) = \sum_{k=1}^S (t_k n_k + t_{kk}) + \beta \left(\sum_{k=1}^S C_k m_k - C_{\text{доп}} \right),$$

где β — неопределенный множитель Лагранжа.

Возьмем частные производные по переменным β и m_k ($k = \overline{1, S}$), тогда, с учетом выражений (3), получим систему из $(S + 1)$ уравнений:

$$\begin{cases} \frac{\partial L(m_k, \beta)}{\partial m_k} = - \frac{\Lambda \prod_{j=1}^{k-1} g_j}{\left(\mu_k m_k - \Lambda \prod_{j=1}^{k-1} g_j \right)^2} + \beta C_k \\ \frac{\partial L(m_k, \beta)}{\partial \beta} = \sum_{k=1}^S C_k m_k - C_{\text{доп}} \end{cases} \quad (4)$$

В результате решения системы уравнений (4) относительно переменных m_k для различных значений $S = \overline{2, S_{\max}}$ определим иерархические структуры $W_S^* = (S^*, m_k^*; k = \overline{1, S})$, $S = \overline{2, S_{\max}}$ удовлетворяющие условиям задачи (1).

Аналогично, решая задачу (2), получим иерархические структуры $W_S^{**} = (S^{**}, m_k^{**}; k = \overline{1, S})$, $S = \overline{2, S_{\max}}$, удовлетворяющие условиям задачи (2), а также оптимальные значения показателей:

$$T^0 = \min T_{\text{sum}}(W_S), C^0 = \min C_{\text{sum}}(W_S), \\ W_S \in \{W_S^*, W_S^{**}; S = \overline{2, S_{\max}}\}.$$

Окончательный выбор опорного варианта иерархической структуры W_{S_0} произведем после нормировки и аддитивной свертки показателей $T_{\text{sum}}, C_{\text{sum}}$:

$$W_{S_0} = \underset{\{W_S^*, W_S^{**}\}}{\operatorname{argmax}} \left\{ \alpha_1 \frac{T^0}{T_{\text{sum}}(W_S)} + \alpha_2 \frac{C^0}{C_{\text{sum}}(W_S)} \right\} = \\ = \underset{\{W_S^*, W_S^{**}\}}{\operatorname{argmax}} \{NOR_{C,T}(W_S)\}, S = \overline{2, S_{\max}}, \quad (5)$$

где α_1, α_2 — коэффициенты равнозначности нормированных показателей времени решения задач управления и стоимости (издержек); $NOR_{C,T}(W_S)$ — значение аддитивной свертки нормированных показателей затрат времени и стоимости СУ.

Таким образом, в результате решения задач (1) и (2) определим оптимальную однородную иерархическую структуру W_{S_0} , принимаемую в качестве опорного варианта для дальнейших исследований.

Оптимизация структуры СУ на основе имитационного моделирования

На основе предложенной постановки задач (1) и (2) и аналитической модели (3) сформируем опорный вариант СУ в виде однородной иерархической структуры W_S , который совершенствуем на основе имитационного моделирования (второй этап оптимизации). При этом СУ из однородной иерархической структуры, как правило, преобразуется в СУ с неоднородной структурой.

На данном этапе оптимизации структуры СУ используем более сложные показатели (чем «время-стоимость»), отражающие целенаправленные процессы, решаемые в СУ, например: пропускная способность СУ, вероятность достижения цели управления, коэффициент сохранения эффективности. Процесс ее преобразования предполагает выполнение нескольких шагов.

Шаг 1. Выполним оценивание показателей эффективности отдельных пунктов обработки информации (узлов) в структуре СУ для ее текущего варианта построения (при первом выполнении — опорный вариант).

Шаг 2. Проведем сравнение показателей. Если полученные показатели эффективности не соответствуют требуемым, то выполним шаг 3. А если соответствует — то структура СУ не требует дальнейшего совершенствования и алгоритм завершается.

Шаг 3. Определим пункты обработки информации, ограничивающие эффективность функционирования СУ — выполним поиск узлов и дуг в графе, «сдерживающие» общий поток информации СУ. В результате найдем путь (множество дуг и узлов) в графе СУ с наименьшей пропускной способностью (находится минимальный разрез графа).

Далее решим задачу перераспределения ресурсов в структуре СУ: выполним изменения количества пунктов обработки информации на уровнях управления и их характеристик; объединение информации в соответствии с логикой работы СУ и перераспределение потоков между пунктами обработки информации с последующим оцениванием показателей эффективности.

Процесс поиска оптимального решения представляет собой проведение нескольких итераций (повторяющихся действий) с шагами 2–4 алгоритма до достижения требуемых показателей эффективности путем направленного перебора. В случае отсутствия искомого результата выбирается новый опорный вариант СУ при менее «жестких» ограничениях задач (1) и (2), т. е. при увеличении допустимого времени решения задач или/и увеличении допустимой стоимости СУ.

Задача определения критического участка в структуре СУ и поиска пунктов обработки информации, ограничивающих поток информации, относится к задачам оценки потоков в сетях и может быть решена на основе определения разреза в графе с минимальной пропускной способностью [20].

Процесс преобразования опорного варианта структуры СУ может быть представлен в виде алгоритма, основанного на нахождении «критических участков» опорного варианта, у которых необходимо улучшить характеристики для достижения требуемых показателей эффективности СУ.

Алгоритм имеет следующие шаги.

Шаг 1. Выполним оценку начального уровня эффективности K_0 опорного варианта СУ. В ходе этой оценки определим начальные значения показателей эффективности $K^{cy} := K_0$ и пунктов управления K_{ks} , где K_{ks} — показатель эффективности k -го узла s -уровня в структуре СУ. В качестве показателей K^{cy} (K_0), K_{ks} , могут быть приняты: пропускная способность СУ и вероятность выполнения задач управления, которые определяются на основе имитационного моделиро-

вания путем подсчета количества решенных задач в каждом пункте управления СУ.

Шаг 2. Проведем проверку условия $K^{СУ} \geq K^{СУтр}$, где $K^{СУтр}$ — требуемое значение показателя эффективности. Если условие выполняется, то при принятых технических решениях СУ удовлетворяет требованиям (завершение работы алгоритма). В противном случае определяются критические участки в структуре СУ (шаг 3).

Шаг 3. Определение «критического участка» в структуре СУ (нахождение минимального разреза графа) и улучшение структуры СУ.

В узлах критического участка структуры СУ последовательно по мере возрастания стоимости путем моделирования исследуем варианты изменения информационных связей между пунктами управления «критического участка» и/или замены технических средств с улучшенными характеристиками. Для нового варианта структуры рассчитаем стоимость C_{sum} в соответствии с выражением (3) и показатель эффективности $K^{СУ}$:

$$K^{СУ} = K^{СУ} + \sum_{(k)} \sum_{(j)} \sum_{(r)} \Delta K_{kjr} \geq K^{СУтр},$$

где ΔK_{kjr} — прирост значения показателя эффективности j -го узла, k -уровня в структуре СУ за счет применения r -го технического средства.

Шаг 4. Осуществим проверку условия завершения алгоритма.

Если выявленный «критический участок» в структуре СУ остался прежним (при этом выполнено условие $K^{СУ} < K^{СУтр}$), то выполним переход к выбору нового опорного варианта (это является признаком окончания работы алгоритма), так как все мероприятия по улучшению структуры исчерпаны.

Если выявленный «критический участок» в структуре СУ — новый (т. е. за счет этого участка улучшение структуры СУ происходит впервые), то выполним переход к следующей итерации, к шагу 2, для улучшения структуры при нахождении нового «критического участка».

Таким образом, сформирован вариант структуры СУ, обеспечивающий минимум ее стоимости при требуемом показателе эффективности:

$$\min_{S(m_k)} C_{sum}(S(m_k)), \text{ при } K^{СУ} \geq K^{СУтр}.$$

Выбранный вариант структуры СУ отвечает критерию достаточности по показателю эффективности.

Пример оптимизации структуры системы управления

В соответствии с двухступенчатой процедурой оптимизации первоначально найдем опорный вариант структуры СУ. По условиям решения задачи выбора опорного варианта структуры (1), (2) известны следующие исходные данные (табл. 1).

Результаты решения системы уравнений (4) относительно переменных m_k для значений $S = 2, 3, 4$ и удовлетворяющие условиям задачи (1), по минимуму времени решения задач в СУ с учетом ограничения по стоимости, представлены в табл. 2. В каждой строке таблицы

Таблица 1. Исходные данные для расчета опорной однородной структуры системы управления

Table 1. Initial data for the calculation of the reference homogeneous control system structure

S	k	g_k	μ_k , 1/ч	t_{kk} , ч	C_k , тыс. усл. ед.
2	1	0,45	10,0	0,02	25
	2	1,0	15,0	0,01	30
3	1	0,3	10,0	0,02	25
	2	1,0	15,0	0,01	30
	3	1,0	15,0	0,01	35
4	1	0,3	10,0	0,02	25
	2	1,0	15,0	0,01	30
	3	1,0	15,0	0,01	35
	4	1,0	15,0	0,01	49

Примечание: $\Lambda = 22,0$ 1/ч; $\alpha_1 = \alpha_2 = 0,5$; $S_{\max} = 4$; $C_{\text{доп}} = 300$ тыс. усл. ед., $T_{\text{доп}} = 0,2$ ч.

для выбранного $C_{\text{доп}}$ представлены: количество пунктов управления (т. е. вариант структуры СУ) на каждом уровне иерархии m_1-m_2 (при $S = 2$), $m_1-m_2-(m_3)$ (при $S = 3$), $m_1-m_2-(m_3)-(m_4)$ (при $S = 4$), а также затраты на создание и эксплуатацию пунктов управления иерархической структуры C_{sum} , суммарное время на решение задач пунктами управления иерархической структуры T_{sum} и значение аддитивной свертки нормированных показателей затрат времени и стоимости СУ — $NOR_{C,T}$. Расчет C_{sum} , T_{sum} и $NOR_{C,T}$ выполним в соответствии с выражениями (3) и (5). Если $T_{sum} > T_{\text{доп}}$ или $C_{sum} > C_{\text{доп}}$, то в ячейке таблицы вместо $NOR_{C,T}$ будет указано соответствующее, невыполненное, условие по $T_{\text{доп}}$ или $C_{\text{доп}}$.

Анализ расчетов показал, что варианты структуры СУ, удовлетворяющие ограничению задачи (1), $C_{sum} \leq C_{\text{доп}}$, соответствуют множеству $W_S^* = (S^*, m_k^* : k = \overline{1, S})$, $S = \overline{2, S_{\max}}$ и в табл. 2 отмечены двойной линией:

$$W_S^* = (9-1, 10-1, 6-2-(1), 7-2-(1), 7-3-(1)).$$

Результаты решения задачи (2), по минимуму стоимости с учетом ограничения по времени решения задач, представлены в табл. 3. В каждой строке таблицы для выбранного $T_{\text{доп}}$ приведены варианты структуры СУ ($m_1-m_2-(m_3)-(m_4)$ и т. д.) по форме, аналогичной табл. 2.

Анализ расчетов (табл. 3), показал, что варианты структуры, удовлетворяющие ограничению задачи (2), $T_{sum} \leq T_{\text{доп}}$, соответствуют множеству $W_S^{**} = (S^{**}, m_k^{**} : k = \overline{1, S})$, $S = \overline{2, S_{\max}}$ — отмечены двойной линией:

$$W_S^{**} = (7-3-(1), 8-2-(1), 7-2-(1), 6-2-(1), 10-1, 9-1).$$

Проведем окончательный выбор опорного варианта иерархической структуры W_{S0} из полученных множеств W_S^* , W_S^{**} по критерию максимума значения аддитивной свертки нормированных показателей затрат времени и стоимости СУ T_{sum} , C_{sum} в соответствии с выражением (5):

$$W_{S0} = \underset{\{W_S^*, W_S^{**}\}}{\operatorname{argmax}} \{NOR_{C,T}(W_S)\} = (S^0 = 3, m_k^0 = (7-3-1))$$

при $NOR_{C,T}(W_{S0}) = 0,519$.

Таблица 2. Результаты расчета опорной однородной структуры системы управления по минимуму времени решения задач с учетом ограничения по стоимости

Table 2. The calculation results of the reference homogeneous control system structure by minimizing the processing time with a cost constraint

$C_{\text{доп}}$, тыс. усл. ед.	Содержание ячейки таблицы	$S = 2$	$S = 3$	$S = 4$
240	$m_1-m_2-(m_3)-(m_4)$ C_{sum} T_{sum} $NOR_{C,T}$	8-1 230 0,201 $T_{\text{sum}} > T_{\text{доп}}$	7-1-(1) 240 0,210 $T_{\text{sum}} > T_{\text{доп}}$	5-1-(1)-(1) 239 0,296 $T_{\text{sum}} > T_{\text{доп}}$
260	$m_1-m_2-(m_3)-(m_4)$ C_{sum} T_{sum} $NOR_{C,T}$	9-1 255 0,195 0,478	6-2-(1) 245 0,176 0,516	5-1-(1)-(1) 239 0,296 $T_{\text{sum}} > T_{\text{доп}}$
280	$m_1-m_2-(m_3)-(m_4)$ C_{sum} T_{sum} $NOR_{C,T}$	10-1 280 0,191 0,465	7-2-(1) 270 0,164 0,518	5-2-(1)-(1) 269 0,262 $T_{\text{sum}} > T_{\text{доп}}$
300	$m_1-m_2-(m_3)-(m_4)$ C_{sum} T_{sum} $NOR_{C,T}$	10-1 280 0,191 0,465	7-3-(1) 300 0,153 0,519	6-2-(1)-(1) 294 0,241 $T_{\text{sum}} > T_{\text{доп}}$
320	$m_1-m_2-(m_3)-(m_4)$ C_{sum} T_{sum} $NOR_{C,T}$	11-1 305 0,188 $C_{\text{sum}} > C_{\text{доп}}$	9-2-(1) 320 0,150 $C_{\text{sum}} > C_{\text{доп}}$	7-2-(2)-(1) 354 0,196 $C_{\text{sum}} > C_{\text{доп}}$

В табл. 2 и 3 оптимальная однородная иерархическая структура СУ W_{S0} , с соответствующими значениями $C_{\text{sum}} = 300$ тыс. усл. ед., $T_{\text{sum}} = 0,153$ ч, $NOR_{C,T} = 0,519$, отмечена двойной «жирной» линией.

Таким образом, найден опорный вариант структуры СУ, в которой получено три уровня управления: 7 пунктов управления на первом уровне, три пункта управления — на втором и один пункт управления — на третьем. Так как 7 пунктов управления первого уровня не распределяются равномерно между пунктами второго уровня, то на один из пунктов управления второго уровня будет замыкаться три пункта управления первого уровня (рис. 2, а). Данная неоднородность будет учтена при имитационном моделировании — на втором этапе оптимизации.

На втором этапе оптимизации структуры СУ использованы показатели эффективности, в качестве которых рассмотрены пропускные способности [21] пунктов управления — K_j^k и СУ в целом — $K^{\text{су}}$. Под пропускной способностью в данном примере понимаем долю задач, решенных на пунктах управления (в СУ) относительно всех поступивших за время моделирования $T_{\text{мод}}$ задач. Зададим требуемые значения пропускных способностей пунктов управления — $K_j^{k \text{ тр}}$, и СУ в целом — $K^{\text{су тр}}$. Кроме того, в отличие от первого этапа, время обслуживания задач на пунктах управления примем распределенным по нормальному закону с параметрами a_j^k и σ_j^k , что устраняет принятое ранее упрощение об экспоненциальном законе распределения времени обслуживания задач.

С учетом исходных данных первого этапа (табл. 1), выбранного опорного варианта структуры СУ $W_S^0 =$

$= (7-3-(1))$ и выражения (3) для λ_{k_2} получим значения исходных данных имитационного моделирования (табл. 4).

Получены результаты имитационного моделирования (пропускные способности пунктов управления) опорного варианта структуры СУ: $K_1^3 = 0,69$; $K_1^2 = K_2^2 = 0,89$; $K_3^2 = 0,84$; $K_1^1 - K_7^1 = 0,76$. В соответствии с алгоритмом поиска определены пункты управления, ограничивающие пропускную способность СУ, а именно x_1^3 с пропускной способностью $K_1^3 = K^{\text{су}} = 0,69$, что меньше требуемого значения $K^{\text{су тр}} = 0,7$.

Данный узел (x_1^3) находится в составе «критического участка» структуры СУ, включающего следующие пункты управления: $x_1^3 - x_3^2 - (x_5^1 - x_7^1)$ (рис. 2, а). Дальнейшее улучшение структуры выполним путем изменения информационных связей между пунктами управления «критического участка» и замены технических средств с улучшенными характеристиками:

- x_5^1 переклещивается на x_1^2 , x_6^1 переклещивается на x_2^2 ;
- x_3^2 исключается из структуры СУ, а оставшийся пункт управления x_7^1 подключается непосредственно к x_1^3 ;

- за счет стоимости исключенного пункта управления x_3^2 , улучшаются характеристики x_1^2 и x_2^2 (уменьшается коэффициент интеграции $g_2 = 0,5$, показывающий какая часть задачи, поступившей на 2-й уровень управления, будет передана на вышестоящий уровень x_1^3) в пределах заданной стоимости СУ $C_{\text{доп}} = 300$ тыс. усл. ед.

С учетом проведенных изменений структура СУ представлена на рис. 2, б, а исходные данные для имитационного моделирования — в табл. 5.

Таблица 3. Результаты расчета опорной однородной структуры системы управления по минимуму стоимости с учетом ограничения по времени решения задач

Table 3. The calculation results of the reference homogeneous control system structure by minimizing the cost with a processing time constraint

$T_{\text{доп}}, \text{ч}$	Содержание ячейки таблицы	$S = 2$	$S = 3$	$S = 4$
0,150	$m_1-m_2-(m_3)-(m_4)$ C_{sum} T_{sum} $NOR_{C,T}$	$C_{\text{sum}} > C_{\text{доп}}$	9-2-(1) 320 0,150 $C_{\text{sum}} > C_{\text{доп}}$	15-4-(3)-(1) 649 0,149 $C_{\text{sum}} > C_{\text{доп}}$
0,155	$m_1-m_2-(m_3)-(m_4)$ C_{sum} T_{sum} $NOR_{C,T}$	$C_{\text{sum}} > C_{\text{доп}}$	7-3-(1) 300 0,153 0,519	14-4-(3)-(1) 624 0,151 $C_{\text{sum}} > C_{\text{доп}}$
0,160	$m_1-m_2-(m_3)-(m_4)$ C_{sum} T_{sum} $NOR_{C,T}$	99-1 2005 0,159 $C_{\text{sum}} > C_{\text{доп}}$	8-2-(1) 295 0,156 0,518	14-3-(2)-(1) 559 0,160 $C_{\text{sum}} > C_{\text{доп}}$
0,165	$m_1-m_2-(m_3)-(m_4)$ C_{sum} T_{sum} $NOR_{C,T}$	66-1 1680 0,165 $C_{\text{sum}} > C_{\text{доп}}$	7-2-(1) 270 0,164 0,518	13-3-(2)-(1) 534 0,163 $C_{\text{sum}} > C_{\text{доп}}$
...
0,195	$m_1-m_2-(m_3)-(m_4)$ C_{sum} T_{sum} $NOR_{C,T}$	10-1 280 0,191 0,465	6-2-(1) 245 0,176 0,516	8-2-(2)-(1) 379 0,188 $C_{\text{sum}} > C_{\text{доп}}$
0,200	$m_1-m_2-(m_3)-(m_4)$ C_{sum} T_{sum} $NOR_{C,T}$	9-1 255 0,195 0,478	6-2-(1) 245 0,176 0,516	7-2-(2)-(1) 354 0,196 $C_{\text{sum}} > C_{\text{доп}}$
0,205	$m_1-m_2-(m_3)-(m_4)$ C_{sum} T_{sum} $NOR_{C,T}$	8-1 230 0,201 $T_{\text{sum}} > T_{\text{доп}}$	5-2-(1) 220 0,203 $T_{\text{sum}} > T_{\text{доп}}$	7-2-(2)-(1) 354 0,196 $C_{\text{sum}} > C_{\text{доп}}$

Таблица 4. Исходные данные для имитационного моделирования опорного варианта структуры системы управления

Table 4. Initial data for the simulation modeling of the reference variant of the control system structure

S	x_j^k	g_k	$\lambda_k, 1/\text{ч}$	$a_j^k, \text{мин}$	$\sigma_j^k, \text{мин}$	$C_k, \text{тыс. усл. ед.}$
3	x_1^3	1,0	6,59	4,0	1,5	35
2	x_1^2, x_2^2 x_3^2	1,0 1,0	1,89 2,83	4,0 4,0	1,5 1,5	30 30
1	x_1^1, \dots, x_7^1	0,3	3,14	6,0	2,0	25

Примечание: $\Lambda = 22,0 \text{ 1/ч}$; $C_{\text{доп}} = 300 \text{ тыс. усл. ед.}$; $K_j^k \text{ тр} = K^{\text{сутр}} = 0,7$; $T_{\text{мод}} = 500 \text{ ч}$.

Таблица 5. Исходные данные для имитационного моделирования скорректированного варианта структуры системы управления

Table 5. Initial data for the simulation modeling of the corrected variant of the control system structure

S	x_j^k	g_k	$\lambda_k, 1/\text{ч}$	$a_j^k, \text{мин}$	$\sigma_j^k, \text{мин}$	$C_k, \text{тыс. усл. ед.}$
3	x_1^3	1,0	3,77	4,0	1,5	35
2	x_1^2, x_2^2	0,5	2,83	4,0	1,5	45
1	x_1^1, \dots, x_7^1	0,3	3,14	6,0	2,0	25

Примечание: $\Lambda = 22,0 \text{ 1/ч}$; $C_{\text{доп}} = 300 \text{ тыс. усл. ед.}$; $K_j^k \text{ тр} = K^{\text{сутр}} = 0,7$; $T_{\text{мод}} = 500 \text{ ч}$

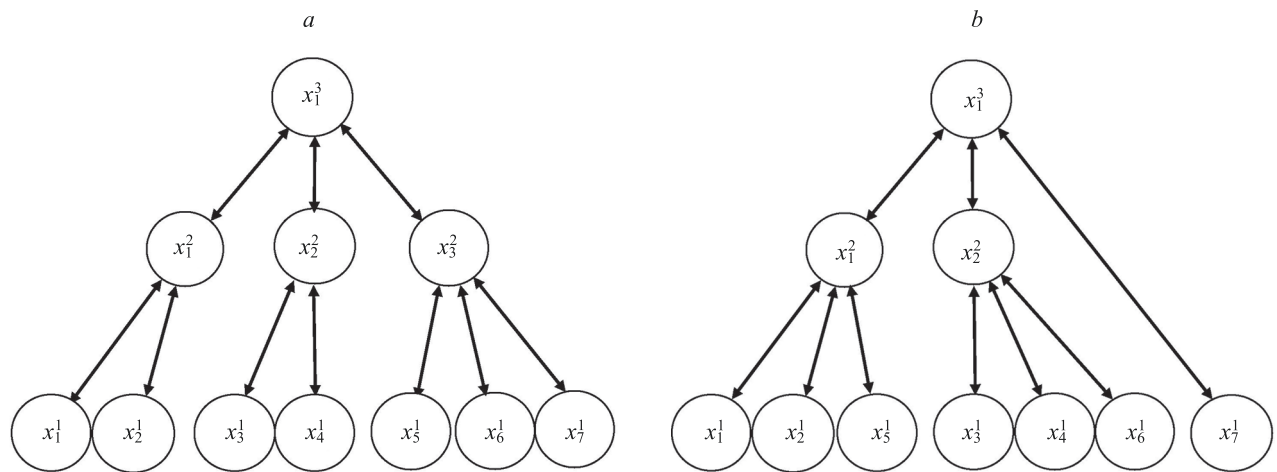


Рис. 2. Опорный вариант структуры системы управления (а), скорректированный (по результатам имитационного моделирования) вариант структуры системы управления (б)

Fig. 2. The reference variant of the control system structure (a), corrected variant of the control system structure (based on the results of simulation modeling) (b)

В результате имитационного моделирования скорректированного варианта структуры СУ получены значения пропускной способности пунктов управления: $K_1^3 = 0,79$; $K_1^2 = K_2^2 = 0,84$; $K_1^1 - K_7^1 = 0,76$. Полученные показатели эффективности пунктов управления — K_j^k и СУ в целом — $K^{СУ}$ больше требуемых значений $K_j^{k\text{тр}} = K^{СУ\text{тр}} = 0,7$.

Таким образом, сформирован вариант структуры СУ, соответствующий критерию достаточности по показателю эффективности — $K^{СУ} \geq K^{СУ\text{тр}}$.

Заключение

Совершенствование технических средств обработки и передачи данных привело к качественному изменению принципов управления в системах целенаправленных процессов, что в значительной мере расширило возможности отдельных элементов в структуре систем управления при решении задач управления. В связи с этим вопросы моделирования и исследования на моде-

лях различных вариантов структур систем управления с учетом взаимного влияния составных частей — актуальны и требуют разработки новых методов исследования с учетом декомпозиции структуры и нахождения новых нетрадиционных подходов.

Предложенный метод представляет собой двухуровневую процедуру оптимизации, позволяющую на основе аналитической модели систем управления выбрать наилучший базовый (опорный) вариант ее построения для частного случая, когда структура является иерархической и однородной. Далее, с помощью декомпозиции структуры, можно осуществить переход к исследованию системы управления с произвольной структурой и последующей ее оптимизации. Метод основан на определении критического участка в структуре систем управления с точки зрения обеспечения требуемой эффективности управления с последующим перераспределением вычислительных ресурсов и формированием очередного плана для исследований до достижения требуемых показателей эффективности.

Литература

1. Иванов Д.А., Иванова М.А., Соколов Б.В. Анализ тенденций изменения принципов управления предприятиями в условиях развития технологий Индустрии 4.0 // Труды СПИИРАН. 2018. № 5(60). С. 97–127. <https://doi.org/10.15622/sp.60.4>
2. Цветков В.Я., Алпатов А.Н. Проблемы распределенных систем // Перспективы науки и образования. 2014. № 6(12). С. 31–36.
3. Шведенко В.Н., Мозохин А.Е. Концепция управления сетевой структурой интеллектуальных устройств в условиях цифровой трансформации энергетической отрасли // Научно-технический вестник информационных технологий, механики и оптики. 2021. Т. 21. № 5. С. 748–754. <https://doi.org/10.17586/2226-1494-2021-21-5-748-754>
4. Tang B., Chen Z., Hefferman G., Wei T., He H., Yang Q. A hierarchical distributed fog computing architecture for big data analysis in smart cities // Proc. of the ASE BigData & SocialInformatics. 2015. P. 28. <https://doi.org/10.1145/2818869.2818898>
5. Oesterreich T.D., Teuteberg F. Understanding the implications of digitisation and automation in the context of Industry 4.0: A triangulation approach and elements of research agenda for the

References

1. Ivanov D.A., Ivanova M.A., Sokolov B.V. Analysis of transformation trends in enterprise management principles in the era of industry 4.0 technology. *SPIIRAS Proceedings*, 2018, no. 5(60), pp. 97–127. (in Russian). <https://doi.org/10.15622/sp.60.4>
2. Tsvetkov V.Y., Alpatov A.N. Problems of distributed systems. *Perspectives of Science and Education*, 2014, no. 6(12), pp. 31–36. (in Russian)
3. Shvedenko V.N., Mozokhin A.E. The concept of managing the network structure of intelligent devices in the digital transformation of the energy industry. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 2021, vol. 21, no. 5, pp. 748–754. (in Russian). <https://doi.org/10.17586/2226-1494-2021-21-5-748-754>
4. Tang B., Chen Z., Hefferman G., Wei T., He H., Yang Q. A hierarchical distributed fog computing architecture for big data analysis in smart cities. *Proc. of the ASE BigData & SocialInformatics*, 2015, pp. 28. <https://doi.org/10.1145/2818869.2818898>
5. Oesterreich T.D., Teuteberg F. Understanding the implications of digitisation and automation in the context of Industry 4.0: A

- construction industry // *Computers in Industry*. 2016. V. 83. P. 121–139. <https://doi.org/10.1016/j.compind.2016.09.006>
6. Чуканов С.Н. Моделирование структуры сложной системы на основе оценивания меры взаимодействия подсистем // Компьютерные исследования и моделирование. 2020. Т. 12. № 4. С. 707–719. <https://doi.org/10.20537/2076-7633-2020-12-4-707-719>
 7. Доронина Ю.В., Скатков А.В. Каскадно-иерархическое моделирование в задачах анализа динамики ресурсных характеристик сложных систем // Информационно-управляющие системы. 2020. № 3(106). С. 48–58. <https://doi.org/10.31799/1684-8853-2020-3-48-58>
 8. Новиков Д.А. Аналитическая сложность и погрешность решения задач управления организационно-техническими системами // Автоматика и телемеханика. 2018. № 5. С. 107–118.
 9. Буков В.Н., Бронников А.М., Агеев А.М., Гамаюнов И.Ф. Аналитический подход к формированию конфигураций технических систем // Автоматика и телемеханика. 2017. № 9. С. 67–83.
 10. Трегубов Р.Б., Алексиков Ю.Г., Сaitов С.И. Представление иерархических многоуровневых маршрутизируемых систем в теоретико-множественном базисе // Труды СПИИРАН. 2016. № 3(46). С. 139–168. <https://doi.org/10.15622/SP.46.15>
 11. Ашимов А.А., Гейда А.С., Лысенко И.В., Юсупов Р.М. Эффективность функционирования и другие операционные свойства систем: задачи и метод оценивания // Труды СПИИРАН. 2018. № 5(60). С. 241–270. <https://doi.org/10.15622/sp.60.9>
 12. Нетес В.А. Двусторонние оценки коэффициента сохранения эффективности систем с выходным эффектом, зависящим от числа исполнительных элементов // Автоматика и телемеханика. 2018. № 11. С. 99–105. <https://doi.org/10.31857/S000523100002746-9>
 13. Доронина Ю.В., Рябовая В.О., Чесноков Д.И. Применение модельно-ориентированного проектирования для решения задачи структурного синтеза // Труды СПИИРАН. 2016. № 6(49). С. 122–143. <https://doi.org/10.15622/sp.49.7>
 14. Багрецов С.А., Везиров В.Н., Львов В.М. и др. Технология синтеза организационных структур сложных систем управления. М. ГУП «ВНИИ межотраслевой информации – федеральный информационно-аналитический центр оборонной промышленности», Академия оборонных отраслей промышленности, ГУП «Межотраслевой центр эргономических исследований и разработок в военной технике «Эргоцентр», 1998. 224 с.
 15. Каляев И.А., Гайдук А.Р., Капустян С.Г. Модели и алгоритмы коллективного управления в группах роботов. М.: Физматлит, 2009. 280 с.
 16. Рапопорт Э.Я., Плешивцева Ю.Э. Метод многокритериальной оптимизации управляемых систем с распределенными параметрами // Труды СПИИРАН. 2018. № 5(60). С. 64–96. <https://doi.org/10.15622/sp.60.3>
 17. Губко М.В. Модели и методы оптимизации структуры иерархических систем обработки информации: диссертация на соискание ученой степени доктора физико-математических наук: 05.13.01 / Институт проблем управления им. В.А. Трапезникова РАН. М., 2014. 372 с.
 18. Кандырин Ю.В., Шкурина Г.Л. Формирование частичных порядков вариантов для выбора оптимальных альтернатив в однородных множествах // Научно-технический вестник информационных технологий, механики и оптики. 2015. Т. 15. № 3. С. 476–482. <https://doi.org/10.17586/2226-1494-2015-15-3-476-482>
 19. Тимофеева Г.А. Вероятностные решения задач условной оптимизации // Труды института математики и механики УРО РАН. 2020. Т. 26. № 1. С. 198–211. <https://doi.org/10.21538/0134-4889-2020-26-1-198-211>
 20. Форд А., Фалкерсон Д. Потоки в сетях. М.: Мир, 1966. 276 с.
 21. Вентцель Е.С. Исследование операций. М.: Советское радио, 1972. 407 с.
 - triangulation approach and elements of research agenda for the construction industry. *Computers in Industry*, 2016, vol. 83, pp. 121–139. <https://doi.org/10.1016/j.compind.2016.09.006>
 6. Chukanov S.N. Modeling the structure of a complex system based on estimation of the measure of interaction of subsystems. *Computer Research and Modeling*, vol. 12, no. 4, pp. 707–719. (in Russian). <https://doi.org/10.20537/2076-7633-2020-12-4-707-719>
 7. Doronina Y., Skatkov A. Cascade-hierarchical modeling in analyzing the dynamics of complex system resource characteristics. *Information and Control Systems*, 2020, no. 3(106), pp. 48–58. <https://doi.org/10.31799/1684-8853-2020-3-48-58>
 8. Novikov D.A. Analytical complexity and errors of solving control problems for organizational and technical systems. *Automation and Remote Control*, 2018, vol. 79, no. 5, pp. 860–869. <https://doi.org/10.1134/s0005117918050077>
 9. Bukov V.N., Bronnikov A.M., Ageev A.M., Gamayunov I.F. An analytic approach to constructing configurations of technical systems. *Automation and Remote Control*, 2017, vol. 78, no. 9, pp. 1600–1613. <https://doi.org/10.1134/s0005117917090053>
 10. Tregubov R.B., Aleksikov Y.G., Saitov S.I. Representation of the hierarchical multilayer routing systems in a set-theoretic base. *SPIIRAS Proceedings*, 2016, no. 3(46), pp. 139–168. (in Russian). <https://doi.org/10.15622/SP.46.15>
 11. Ashimov A.A., Geida A.S., Lysenko I.V., Yusupov R.M. System functioning efficiency and other system operational properties: research problems, evaluation method. *SPIIRAS Proceedings*, 2018, no. 5(60), pp. 241–270. (in Russian). <https://doi.org/10.15622/sp.60.9>
 12. Netes V.A. Bilateral estimates of the effectiveness retention ratio of the systems with output effect depending on the number of executive elements. *Automation and Remote Control*, 2018, vol. 79, no. 11, pp. 2033–2037. <https://doi.org/10.1134/S0005117918110073>
 13. Doronina J.V., Ryabovaya V.O., Chesnokov D.I. Using model-based design for solving structural synthesis problems. *SPIIRAS Proceedings*, 2016, no. 6(49), pp. 122–143. (in Russian). <https://doi.org/10.15622/sp.49.7>
 14. Bagretov S.A., Vezirov V.N., L'vov V.M. et al. *Synthesis Technology of the Organizational Framework of the Complex Control Systems*. Moscow, VIMI of Federal State Unitary Enterprise, “Ergocentr”, 1998, 224 p. (in Russian)
 15. Kaliev I.A., Gaiduk A.R., Kapustian S.G. *Models and Algorithms of the Shared Control in Robot Teams*. Moscow, Fizmatlit Publ., 2009, 280 p. (in Russian)
 16. Rapoport E.Y., Pleshivtseva Y.E. Method of multiobjective optimization of controlled systems with distributed parameters. *SPIIRAS Proceedings*, 2018, no. 5(60), pp. 64–96. (in Russian). <https://doi.org/10.15622/sp.60.3>
 17. Gubko M.V. *Model and Methods for Optimizing the Structure of Hierarchical Information Processing Systems*. Thesis for the degree of doctor of physical and mathematical sciences. Institute of Control Sciences V.A. Trapeznikov Academy of Sciences. Moscow, 2014, 327 p. (in Russian)
 18. Kandyrin Yu.V., Shkurina G.L. Creation of partial orders of variants for selection of optimal alternatives in homogeneous sets. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 2015, vol. 15, no. 3, pp. 476–482. <https://doi.org/10.17586/2226-1494-2015-15-3-476-482>
 19. Timofeeva G.A. Probabilistic solutions of conditional optimization problems. *Trudy Instituta Matematiki i Mekhaniki UrO RAN*, 2020, vol. 26, no. 1, pp. 198–211. (in Russian). <https://doi.org/10.21538/0134-4889-2020-26-1-198-211>
 20. Ford L.R., Fulkerson D.R. *Flows in Networks*. Princeton University Press, Oxford University Press, 1962, 212 p.
 21. Venttsel E.S. *Operations Research*. Moscow, Sovetskoe Radio Publ., 1972. 407 c. (in Russian)

Авторы


Онуфрей Андрей Юрьевич — доктор технических наук, профессор, ведущий научный сотрудник, Военно-космическая академия имени А.Ф. Можайского, Санкт-Петербург, 197198, Российская Федерация, <https://orcid.org/0000-0002-6189-5886>, vka@mil.ru


Разумов Александр Владимирович — доктор технических наук, профессор, старший научный сотрудник, Военно-космическая академия имени А.Ф. Можайского, Санкт-Петербург, 197198, Российская Федерация, <https://orcid.org/0000-0003-1208-2368>, vka@mil.ru

Authors

Andrey Yu. Onufrey — D.Sc., Professor, Leading Researcher, Mozhaisky Military Aerospace Academy, Saint Petersburg, 197198, Russian Federation, <https://orcid.org/0000-0002-6189-5886>, vka@mil.ru

Alexander V. Razumov — D.Sc., Professor, Senior Researcher, Mozhaisky Military Aerospace Academy, Saint Petersburg, 197198, Russian Federation, <https://orcid.org/0000-0003-1208-2368>, vka@mil.ru

Какаев Виталий Викторович — доктор военных наук, доцент, начальник управления, Военно-космическая академия имени А.Ф. Можайского, Санкт-Петербург, 197198, Российская Федерация,  57195929078, <https://orcid.org/0000-0001-9898-2740>, vka@mil.ru

Vitaliy V. Kakaev — D.Sc. (Military Science), Associate Professor, Principal Staff Scientist, Mozhaisky Military Aerospace Academy, Saint Petersburg, 197198, Russian Federation,  57195929078, <https://orcid.org/0000-0001-9898-2740>, vka@mil.ru

Статья поступила в редакцию 12.06.2022
Одобрена после рецензирования 07.12.2022
Принята к печати 18.01.2023

Received 12.06.2022
Approved after reviewing 07.12.2022
Accepted 18.01.2023



Работа доступна по лицензии
Creative Commons
«Attribution-NonCommercial»

doi: 10.17586/2226-1494-2023-23-1-54-61

УДК 681.51

Метод идентификации параметров синусоидального сигнала с неизвестной переменной амплитудой

Хак Тунг Нгуен¹✉, Сергей Михайлович Власов²

^{1,2} Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация

¹ nguyenkhactunghvhq1994@gmail.com✉, <https://orcid.org/0000-0001-6430-1927>

² smvlasov@itmo.ru, <https://orcid.org/0000-0002-8345-7553>

Аннотация

Предмет исследования. Предложен новый метод идентификации параметров синусоидального сигнала с неизвестной переменной амплитудой. Проблема оценивания параметров синусоидальных сигналов актуальна в задачах динамического позиционирования и компенсации возмущений для синтеза законов управления, учитывающих внешние возмущающие воздействия. В предлагаемом методе устранено ограничение на амплитуду сигнала. В отличие от известных подходов, где амплитуда должна быть фиксированной, в рассмотренном методе амплитуда сигнала может быть переменной. **Метод.** Для реализации представленного метода идентификации применены жорданова форма матрицы и операторы запаздывания. В ходе параметризации сформирована регрессионная модель, содержащая неизвестные стационарные параметры. Для поиска неизвестных параметров применен метод динамического расширения регрессора и смешивания. **Основные результаты.** Результаты компьютерного моделирования показали работоспособность разработанного алгоритма. Результаты моделирования подтвердили сходимость оценивания параметров к истинным значениям. **Практическая значимость.** Подход может быть применен для широкого класса прикладных задач, связанных с компенсацией возмущений в системах виброзащиты, системах мониторинга при определении параметров высотных или большепролетных строительных сооружений, в системах управления робототехническими объектами.

Ключевые слова

синусоидальные сигналы, идентификация, переменная амплитуда, жорданова форма матрицы, линейная регрессионная модель

Ссылка для цитирования: Нгуен Х.Т., Власов С.М. Метод идентификации параметров синусоидального сигнала с неизвестной переменной амплитудой // Научно-технический вестник информационных технологий, механики и оптики. 2023. Т. 23, № 1. С. 54–61. doi: 10.17586/2226-1494-2023-23-1-54-61

Method for identification of sinusoidal signal parameters with variable unknown amplitude

Khac T. Nguyen¹✉, Sergey M. Vlasov²

^{1,2} ITMO University, Saint Petersburg, 197101, Russian Federation

¹ nguyenkhactunghvhq1994@gmail.com✉, <https://orcid.org/0000-0001-6430-1927>

² smvlasov@itmo.ru, <https://orcid.org/0000-0002-8345-7553>

Abstract

A new method proposed for identifying the parameters of sinusoidal signal with unknown variable amplitude. The problem of estimating the parameters of sinusoidal signals is relevant in the problems of dynamic positioning and disturbance compensation, for the synthesis of control laws that take into account external disturbances. In the proposed method, the restriction on the signal amplitude is removed. In contrast to known approaches, where the amplitude must be fixed, in the proposed method the signal amplitude can be variable. To implement the proposed identification algorithm, the Jordan matrix form and delay operators are used. During parameterization, a regression model is formed containing unknown stationary parameters. To search for unknown parameters, the method of dynamic expansion of the

regressor and mixing is used. The results of computer simulation demonstrate the efficiency of the proposed algorithm. The simulation results confirmed the convergence of parameter estimation to the true values. The proposed approach can be applied to a wide class of applied problems related to disturbance compensation in vibration protection systems, monitoring systems in determining the parameters of high-rise or large-span building structures, and in robotic object control systems.

Keywords

sinusoidal signal, identification, variable amplitude, Jordan form of the matrix, linear regression model

For citation: Nguyen K.T., Vlasov S.M. Method for identification of sinusoidal signal parameters with variable unknown amplitude. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 2023, vol. 23, no. 1, pp. 54–61 (in Russian). doi: 10.17586/2226-1494-2023-23-1-54-61

Введение

В работе рассмотрена задача идентификации параметров синусоидального сигнала с неизвестной нестационарной амплитудой. Проблема идентификации частоты синусоидальных сигналов изучена для постоянных [1–7] и переменных [8–13] параметров. Задача идентификации частоты синусоидальных сигналов является очень важной и актуальной, решение которой может найти в большом количестве практических приложений: в системах динамического позиционирования и компенсации возмущений, системах виброзащиты и мониторинга при определении параметров высотных или большепролетных строительных сооружений.

В работе [8] представлен алгоритм оценивания нестационарной частоты незашумленного синусоидального сигнала. Предположено, что амплитуда и частота синусоидального сигнала неизвестные функции времени, но являются решениями линейных стационарных дифференциальных уравнений с известными параметрами. Основной недостаток данного подхода состоит в снятии ограничений на стационарность амплитуды и частоты, но фаза синусоидального сигнала — постоянна.

В [10] рассмотрен алгоритм идентификации линейно меняющейся частоты синусоидального сигнала, имеющего неизвестные амплитуду и фазу. Предложен подход к идентификации линейно меняющейся частоты, предусматривающий итеративную фильтрацию (с использованием линейных устойчивых звеньев первого порядка) измеряемого синусоидального сигнала, которая позволяет получить простую линейную регрессионную модель с одним неизвестным постоянным параметром. Ограничением этого метода является то, что амплитуда и фаза постоянны.

В работе [11] рассмотрена задача оценивания параметров синусоидального сигнала, частота которого изменяется как полиномиальная функция времени. Для решения этой задачи предложена параметризация, позволяющая сформировать уравнения линейной регрессии относительно неизвестных параметров. Оценки формируются с использованием метода динамического расширения регрессора. В [12] предложен метод оценивания частоты синусоидального сигнала с переменной амплитудой. Амплитуда предполагается известной функцией в пределах постоянного множителя, фаза также известна. На первом этапе предложен подход к параметризации синусоидального сигнала и получена модель линейной регрессии третьего порядка. На сле-

дующем этапе создан алгоритм оценивания на основе стандартного градиентного подхода. Общий недостаток этих подходов в том, что амплитуда — известная функция времени.

В настоящей работе рассмотрена задача идентификации параметров синусоидального сигнала с неизвестной переменной амплитудой. Модель линейной регрессии второго порядка построена с использованием операторов задержки. Предположим, что частота, смещение и фаза синусоидального сигнала — постоянные и неизвестные параметры, а амплитуда — переменная и неизвестная функция. Необходимо синтезировать алгоритм идентификации, который обеспечит сходимость настроенной оценки параметров к истинным значениям.

Предложен новый метод идентификации параметров синусоидального сигнала с переменными параметрами, который повысит качество оценивания параметров исходного сигнала, и обеспечит экспоненциальную сходимость к нулю ошибок оценивания. В отличие от аналогичных, в разработанном методе нет ограничения, которое требует измерения переменных частей сигнала. Представленный метод робастный по отношению к шумам, в отличие от аналогов.

Постановка задачи

Рассмотрим измеряемый сигнал

$$y(t) = A(t)\mu e^{-\beta t} \sin(\omega t + \varphi), \quad (1)$$

где μ — постоянный параметр; $A(t)$ — переменная амплитуда; $\beta > 0$ — коэффициент затухания; $\omega \in \mathbb{R}$ и $\varphi \in \mathbb{R}$ — неизвестные частота и фаза; t — время. Параметры μ , β , ω , φ и функция $A(t)$ считаются неизвестными.

Требуется синтезировать алгоритм оценивания частоты $\hat{\omega}(t)$ и коэффициента затухания $\hat{\beta}(t)$, обеспечивающих выполнение условий

$$\lim_{t \rightarrow \infty} (\omega - \hat{\omega}(t)) = 0, \\ \lim_{t \rightarrow \infty} (\beta - \hat{\beta}(t)) = 0.$$

Перепишем уравнение (1) в удобном виде

$$y(t) = A(t)Y(t),$$

где $Y(t) = \mu e^{-\beta t} \sin(\omega t + \varphi)$.

Введем допущение.

Допущение 1. Функция $A(t)$ является выходом линейного генератора

$$\begin{aligned} A(t) &= \mathbf{h}^T \mathbf{\Psi}(t), \\ \dot{\mathbf{\Psi}}(t) &= \mathbf{G} \mathbf{\Psi}(t), \end{aligned}$$

где $\mathbf{\Psi} \in \mathbb{R}^q$ — матрица состояния генератора с неизвестным начальным значением $\mathbf{\Psi}(0)$; $\mathbf{G} \in \mathbb{R}^{q \times q}$ — матрица известных постоянных коэффициентов; $\mathbf{h} \in \mathbb{R}^q$ — вектор соответствующей размерности.

Также допустим, что частота $A(t)$ изменяется по синусоидальному закону, тогда матрица \mathbf{G} и вектор $\mathbf{\Psi}(0)$ примут вид

$$\mathbf{h} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \mathbf{G} = \begin{bmatrix} 0 & 1 \\ -\gamma^2 & 0 \end{bmatrix}, \mathbf{\Psi}(0) = \begin{bmatrix} 0 \\ \theta \end{bmatrix},$$

где γ — известный параметр; θ — неизвестный параметр.

Допущение 2. Функция $Y(t)$ является выходом линейного генератора

$$Y(t) = \mathbf{H}^T \xi(t), \quad (2)$$

$$\dot{\xi}(t) = \mathbf{\Gamma} \xi(t), \quad (3)$$

где $\xi \in \mathbb{R}^p$ — вектор состояния генератора; $\mathbf{\Gamma} \in \mathbb{R}^{p \times p}$ — матрица постоянных коэффициентов; $\mathbf{H} \in \mathbb{R}^p$ — вектор соответствующей размерности.

Для векторно-матричной формы выражения (2)–(3) имеют вид [14]:

$$\xi = \begin{bmatrix} \xi_1 \\ \xi_2 \end{bmatrix}, \mathbf{\Gamma} = \begin{bmatrix} 0 & 1 \\ -\omega^2 - \beta^2 & -2\beta \end{bmatrix}, \mathbf{H} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}.$$

Параметризация синусоидального сигнала

Рассмотрим задачу построения модели линейной регрессии с измеряемыми переменными и вектором постоянных параметров, зависящих от неизвестной частоты ω и коэффициента затухания β .

Наряду с измеряемым сигналом $y(t)$ рассмотрим запаздывающие сигналы

$$y_1(t) = \begin{cases} y(t - d_1), & t \geq d_1, \\ 0, & t < d_1, \end{cases} \quad (4)$$

$$y_2(t) = \begin{cases} y(t - d_2), & t \geq d_2, \\ 0, & t < d_2, \end{cases} \quad (5)$$

где $d_1, d_2 \in \mathbb{R}_+$ — постоянные величины запаздывания.

Для сигналов (4) и (5) запишем соотношения:

$$y_1(t) = A_1(t) Y_1(t), \quad (6)$$

$$y_2(t) = A_2(t) Y_2(t), \quad (7)$$

где

$$A_1(t) = A(t - d_1), A_2(t) = A(t - d_2), \quad (8)$$

$$Y_1(t) = Y(t - d_1), Y_2(t) = Y(t - d_2). \quad (9)$$

Утверждение 1. Измеряемые сигналы (1), (8) и (9) связаны тождеством:

$$\Xi(t) = \mathbf{\Psi}^T(t) \mathbf{\Theta}, \quad (10)$$

где $\Xi \in \mathbb{R}^1$; $\mathbf{\Psi} \in \mathbb{R}^{6 \times 6}$ — измеряемые функции; $\mathbf{\Theta} \in \mathbb{R}^6$ — вектор неизвестных параметров.

Из допущений 1 и 2 запишем функции:

$$A(t) = \mathbf{h}^T e^{\mathbf{G}t} \mathbf{\Psi}(0), \quad (11)$$

$$Y(t) = \mathbf{H}^T e^{\mathbf{\Gamma}t} \xi(0).$$

На основании выражения (8) получим:

$$A_1(t) = A(t - d_1) = \mathbf{h}^T e^{\mathbf{G}t} e^{-\mathbf{G}d_1} \mathbf{\Psi}(0), \quad (12)$$

$$A_2(t) = A(t - d_2) = \mathbf{h}^T e^{\mathbf{G}t} e^{-\mathbf{G}d_2} \mathbf{\Psi}(0), \quad (13)$$

а из выражений (9):

$$Y_1(t) = Y(t - d_1) = \mathbf{H}^T e^{\mathbf{\Gamma}t} e^{-\mathbf{\Gamma}d_1} \xi(0) = \mathbf{H}^T e^{-\mathbf{\Gamma}d_1} \xi(t), \quad (14)$$

$$Y_2(t) = Y(t - d_2) = \mathbf{H}^T e^{\mathbf{\Gamma}t} e^{-\mathbf{\Gamma}d_2} \xi(0) = \mathbf{H}^T e^{-\mathbf{\Gamma}d_2} \xi(t). \quad (15)$$

Поставив (14) и (15) в (6) и (7) соответственно, получим

$$y_1(t) = A_1(t) \mathbf{H}^T e^{-\mathbf{\Gamma}d_1} \xi(t),$$

$$y_2(t) = A_2(t) \mathbf{H}^T e^{-\mathbf{\Gamma}d_2} \xi(t).$$

Утверждение 2. Сигнал (11) связан соотношением:

$$A(t) = \gamma^{-1} \sin(\gamma t) \theta.$$

Доказательство утверждения 2. Преобразуем $e^{\mathbf{G}t}$ в уравнении (11) аналогично подходу, описанному в работе [15]:

$$e^{\mathbf{G}t} = \begin{bmatrix} \cos(\gamma t) & \gamma^{-1} \sin(\gamma t) \\ -\gamma \sin(\gamma t) & \cos(\gamma t) \end{bmatrix}. \quad (16)$$

Поставив соотношение (16) в (11), получим

$$A(t) = \gamma^{-1} \sin(\gamma t) \theta, \quad (17)$$

что и требовалось доказать для утверждения 2.

Аналогично найдем:

$$e^{-\mathbf{G}d_1} = \begin{bmatrix} \cos(\gamma d_1) & -\gamma^{-1} \sin(\gamma d_1) \\ \gamma \sin(\gamma d_1) & \cos(\gamma d_1) \end{bmatrix}, \quad (18)$$

$$e^{-\mathbf{G}d_2} = \begin{bmatrix} \cos(\gamma d_2) & -\gamma^{-1} \sin(\gamma d_2) \\ \gamma \sin(\gamma d_2) & \cos(\gamma d_2) \end{bmatrix}, \quad (19)$$

Для сокращения числа параметров примем $d_2 = 2d_1 = 2d$.

Поставим выражения (17) и (18) в уравнение (12), которое будет иметь вид:

$$A_1(t) = \gamma^{-1} \sin(\gamma t - \gamma d) \theta,$$

а (17) и (19) в уравнение (13):

$$A_2(t) = \gamma^{-1} \sin(\gamma t - 2\gamma d) \theta.$$

Утверждение 3. Сигнал (14) связан соотношением:

$$Y_1(t) = e^{\beta d} [\cos(\omega d) + \beta \omega^{-1} \sin(\omega d) - \omega^{-1} \sin(\omega d)] \xi(t).$$

Доказательство утверждения 3. Применим жорданову форму для преобразователя $e^{-\mathbf{\Gamma}d}$.

Вычислим собственные значения матрицы Γ :

$$\begin{aligned}\det(\Gamma - \lambda \mathbf{I}) &= \begin{vmatrix} 0 - \lambda & 1 \\ -\omega^2 - \beta^2 & -2\beta - \lambda \end{vmatrix} = 0, \\ \Rightarrow \lambda^2 + 2\lambda\gamma + \omega^2 + \beta^2 &= 0 \\ \Rightarrow \lambda_1 = -\beta + i\omega, \lambda_2 = -\beta - i\omega,\end{aligned}$$

где \mathbf{I} — единичная матрица.

Для каждого из собственных значений получим собственные векторы. Для числа $\lambda_1 = -\beta + i\omega$, имеем:

$$\begin{aligned}\begin{bmatrix} \beta - i\omega & 1 \\ -\omega^2 - \beta^2 & -\beta - i\omega \end{bmatrix} \begin{bmatrix} x_1 \\ y_1 \end{bmatrix} &= 0, \Rightarrow \\ \Rightarrow \begin{cases} (\beta - i\omega)x_1 + y_1 = 0 \\ (-\omega^2 - \beta^2)x_1 + (-\beta - i\omega)y_1 = 0 \end{cases} \\ \Rightarrow y_1 &= (i\omega - \beta)x_1.\end{aligned}$$

Полагая $x_1 = 1$, найдем собственный вектор $\mathbf{V}_1 = (x_1, y_1)^T$:

$$x_1 = 1, \Rightarrow y_2 = -\gamma + i\omega, \Rightarrow \mathbf{V}_2 = \begin{bmatrix} 1 \\ -\beta + i\omega \end{bmatrix}$$

и аналогично — вектор $\mathbf{V}_2 = (x_2, y_2)^T$, ассоциированный с собственным значением $\lambda_2 = -\beta - i\omega$, получим

$$\begin{aligned}\begin{bmatrix} \gamma + i\omega & 1 \\ -\omega^2 - \beta^2 & -\beta + i\omega \end{bmatrix} \begin{bmatrix} x_2 \\ y_2 \end{bmatrix} &= 0, \Rightarrow \begin{cases} (\beta + i\omega)x_2 + y_2 = 0 \\ -\omega^2 x_2 + i\omega y_2 = 0 \end{cases} \\ \Rightarrow y_2 &= -(\beta + i\omega)x_2.\end{aligned}$$

При $x_2 = 1$ найдем собственный вектор $\mathbf{V}_2 = (x_2, y_2)^T$:

$$x_2 = 1, \Rightarrow y_2 = -\beta - i\omega, \Rightarrow \mathbf{V}_2 = \begin{bmatrix} 1 \\ -\beta - i\omega \end{bmatrix}.$$

Составим матрицу \mathbf{V} из найденных векторов \mathbf{V}_1 и \mathbf{V}_2 :

$$\mathbf{V} = \begin{bmatrix} 1 & 1 \\ (i\omega - \beta) & -(\beta + i\omega) \end{bmatrix}.$$

Запишем жорданову форму \mathbf{J}_Γ для заданной матрицы Γ , используя форму

$$\mathbf{J}_\Gamma = \mathbf{V}^{-1}\mathbf{\Gamma}\mathbf{V} = \frac{1}{-2i\omega} \begin{bmatrix} -(\beta + i\omega) & -1 \\ -(i\omega - \beta) & 1 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ -\omega^2 - \beta^2 & -2\beta \end{bmatrix} \times$$

$$\times \begin{bmatrix} 1 & 1 \\ (i\omega - \beta) & -(\beta + i\omega) \end{bmatrix},$$

$$\mathbf{J}_\Gamma = \mathbf{V}^{-1}\mathbf{\Gamma}\mathbf{V} = \frac{1}{-2i\omega} \begin{bmatrix} \beta^2 + \omega^2 & \beta - i\omega \\ -\beta^2 - \omega^2 & -\beta - i\omega \end{bmatrix} \times$$

$$\times \begin{bmatrix} 1 & 1 \\ i\omega - \beta & -\beta - i\omega \end{bmatrix},$$

$$\mathbf{J}_\Gamma = \mathbf{V}^{-1}\mathbf{\Gamma}\mathbf{V} = \begin{bmatrix} i\omega - \beta & 0 \\ 0 & -i\omega - \beta \end{bmatrix}.$$

Составим матрицу $e^{\mathbf{J}_\Gamma d}$ в следующем виде

$$e^{\mathbf{J}_\Gamma d} = e^{-\beta d} \begin{bmatrix} e^{i\omega d} & 0 \\ 0 & e^{-i\omega d} \end{bmatrix}.$$

Вычислим матричную экспоненту $e^{\Gamma d}$

$$\begin{aligned}e^{\Gamma d} &= \mathbf{V}e^{\mathbf{J}_\Gamma d}\mathbf{V}^{-1} = e^{-\beta d} \begin{bmatrix} 1 & 1 \\ i\omega - \beta & -i\omega - \beta \end{bmatrix} \begin{bmatrix} e^{i\omega d} & 0 \\ 0 & e^{-i\omega d} \end{bmatrix} \times \\ &\times \frac{1}{-2i\omega} \begin{bmatrix} -i\omega - \beta & -1 \\ i\omega - \beta & 1 \end{bmatrix},\end{aligned}$$

$$e^{\Gamma d} = \frac{e^{-\beta d}}{-2i\omega} \begin{bmatrix} e^{i\omega d} & e^{-i\omega d} \\ (i\omega - \beta)e^{i\omega d} & (-i\omega - \beta)e^{-i\omega d} \end{bmatrix} \begin{bmatrix} -i\omega - \beta & -1 \\ -i\omega + \beta & 1 \end{bmatrix},$$

$$e^{\Gamma d} = e^{-\beta d} \begin{bmatrix} \frac{e^{i\omega d} + e^{-i\omega d}}{2} + \omega^{-1} \frac{e^{i\omega d} - e^{-i\omega d}}{2i} & \\ + \beta\omega^{-1} \frac{e^{i\omega d} - e^{-i\omega d}}{2i} & \\ -(\omega^2 + \beta^2) \omega^{-1} \times \frac{e^{i\omega d} - e^{-i\omega d}}{2i} - & \\ \times \frac{e^{i\omega d} + e^{-i\omega d}}{2} & - \beta\omega^{-1} \frac{e^{i\omega d} - e^{-i\omega d}}{2i} \end{bmatrix}.$$

Экспоненциальные функции $e^{i\omega d}$, $e^{-i\omega d}$ разложим по формуле Эйлера

$$\frac{e^{i\omega d} + e^{-i\omega d}}{2} = \cos(\omega d), \quad \frac{e^{i\omega d} - e^{-i\omega d}}{2i} = \sin(\omega d).$$

Получаем $e^{\Gamma d}$ в виде:

$$\begin{aligned}e^{\Gamma d} &= e^{-\beta d} \begin{bmatrix} \cos(\omega d) + \beta\omega^{-1}\sin(\omega d) & \omega^{-1}\sin(\omega d) \\ -(\omega^2 + \beta^2) \omega^{-1} \times \sin(\omega d) & \cos(\omega d) + \beta\omega^{-1}\sin(\omega d) \end{bmatrix}, \\ e^{\Gamma d} &= e^{\beta d} \begin{bmatrix} \cos(\omega d) + \beta\omega^{-1}\sin(\omega d) & \omega^{-1}\sin(\omega d) \\ -(\omega^2 + \beta^2) \omega^{-1} \times \sin(\omega d) & \cos(\omega d) - \beta\omega^{-1}\sin(\omega d) \end{bmatrix}.\end{aligned}\quad (20)$$

Из уравнений (14) и (20) запишем соотношение вида:

$$Y_1(t) = e^{\beta d} \begin{bmatrix} \cos(\omega d) + \beta\omega^{-1}\sin(\omega d) & -\omega^{-1}\sin(\omega d) \end{bmatrix} \xi(t). \quad (21)$$

Утверждение 3 доказано.

Аналогично найдем:

$$Y_2(t) = e^{2\beta d} \begin{bmatrix} \cos(2\omega d) + \beta\omega^{-1}\sin(2\omega d) & -\omega^{-1}\sin(2\omega d) \end{bmatrix} \xi(t). \quad (22)$$

Подставим уравнения (21), (22) в уравнения (4) и (5) и представим в следующем матричном виде

$$\begin{bmatrix} y_1(t) \\ y_2(t) \end{bmatrix} = \begin{bmatrix} A_1(t)e^{\beta d}(c_1 + \beta\omega^{-1}s_1) & -\omega^{-1}A_1(t)s_1e^{\beta d} \\ A_2(t)e^{2\beta d}(c_2 + \beta\omega^{-1}s_2) & -\omega^{-1}A_2(t)s_2e^{2\beta d} \end{bmatrix} \xi(t),$$

где $s_1 = \sin(\omega d)$; $c_1 = \cos(\omega d)$; $s_2 = \sin(2\omega d)$; $c_2 = \cos(2\omega d)$.

Рассмотрим следующее выражение:

$$\mathbf{B} = \Phi \xi, \quad (23)$$

$$\text{где } \mathbf{B} = \begin{bmatrix} y_1(t) \\ y_2(t) \end{bmatrix},$$

$$\Phi = \begin{bmatrix} A_1(t)e^{\beta d}(c_1 + \beta\omega^{-1}s_1) & -\omega^{-1}A_1(t)s_1e^{\beta d} \\ A_2(t)e^{2\beta d}(c_2 + \beta\omega^{-1}s_2) & -\omega^{-1}A_2(t)s_2e^{2\beta d} \end{bmatrix}.$$

Запишем (23) в виде:

$$\Phi^{-1}\mathbf{B} = \xi = \frac{1}{\det(\Phi)} \text{adj}(\Phi)\mathbf{B} \Rightarrow \text{adj}(\Phi)\mathbf{B} = \det(\Phi)\xi. \quad (24)$$

Умножив уравнение (24) на $A(t)\mathbf{H}^T$, получим

$$A(t)\mathbf{H}^T \text{adj}(\Phi)\mathbf{B} = \det(\Phi)y(t), \quad (25)$$

$$\text{adj}(\Phi) = \begin{bmatrix} -\omega^{-1}A_2(t)s_2e^{2\beta d} & \omega^{-1}A_1(t)s_1e^{\beta d} \\ -A_2(t)e^{2\beta d}(c_2 + \beta\omega^{-1}s_2) & A_1(t)e^{\beta d}(c_1 + \beta\omega^{-1}s_1) \end{bmatrix}, \quad (26)$$

$$\det(\Phi) = -\frac{1}{\omega}A_1(t)A_2(t)e^{3\beta d}s_2c_1 + \frac{1}{\omega}A_1(t)A_2(t)e^{3\beta d}s_1c_2,$$

$$\det(\Phi) = -\frac{1}{\omega}s_1A_1(t)A_2(t)e^{3\beta d}. \quad (27)$$

Обозначим:

$$A(t) = \gamma^{-1}\theta a, \quad (28)$$

$$A_1(t) = \gamma^{-1}\theta a_1, \quad (29)$$

$$A_2(t) = \gamma^{-1}\theta a_2, \quad (30)$$

где $a = \gamma^{-1}\sin(\gamma t)$, $a_1 = \gamma^{-1}\sin(\gamma t - \gamma d)$, $a_2 = \gamma^{-1}\sin(\gamma t - 2\gamma d)$.

Выразим параметры c_2 , s_2 через c_1 , s_1 , используя тригонометрические формулы для двойного угла:

$$c_2 = 2c_1^2 - 1, \quad s_2 = 2c_1s_1. \quad (31)$$

Из выражений (25)–(31) получим уравнение

$$\begin{aligned} -\omega^{-1}e^{2\beta d}2s_1c_1y_1(t)\theta^2aa_2 + \omega^{-1}e^{\beta d}s_1y_2(t)\theta^2aa_1 = \\ = -\omega^{-1}e^{3\beta d}s_1y(t)\theta^2a_1a_2. \end{aligned}$$

Разделив обе части выражения (31) на $-\omega^{-1}e^{\beta d}s_1\theta^2$, запишем

$$2e^{\beta d}c_1y_1(t)aa_2 - y_2(t)aa_1 = y(t)a_1a_2e^{2\beta d}. \quad (32)$$

Выражение (32) представим в виде линейной регрессионной модели (10):

$$\Xi(t) = \Psi^T(t)\Theta, \quad (33)$$

где $\Xi(t) = y_2aa_1$ — измеряемая функция; $\Psi^T(t) = \begin{bmatrix} 2y_1aa_2 \\ -y_2aa_1 \end{bmatrix}$ — регрессор; $\Theta = \begin{bmatrix} e^{\beta d}c_1 \\ e^{2\beta d} \end{bmatrix}$ — вектор неизвестных параметров.

Алгоритм оценивания неизвестных параметров регрессионной модели

Для оценивания параметров вектора Θ модели (33) применим метод динамического расширения регрессора (Dynamic Regressor Extension and Mixing, DREM) [16].

Применим блок запаздывания v_i , $i = \overline{1, q-1}$ для известных элементов регрессионной модели (33):

$$\Xi(t - v_i) = \eta^T(t - v_i)\Theta,$$

где $i = \overline{1, 6}$.

Обозначим

$$\mathfrak{v}_e = \gamma_e\Theta, \quad (34)$$

где $\mathfrak{v}_e = [\Psi(t) \quad \Psi(t - v_1) \quad \dots \quad \Psi(t - v_i)]^T$, $\gamma_e = [\eta_1^T(t) \quad \eta_1^T(t - v_1) \quad \dots \quad \eta_i^T(t - v_i)]$.

Умножив (34) на $\text{adj}(\eta_e(t))$, получим

$$\mathfrak{v}(t) = \Delta(t)\Theta, \quad (35)$$

где $\Delta(t) = \det(\eta_e(t)) \in \mathbb{R}^1$; $\mathfrak{v}(t) = \text{adj}(\eta_e)\mathfrak{v}_e(t) \in \mathbb{R}^2$.

Запишем уравнение (35) покомпонентно:

$$\mathfrak{v}_i(t) = \Delta(t)\Theta_i, \quad i = \overline{1, 2}.$$

Представим алгоритм оценивания параметров Θ_i в виде:

$$\hat{\Theta}_i(t) = -\kappa_i \Delta(t)(\mathfrak{v}_i(t) - \Delta(t)\hat{\Theta}_i),$$

где κ_i — настраиваемый параметр, $i = \overline{1, 2}$.

Алгоритм идентификации параметров исходного сигнала. Для оценивания коэффициента затухания воспользуемся функцией $\ln(\cdot)$ на основе параметра $\hat{\Theta}_2$ из модели (33)

$$\hat{\beta}(t) = \frac{1}{d} \ln \sqrt{\hat{\Theta}_2(t)}.$$

Для оценивания частоты воспользуемся функцией $\arccos(\cdot)$ на основе параметров $\hat{\Theta}_1$ и $\hat{\Theta}_2$ из модели (33)

$$\hat{\omega}(t) = \frac{1}{d} \arccos \left(\frac{\hat{\Theta}_1(t)}{\sqrt{\hat{\Theta}_2(t)}} \right).$$

Математическое моделирование

Рассмотрим результаты численного моделирования, иллюстрирующие эффективность предложенного алгоритма оценивания частоты несмещенного гармонического сигнала с постоянными параметрами. Моделирование выполнено с использованием программной среды MATLAB Simulink.

Пример 1. Рассмотрим сигнал вида

$$y(t) = A(t)e^{-0,2t}\sin(5t - 1). \quad (36)$$

Для функции $A(t)$ принимаются следующие параметры

$$\mathbf{G} = \begin{bmatrix} 0 & 1 \\ -4 & 0 \end{bmatrix}, \quad \Psi(0) = \begin{bmatrix} 0 \\ 4 \end{bmatrix}.$$

Приведем значения параметров для предложенного метода: $d = 0,1$, $v = 0,2$, $\kappa_i = 10^8$.

На рис. 1 показаны результаты оценивания параметров сигнала (36).

Пример 2. Рассмотрим сигнал вида

$$y(t) = A(t)e^{-1t}\sin(10t + 1). \quad (37)$$

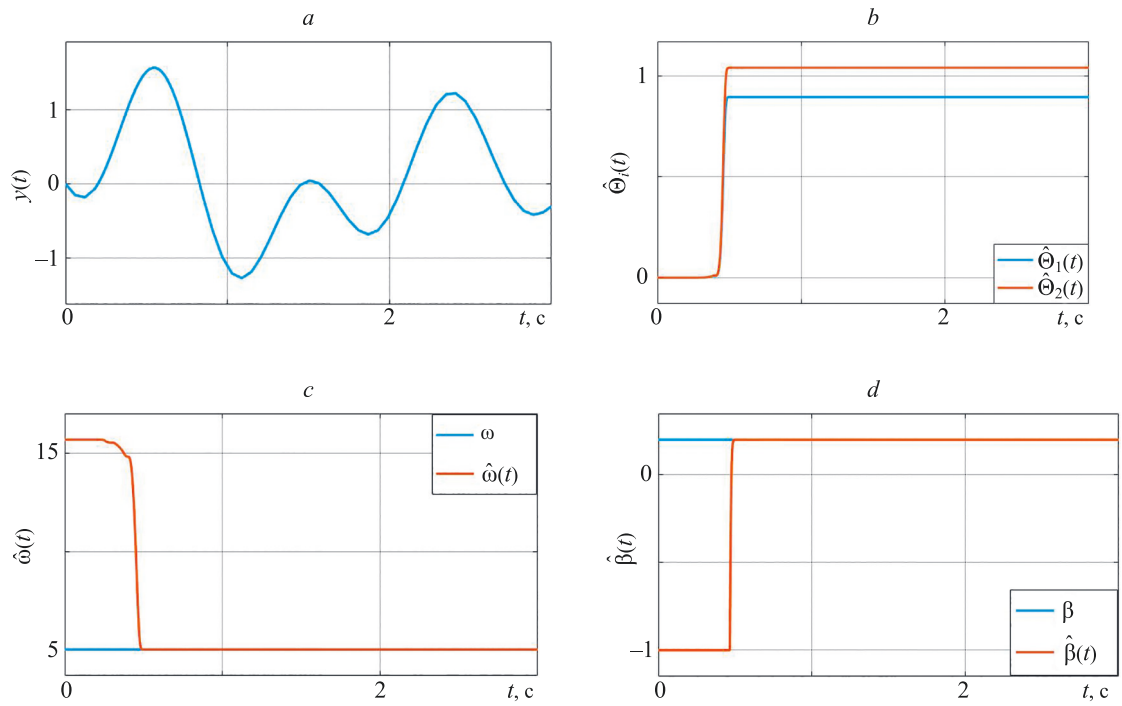


Рис. 1. График сигнала $y(t)$ (a); графики оценок: параметров $\hat{\theta}_i(t)$ (b), частоты $\hat{\omega}(t)$ при $\omega = 5$ (c) и коэффициента затухания $\hat{\beta}(t)$ при $\beta = 0,2$ (d)

Fig. 1. Signal waveform $y(t)$ (a); parameter estimation plot $\hat{\theta}_i(t)$ (b); frequency estimation plot $\hat{\omega}(t)$ at $\omega = 5$ (c); damping factor estimation plot $\hat{\beta}(t)$ at $\beta = 0.2$ (d)

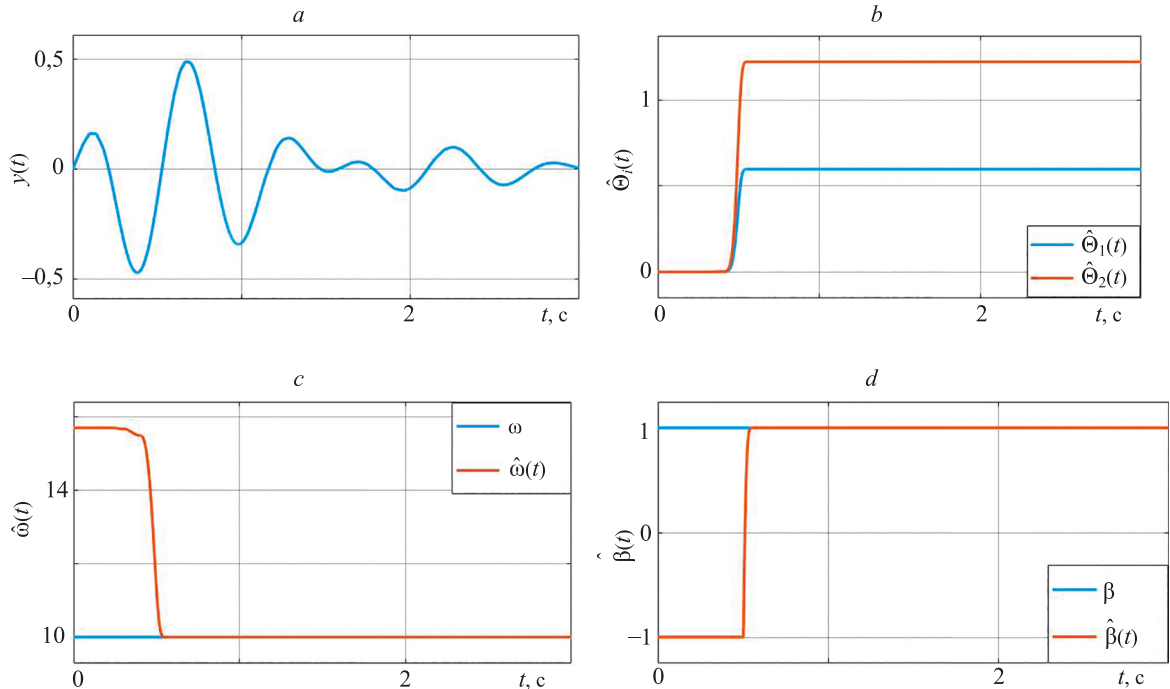


Рис. 2. График сигнала $y(t)$ (a); графики оценок: параметров $\hat{\theta}_i(t)$ (b), частоты $\hat{\omega}(t)$ при $\omega = 3$ (c) и коэффициента затухания $\hat{\beta}(t)$ при $\beta = 1$ (d)

Fig. 2. Signal waveform $y(t)$ (a); parameter estimation $\hat{\theta}_i(t)$ (b); frequency estimation plot $\hat{\omega}(t)$ at $\omega = 3$ (c); damping factor estimation plot $\hat{\beta}(t)$ at $\beta = 1$ (d)

Для функции $A(t)$ используем параметры

$$\mathbf{G} = \begin{bmatrix} 0 & 1 \\ -9 & 0 \end{bmatrix}, \quad \Psi(0) = \begin{bmatrix} 0 \\ 2 \end{bmatrix}.$$

Приведем значения параметров для предложенного метода: $d = 0,1$, $v = 0,2$, $\kappa_i = 10^8$.

На рис. 2 показаны результаты оценивания параметров сигнала (37).

На рис. 1, 2 приведены результаты оценивания параметров сигналов (36) и (37). Как видно из графиков, предложенный алгоритм оценивания обеспечивает экспоненциальную сходимость к истинным значениям оценивания параметров сигнала $y(t)$.

Заключение

В работе предложен новый метод идентификации параметров синусоидального сигнала с переменной неизвестной амплитудой. Амплитуда синусоидального сигнала изменяется по гармоническому закону с неизвестными параметрами. Предложен подход для параметризации синусоидального сигнала с использованием операторов запаздывания и жорданова форма матрицы. Выполнена оценка вектора параметров регрессионных моделей с использованием метода динамического расширения регрессора. Результаты моделирования подтверждают сходимость оценивания параметров к истинным значениям. В дальнейшем планируется расширить область применения алгоритма на случай переменных частот и фаз.

Литература

1. Pyrkin A.A., Bobtsov A.A., Efimov D., Zolghadri A. Frequency estimation for periodical signal with noise in finite time // *Proc. of the 50th IEEE Conference on Decision and Control and European Control Conference*, 2011. P. 3646–3651. <https://doi.org/10.1109/CDC.2011.6160655>
2. Aranovskiy S., Bobtsov A., Ortega R., Pyrkin A. Improved transients in multiple frequencies estimation via dynamic regressor extension and mixing // *IFAC-PapersOnLine*, 2016. V. 49. N 13. P. 99–104. <https://doi.org/10.1016/j.ifacol.2016.07.934>
3. Bobtsov A., Lyamin A., Romasheva D. Algorithm of parameters' identification of polyharmonic function // *IFAC Proceedings Volumes*, 2002. V. 35. N 1. P. 439–443. <https://doi.org/10.3182/20020721-6-ES-1901.01059>
4. Marino R., Tomei R. Global estimation of n unknown frequencies // *IEEE Transactions on Automatic Control*, 2002. V. 47. N 8. P. 1324–1328. <https://doi.org/10.1109/TAC.2002.800761>
5. Bodson M., Douglas S.C. Adaptive algorithms for the rejection of sinusoidal disturbances with unknown frequency // *Automatica*, 1997. V. 33. N 12. P. 2213–2221. [https://doi.org/10.1016/S0005-1098\(97\)00149-0](https://doi.org/10.1016/S0005-1098(97)00149-0)
6. Khac T., Vlasov S.M., Iureva R.A. Estimating the frequency of the sinusoidal signal using the parameterization based on the delay operators // *Proc. of the 18th International Conference on Informatics in Control, Automation and Robotics (ICINCO)*, 2021. P. 656–660. <https://doi.org/10.5220/0010536506560660>
7. Севастеева Е.С., Чернов В.А., Бобцов А.А. Алгоритм увеличения скорости идентификации частоты синусоидального сигнала // *Известия вузов. Приборостроение*, 2019. Т. 62. № 9. С. 767–771. <https://doi.org/10.17586/0021-3454-2019-62-9-767-771>
8. Бобцов А.А., Николаев Н.А., Оськина О.В., Низовцев С.И. Идентификация нестационарного параметра незашумленного синусоидального сигнала // *Автоматика и телемеханика*, 2022. № 7. С. 137–151. <https://doi.org/10.31857/S000523102207008X>
9. Низовцев С.И., Шаветов С.В., Пыркин А.А. Алгоритм оценивания переменной частоты синусоидального сигнала // *Известия вузов. Приборостроение*, 2021. Т. 64. № 9. С. 789–793. <https://doi.org/10.17586/0021-3454-2021-64-9-789-793>
10. Ле Ван Туан, Коротина М.М., Бобцов А.А., Арановский С.В. Алгоритм идентификации линейно меняющейся частоты синусоидального сигнала // *Научно-технический вестник информационных технологий, механики и оптики*, 2019. Т. 19. № 1. С. 52–58. <https://doi.org/10.17586/2226-1494-2019-19-1-52-58>
11. Коротина М.М., Арановский С.В., Бобцов А.А. Оценивание параметров синусоидального сигнала со степенной функцией частоты // *Известия вузов. Приборостроение*, 2020. Т. 63. № 6. С. 507–514. <https://doi.org/10.17586/0021-3454-2020-63-6-507-514>
12. Vedyakov A.A., Vediaikova A.O., Bobtsov A.A., Pyrkin A.A., Aranovskiy S.V. Frequency estimation of a sinusoidal signal with time-varying amplitude // *IFAC-PapersOnLine*, 2017. V. 50. N 1. P. 12880–12885. <https://doi.org/10.1016/j.ifacol.2017.08.1940>

References

1. Pyrkin A.A., Bobtsov A.A., Efimov D., Zolghadri A. Frequency estimation for periodical signal with noise in finite time. *Proc. of the 50th IEEE Conference on Decision and Control and European Control Conference*, 2011, pp. 3646–3651. <https://doi.org/10.1109/CDC.2011.6160655>
2. Aranovskiy S., Bobtsov A., Ortega R., Pyrkin A. Improved transients in multiple frequencies estimation via dynamic regressor extension and mixing. *IFAC-PapersOnLine*, 2016, vol. 49, no. 13, pp. 99–104. doi.org/10.1016/j.ifacol.2016.07.934
3. Bobtsov A., Lyamin A., Romasheva D. Algorithm of parameters' identification of polyharmonic function. *IFAC Proceedings Volumes*, 2002, vol. 35, no. 1, pp. 439–443. <https://doi.org/10.3182/20020721-6-ES-1901.01059>
4. Marino R., Tomei R. Global estimation of n unknown frequencies. *IEEE Transactions on Automatic Control*, 2002, vol. 47, no. 8, pp. 1324–1328. <https://doi.org/10.1109/TAC.2002.800761>
5. Bodson M., Douglas S.C. Adaptive algorithms for the rejection of sinusoidal disturbances with unknown frequency. *Automatica*, 1997, vol. 33, no. 12, pp. 2213–2221. [doi.org/10.1016/S0005-1098\(97\)00149-0](https://doi.org/10.1016/S0005-1098(97)00149-0)
6. Khac T., Vlasov S.M., Iureva R.A. Estimating the frequency of the sinusoidal signal using the parameterization based on the delay operators. *Proc. of the 18th International Conference on Informatics in Control, Automation and Robotics (ICINCO)*, 2021, pp. 656–660. <https://doi.org/10.5220/0010536506560660>
7. Sevasteeva E.S., Chernov V.A., Bobtsov A.A. Algorithm for increasing the speed of sinusoidal signal frequency identification. *Journal of Instrument Engineering*, 2019, vol. 62, no. 9, pp. 767–771. (in Russian). <https://doi.org/10.17586/0021-3454-2019-62-9-767-771>
8. Bobtsov A.A., Nikolaev N.A., Oskina O.V., Nizovtsev S.I. Identification of a time-varying parameter of a noiseless sinusoidal signal. *Automation and Remote Control*, 2022, vol. 83, no. 7, pp. 1123–1135. <https://doi.org/10.1134/s0005117922070086>
9. Nizovtsev S.I., Shavetov S.V., Pyrkin A.A. Algorithm for estimating the variable frequency of sinusoidal signal. *Journal of Instrument Engineering*, 2021, vol. 64, no. 9, pp. 789–793. (in Russian). <https://doi.org/10.17586/0021-3454-2021-64-9-789-793>
10. Tuan Le Van, Korotina M.M., Bobtsov A.A., Aranovskiy S.V. New identification algorithm for linearly varying frequency of sinusoidal signal. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 2019, vol. 19, no. 1, pp. 52–58. (in Russian). <https://doi.org/10.17586/2226-1494-2019-19-1-52-58>
11. Korotina M.M., Aranovskiy S.V., Bobtsov A.A. Parameters estimation of sinusoidal signal with power-law frequency dependence. *Journal of Instrument Engineering*, 2020, vol. 63, no. 6, pp. 507–514. (in Russian). <https://doi.org/10.17586/0021-3454-2020-63-6-507-514>
12. Vedyakov A.A., Vediaikova A.O., Bobtsov A.A., Pyrkin A.A., Aranovskiy S.V. Frequency estimation of a sinusoidal signal with time-varying amplitude. *IFAC-PapersOnLine*, 2017, vol. 50, no. 1, pp. 12880–12885. <https://doi.org/10.1016/j.ifacol.2017.08.1940>

13. Vedyakov A.A., VEDIKOVA A.O., Bobtsov A.A., Pyrkin A.A., Kakanov M.A. Frequency estimation of a sinusoidal signal with time-varying amplitude and phase // *IFAC-PapersOnLine*. 2018. V. 51. N 32. P. 663–668. <https://doi.org/10.1016/j.ifacol.2018.11.501>
14. Нгуен Х., Власов С.М., Пыркин А.А. Алгоритм идентификации параметров синусоидального сигнала с экспоненциально затухающей амплитудой // *Мехатроника, автоматизация, управление*. 2022. Т. 23. № 3. С. 125–131. <https://doi.org/10.17587/mau.23.125-131>
15. Нгуен Х., Власов С.М., Пыркин А.А., Попков И.В. Новый алгоритм идентификации частоты синусоидального сигнала с постоянными параметрами // *Научно-технический вестник информационных технологий, механики и оптики*. 2022. Т. 22. № 1. С. 18–24. <https://doi.org/10.17586/2226-1494-2022-22-1-18-24>
16. Aranovskiy S., Bobsov A., Ortega R., Pyrkin A. Performance enhancement of parameter estimators via dynamic regressor extension and mixing // *IEEE Transactions on Automatic Control*. 2017. V. 62. N 7. P. 3546–3550. <https://doi.org/10.1109/tac.2016.2614889>
13. Vedyakov A.A., VEDIKOVA A.O., Bobtsov A.A., Pyrkin A.A., Kakanov M.A. Frequency estimation of a sinusoidal signal with time-varying amplitude and phase. *IFAC-PapersOnLine*, 2018, vol. 51, no. 32, pp. 663–668. <https://doi.org/10.1016/j.ifacol.2018.11.501>
14. Nguyen Kh., Vlasov S.M., Pyrkin A.A. Algorithm for identification of parameters sinusoidal signal with the exponentially damping amplitude. *Mekhatronika, Avtomatizatsiya, Upravlenie*, 2022, vol. 23, no. 3, pp. 125–131. (in Russian). <https://doi.org/10.17587/mau.23.125-131>
15. Nguyen Kh.T., Vlasov S.M., Pyrkin A.A., Popkov I.V. A new algorithm for the identification of sinusoidal signal frequency with constant parameters. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 2022, vol. 22, no. 1, pp. 18–24. (in Russian). <https://doi.org/10.17586/2226-1494-2022-22-1-18-24>
16. Aranovskiy S., Bobsov A., Ortega R., Pyrkin A. Performance enhancement of parameter estimators via dynamic regressor extension and mixing. *IEEE Transactions on Automatic Control*, 2017, vol. 62, no. 7, pp. 3546–3550. <https://doi.org/10.1109/tac.2016.2614889>

Авторы

Нгуен Хак Тунг — аспирант, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, [sc 57222389082](https://orcid.org/0000-0001-6430-1927), <https://orcid.org/0000-0001-6430-1927>, nguyenkhactunghvq1994@gmail.com

Власов Сергей Михайлович — кандидат технических наук, доцент, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, [sc 55355689600](https://orcid.org/0000-0002-8345-7553), <https://orcid.org/0000-0002-8345-7553>, smvlasov@itmo.ru

Authors

Khac T. Nguyen — PhD student, ITMO University, Saint Petersburg, 197101, Russian Federation, [sc 57222389082](https://orcid.org/0000-0001-6430-1927), <https://orcid.org/0000-0001-6430-1927>, nguyenkhactunghvq1994@gmail.com

Sergey M. Vlasov — PhD, Associate Professor, ITMO University, Saint Petersburg, 197101, Russian Federation, [sc 55355689600](https://orcid.org/0000-0002-8345-7553), <https://orcid.org/0000-0002-8345-7553>, smvlasov@itmo.ru

Статья поступила в редакцию 18.11.2022
Одобрена после рецензирования 22.12.2022
Принята к печати 16.01.2023

Received 18.11.2022
Approved after reviewing 22.12.2022
Accepted 16.01.2023



Работа доступна по лицензии
Creative Commons
«Attribution-NonCommercial»

НОВЫЕ МАТЕРИАЛЫ И НАНОТЕХНОЛОГИИ MATERIAL SCIENCE AND NANOTECHNOLOGIES

doi: 10.17586/2226-1494-2023-23-1-62-67

Improvement of the automatic temperature stabilisation process in the cryovacuum unit

Oleg Yu. Golikov^{1✉}, Darkhan Yerezhep², Dmitriy Yu. Sokolov³

^{1,2,3} Al-Farabi Kazakh National University, Almaty, 050040, Republic of Kazakhstan

¹ golikov@physics.kz✉, <https://orcid.org/0000-0002-6691-8346>

² darhan_13@physics.kz, <https://orcid.org/0000-0002-2232-2911>

³ yasnyisokol@gmail.com, <https://orcid.org/0000-0001-7966-1140>

Abstract

This study concerns the issues of temperature stabilization in units used to research the properties of molecules at low and ultra-low temperatures. This research is relevant due to the need to increase the speed and accuracy of the data obtained. Using the LabView graphical programming environment tools, a control program was created for the LakeShore 325 thermocontroller which reacts when the current temperature is close to the control point temperature set by the researcher. By adding controls for the heating element power and PID controller boot times, it is possible to use them more flexibly. The method was verified for the temperature control points of 40 K, 100 K, 150 K and 200 K. A comparison of the proposed temperature stabilization program with the standard PID controller solution demonstrates the advantages of the former. The speed of reaching the control points was doubled. The digitalization of the LakeShore 325 thermocontroller makes it possible to work further on improving temperature stabilization. The resulting increase in the accuracy–time stabilization ratio makes it possible for those who conduct low-temperature experiments to improve the quality of their measurements dramatically. The introduction of a digital version of the temperature control device opens up possibilities for further automation of cryovacuum units by linking the thermal control program with other programs, for example, recording the spectra at specific temperature values.

Keywords

automation, temperature control, low temperatures, PID-controllers, IR spectroscopy, programming

Acknowledgements

These studies have been carried out with the financial support of the Ministry of Education and Science of the Republic of Kazakhstan under grant AP08855681.

For citation: Golikov O.Yu., Yerezhep D., Sokolov D.Yu. Improvement of the automatic temperature stabilisation process in the cryovacuum unit. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 2023, vol. 23, no. 1, pp. 62–67. doi: 10.17586/2226-1494-2023-23-1-62-67

УДК 681.5.073; 536.581.3; 536.587

Улучшение процесса автоматической стабилизации температуры в криовакуумной установке

Олег Юрьевич Голиков^{1✉}, Дархан Ережеп², Дмитрий Юрьевич Соколов³

^{1,2,3} Казахский национальный университет им. аль-Фараби, Алматы, 050040, Республика Казахстан

¹ golikov@physics.kz✉, <https://orcid.org/0000-0002-6691-8346>

² darhan_13@physics.kz, <https://orcid.org/0000-0002-2232-2911>

³ yasnyisokol@gmail.com, <https://orcid.org/0000-0001-7966-1140>

Аннотация

Предмет исследования. Рассмотрены вопросы температурной стабилизации в установках, предназначенных для выполнения исследований свойств молекул при низких и сверхнизких температурах. Актуальность работы обусловлена необходимостью повышения скорости и точности получаемых данных, на которые в основном оказывает влияние температура исследования. **Метод.** С помощью инструментов программирования графической среды LabView создана управляющая программа для термоконтроллера LakeShore 325, реагирующая на приближение (рабочей) температуры к температуре (заданной) контрольной точки. Добавление элементов управления мощностью нагревательного элемента и временем включения PID-регулятора позволяет использовать их более гибко. Проведена верификация метода стабилизации для контрольных точек температуры 40, 100, 150 и 200 К. **Основные результаты.** Сравнение предложенной программы стабилизации температуры со стандартным решением в виде PID-регулятора показало его преимущество. Получено увеличение скорости достижения контрольных точек до двух порядков. Цифровизация термоконтроллера LakeShore 325 дала возможность выполнять дальнейшие работы по совершенствованию температурной стабилизации. **Практическая значимость.** Полученное увеличение соотношения точность–время при стабилизации позволило в разы улучшить качество проводимых измерений в области низких температур. Внедрение цифровой версии терморегулирующего прибора открывает возможности для дальнейшей автоматизации криовакуумных установок с помощью объединения программы термоконтроля с другими программами, регистрирующими, например, спектры при определенных значениях температуры.

Ключевые слова

автоматизация, контроль температуры, низкие температуры, PID-регуляторы, ИК спектроскопия, программирование

Благодарности

Работа была выполнена при финансовой поддержке Министерства образования и науки Республики Казахстан в рамках гранта AP08855681.

Ссылка для цитирования: Голиков О.Ю., Ережеп Д., Соколов Д.Ю. Улучшение процесса автоматической стабилизации температуры в криовакуумной установке // Научно-технический вестник информационных технологий, механики и оптики. 2023. Т. 23, № 1. С. 62–67 (на англ. яз.). doi: 10.17586/2226-1494-2023-23-1-62-67

Introduction

Over the last decades, studying the behavior of various molecules at low and ultra-low temperatures has been of great scientific interest [1, 2]. Firstly, observing molecular complexes at negative temperatures allows the thermodynamics of the processes involved to be determined, the simplest stages of complex reactions to be determined, and the selectivity of the processes to be ensured. The processes of the formation of glass-like states of various molecules are important and relevant [3–7]. Secondly, ultra-low temperatures significantly affect the physical and mechanical (ultimate strength and endurance) as well as thermophysical (heat capacity, thermal conductivity, linear thermal expansion coefficient) properties of some substances [8, 9]. Due to low temperatures, the material experiences brittle fracture or changes one or more of its thermophysical properties. And thirdly, the study of infrared (IR) spectra of molecules and molecular compounds at low temperatures allows us to answer a number of questions and describe a number of processes occurring in open space [10–21].

So far, astrochemical studies conducted in ground-based and space laboratories have provided IR evidence for about a dozen extraterrestrial molecules and ions in the solid phase [19]. H₂O ice and ice mixtures consisting primarily of H₂O compounds have been found to be important constituents of interstellar particles, comets and a number of planetary satellites [21].

Identification of extraterrestrial H₂O [14, 20], CH₃OH [11, 12], C₂H₅OH [13, 15, 16], NH₃ [22], CO₂ [23] and CH₄ in their condensed state was performed by comparing the spectra taken in the near- and mid-infrared ranges (100 m⁻¹ to 4 m⁻¹) with the spectra of the same molecules obtained in the ground-based laboratories. The obtained IR spectra of low molecular weight amorphous and crystalline solids, focusing on amorphous ice below 100 K, are used to study the low-temperature chemistry of interstellar clouds and extraterrestrial solar system objects such as ice-covered moons and trans-Neptunian objects [17, 18].

In spite of the fact that the IR spectra of molecules in various low-temperature ranges have been studied, an important challenge is the stabilization of the temperature in the research unit and the minimum temperature step of obtaining the IR spectra of the sample under study. At the moment, the problem of stabilization is solved by using a proportional–integral–derivative (PID) temperature controllers with a heating element until the temperature points set by the experimenter are reached. However, such temperature controllers are not flexible enough and often lack a graphical interface and special software that would make it possible to configure them more flexibly for specific tasks. As a result, the accuracy of the experiment is negatively affected, and it is not possible to obtain infrared spectra at certain temperatures or with small temperature steps; or it takes the researcher a lot of time to do so.

The aim of this research is the temperature stabilization of the cryogenic vacuum spectrophotometer unit at Al-

Farabi Kazakh National University in order to solve the problems highlighted above. The results pursued include reducing the time of the experiment, reducing the temperature step of obtaining IR spectra of various molecules at low temperatures and creating more flexible conditions for future experiments.

Methodology

Experiments to verify the work done were performed on a universal cryogenic vacuum spectrophotometer unit in the cryolaboratory at Al-Farabi Kazakh National University. LakeShore 325 thermal controller (LakeShore Cryotronics, USA) is used as a temperature controller in this cryovacuum unit. This temperature controller model is a two-channel regulator, capable of working with various types of temperature sensors (diodes, resistance temperature detectors, thermocouples). In addition to the sensors, the controller also features two independent PID control loops with 25W and 2W heater output, controlling either 50-ohm or 25-ohm loads for optimal flexibility in temperature control of the cryogenic unit.

The PID controller integrated into the thermal controller plays a special role in achieving our objectives as it is used for the initial stabilization of the heating element temperature. The PID controller is a feedback device in the control loop of the system which generates the control signal to produce an output signal with the required accuracy. The control signal is the sum of three other signals: the error signal (proportional to the difference between the input and feedback signals), the integral of the error signal and its derivative.

Therefore, the output signal of the controller u is defined as

$$u(t) = P + I + D = K_p e(t) + K_i \int_0^t e(\tau) d\tau + K_d \frac{de}{dt},$$

$$e(t) = r(t) - y(t),$$

where $r(t)$ is the input signal; $y(t)$ is the output sum signal; $e(t)$ is the discrepancy or deviation of the value from the original signal; P is the difference component; I is the time-integral component; D is the time-differentiation component; t is the current moment in time; τ is the time step; and K_p , K_i and K_d are the amplification factors of the corresponding controller components.

Despite the advantages and seemingly complete operability of the PID controller as a stabilizing component of the temperature controller, it was observed during the experiments that it took quite a long time to reach the set temperature. Besides, when approaching the set temperature, significant fluctuations were observed near it, sometimes up to 5 K, which can cause significant deviations in the final results when conducting an experiment. The hypothesis was made that the insufficient accuracy and speed of the temperature stabilization performed by the PID controller is related to its inability to “predict” the behavior of the heating element within a wide range. Most likely due to the uneven heating after switching off the cooling unit of the machine, the PID controller is unable to reduce the

heating power in time, constantly “jumping over” the set temperature.

Therefore, the automation element to be introduced into the system must first address the issue of predicting the speed at which the sample substrate is heated in order to minimize the time it takes for the PID controller to stabilize the temperature near the control point set by the experimenter. So, this automation element must have direct control over the LakeShore 325 thermal controller, be able to change the power of the heater, the gains of the PID controller and, if possible, switch on and off the McMahon machine that is used to lower the temperature of the substrate and the experimental chamber of the unit.

The software suite LabVIEW (National Instruments, USA) was chosen as the environment in which the automation would be performed. This development environment allows executing programs written in graphical programming language G. This choice was motivated by the simplicity of the development of applications in the LabVIEW environment, a wide range of features as well as the availability of a large number of ready solutions, which can simplify the process of writing the program required to fulfill the task.

Results

The stabilization of the PID controller was improved by adding elements to the program that react to the current temperature of the heated substrate approaching the temperature set by the experimenter. It is known that PID controllers also react to approaching the control point by reducing the power of the heater if P , I and D parameters are properly adjusted. However, this process either takes too long or is characterized by “jumping over” the control point (up to 1.5–2 K) followed by harmonic equalization of temperature. Both options are not good enough for our research carried out on a cryogenic vacuum spectrophotometer, since in the first case, the duration of the experiment would be significantly increased by the slow speed of reaching the control points, and in the second case, the spectrum close to the control point would be obtained with fluctuations of more than 1 K which is also undesirable.

Thus, the proposed solution solves the two disadvantages of the options mentioned above: it allows reaching the reference temperature quickly and at the same time, makes the fluctuations near it less severe. This is achieved by activating the PID controller later in relation to the modes with the default settings of the P , I and D parameters. As a result, the control temperature is reached more quickly (due to the longer operation of the heater at 100 % power). The PID controller is activated by comparing the control temperature with the current temperature, incremented by a small value, which has been calculated experimentally and assumed to be 0.5 K. When the current temperature reaches the reference temperature with an increment value, the heating power is reset to 5 % and then, after a short delay, the PID-controller is activated. After the power of the heater is reduced, the temperature of the substrate starts to decrease, but the activated PID controller soon stabilizes the temperature close to the set control point.

To confirm the improved performance of the modified temperature stabilization algorithm, it is necessary to compare it with the traditional method of regulation using the PID controller integrated in the LakeShore 325 thermocontroller. It is important to compare the performance of the two stabilizing components over a wide range, capturing both low (below 100 K) and high (100–200 K) temperature ranges. Fig. 1 and 2 show the graphs comparing the temperature stabilization of the standard PID controller of the LakeShore 325 thermocontroller with the default settings of $P = 50$, $I = 20$, and $D = 5$ and the improved (i-PID) controller with the adjustments mentioned above for the temperatures of 40 K, 100 K, 150 K, and 200 K. The temperature-time dependencies were obtained 10 K before the control point in order to investigate a more complex stabilization process than with a small step.

As can be seen from the graphs, in three of the four cases, the heating up to the control point was faster with the improved PID controller. Slower heating was observed only when the reference point of 100 K was reached. Obviously, faster heating of the substrate is achieved by activating the

PID controller later, thus keeping its power at 100 % until the very last moment, and by using it more flexibly close to the control point.

In all cases, stabilization near the control point was faster with the improved PID controller. The advantage of the i-PID controller is particularly noticeable in Fig. 1 where stabilization near 40 K was achieved about three times more quickly than with the standard PID controller. The difference in the speed of reaching the control points in Fig. 1 and 2 is presented in the form of timestamps. It can be seen that in the case of stabilization to 40 K, the method has an advantage of 18 seconds (Fig. 1, *a*); for stabilization to 100 K, it has an advantage of 165 seconds (Fig. 1, *b*); for 150 K, it has an advantage of 60 seconds (Fig. 2, *a*); and for 200 K, it has an advantage of 70 seconds (Fig. 2, *b*).

Reaching the 100 K and 200 K control points (Fig. 1, *b* and Fig. 2, *b*) with the i-PID controller was about twice as fast as with the standard PID controller. Once the control point of 150 K was reached, there was only a small advantage in using the i-PID controller — primarily the smaller deviation. The faster stabilization of the i-PID controller can be explained by the correctly selected timing

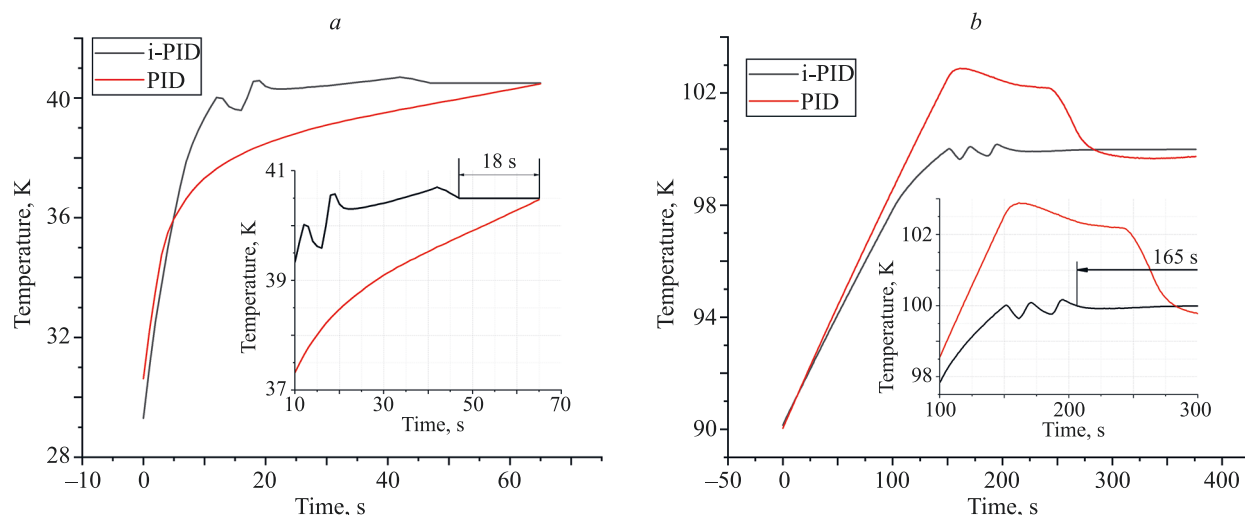


Fig. 1. Standard (PID) and improved (i-PID) thermocontrollers stabilization speed close to the control points of 40 K (*a*) and 100 K (*b*)

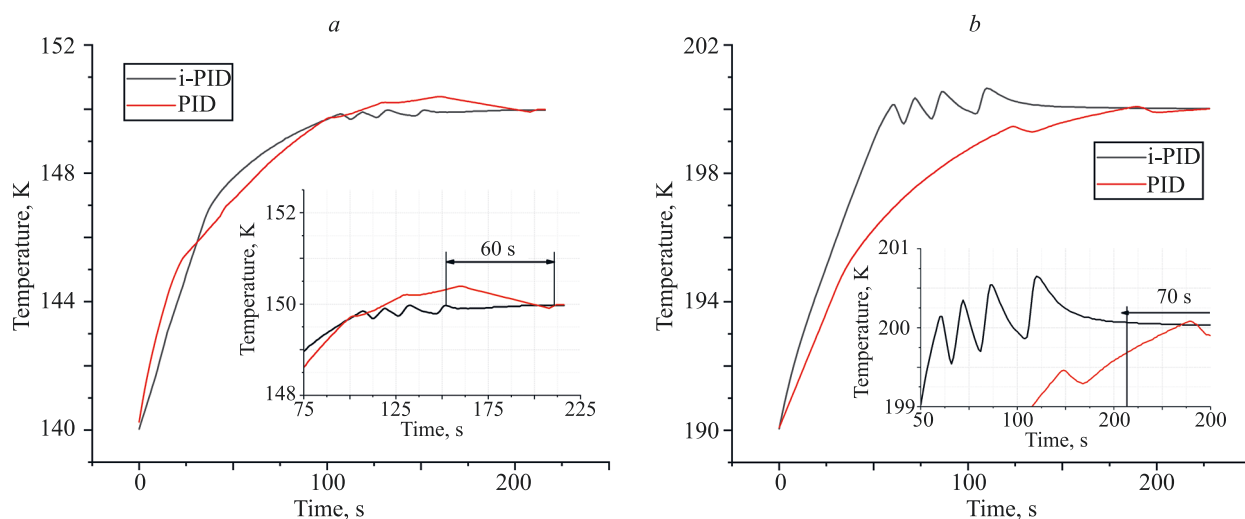


Fig. 2. Standard (PID) and improved (i-PID) thermocontrollers stabilization speed close to the control points of 150 K (*a*) and 200 K (*b*)

of the increase and decrease of the power of the heating element, simultaneously with controlling the PID controller by turning it on and off at the right time.

It is worth noting that it is also possible to reach the control points more quickly when using a standard PID controller by changing the P and I values and removing the D component. However, with this approach, the fluctuation was found to be too long and significant (about 1–2 K), this prevented the IR spectra from being obtained immediately after reaching the control point and required an equally long wait for the stabilization as with the option of the parameters $P = 50$, $I = 20$ and $D = 5$ explored in this article.

Conclusion

The results achieved in this research may seem insignificant. However, for further work on the automation and improvement of the operation of the experimental cryovacuum unit, the obtained gain of 20–160 seconds

is significant, not to mention obtaining the temperature change not exceeding 0.5 K at the moment of reaching the control points.

Despite the versatility and accuracy of stabilization, the PID controller can still be improved by incorporating a control element capable of switching the PID controller on and off within designated ranges and independently (without relying on the PID controller) changing the power of the heating element. This has been demonstrated in this research.

An important issue is the response of the demonstrated i-PID controller to the heating step value (the temperature range between the current temperature and the set control point) on which the stabilization time strongly depends. In this research, only the case of a relatively large heating step has been considered implying more complex stabilization conditions. Nevertheless, it is worth investigating whether it is possible to reach and stabilize the temperature near the control point even faster with smaller heating steps by varying the selected stabilization factor.

References

1. Lee Y., Halperin W.P. Recent progress and new challenges in quantum fluids and solids. *Journal of Low Temperature Physics*, 2017, vol. 189, no. 1, pp. 1–14. <https://doi.org/10.1007/s10909-017-1800-4>
2. Jones A.T., Scheller C.P., Prance J.R., Kalyoncu Y.B., Zumbühl D.M., Haley R.P. Progress in cooling nanoelectronic devices to ultra-low temperatures. *Journal of Low Temperature Physics*, 2020, vol. 201, no. 5, pp. 772–802. <https://doi.org/10.1007/s10909-020-02472-9>
3. Debenedetti P.G., Stillinger F.H. Supercooled liquids and the glass transition. *Nature*, 2001, vol. 410, no. 6825, pp. 259–267. <https://doi.org/10.1038/35065704>
4. Cavagna A. Supercooled liquids for pedestrians. *Physics Reports*, 2009, vol. 476, no. 4–6, pp. 51–124. <https://doi.org/10.1016/j.physrep.2009.03.003>
5. Ediger M.D., Angell C.A., Nagel S.R. Supercooled liquids and glasses. *The Journal of Physical Chemistry*, 1996, vol. 100, no. 31, pp. 13200–13212. <https://doi.org/10.1021/jp953538d>
6. Hodge I.M. Enthalpy relaxation and recovery in amorphous materials. *Journal of Non-Crystalline Solids*, 1994, vol. 169, no. 3, pp. 211–266. [https://doi.org/10.1016/0022-3093\(94\)90321-2](https://doi.org/10.1016/0022-3093(94)90321-2)
7. Yokoyama D. Molecular orientation in small-molecule organic light-emitting diodes. *Journal of Materials Chemistry*, 2011, vol. 21, no. 48, pp. 19187–19202. <https://doi.org/10.1039/C1JM13417E>
8. Galliou S., Imbaud J., Goryachev M., Bourquin R., Abbé P. Losses in high quality quartz crystal resonators at cryogenic temperatures. *Applied Physics Letters*, 2011, vol. 98, no. 9, pp. 091911. <https://doi.org/10.1063/1.3559611>
9. Buehler W.J., Gilfrich J.V., Wiley R.C. Effect of low-temperature phase changes on the mechanical properties of alloys near composition TiNi. *Journal of Applied Physics*, 1963, vol. 34, no. 5, pp. 1475–1477. <https://doi.org/10.1063/1.1729603>
10. Öberg K.I., Garrod R.T., van Dishoeck E.F., Linnartz H. Formation rates of complex organics in UV irradiated CH₃OH-rich ices. I. Experiments. *Astronomy & Astrophysics*, 2009, vol. 504, no. 3, pp. 891–913. <http://doi.org/10.1051/0004-6361/200912559>
11. Parise B., Castets A., Herbst E., Caux E., Ceccarelli C., Mukhopadhyay I., Tielens A.G.G.M. First detection of triply-deuterated methanol. *Astronomy & Astrophysics*, 2004, vol. 416, no. 1, pp. 159–163. <http://doi.org/10.1051/0004-6361:20034490>
12. Drobyshev A., Aldiyarov A., Sokolov D., Shinbaeva A., Nurmukan A. IR Spectrometry studies of methanol cryovacuum condensates. *Low Temperature Physics*, 2019, vol. 45, no. 4, pp. 441–451. <https://doi.org/10.1063/1.5093525>
13. Aldiyarov A., Aryutkina M., Drobyshev A., Kaikanov M., Kurnosov V. Investigation of dynamic glass transitions and structural transformations in cryovacuum condensates of ethanol. *Low*

Литература

1. Lee Y., Halperin W.P. Recent progress and new challenges in quantum fluids and solids // *Journal of Low Temperature Physics*. 2017. V. 189. N 1. P. 1–14. <https://doi.org/10.1007/s10909-017-1800-4>
2. Jones A.T., Scheller C.P., Prance J.R., Kalyoncu Y.B., Zumbühl D.M., Haley R.P. Progress in cooling nanoelectronic devices to ultra-low temperatures // *Journal of Low Temperature Physics*. 2020. V. 201. N 5. P. 772–802. <https://doi.org/10.1007/s10909-020-02472-9>
3. Debenedetti P.G., Stillinger F.H. Supercooled liquids and the glass transition // *Nature*. 2001. V. 410. N 6825. P. 259–267. <https://doi.org/10.1038/35065704>
4. Cavagna A. Supercooled liquids for pedestrians // *Physics Reports*. 2009. V. 476. N 4–6. P. 51–124. <https://doi.org/10.1016/j.physrep.2009.03.003>
5. Ediger M.D., Angell C.A., Nagel S.R. Supercooled liquids and glasses // *The Journal of Physical Chemistry*. 1996. V. 100. N 31. P. 13200–13212. <https://doi.org/10.1021/jp953538d>
6. Hodge I.M. Enthalpy relaxation and recovery in amorphous materials // *Journal of Non-Crystalline Solids*. 1994. V. 169. N 3. P. 211–266. [https://doi.org/10.1016/0022-3093\(94\)90321-2](https://doi.org/10.1016/0022-3093(94)90321-2)
7. Yokoyama D. Molecular orientation in small-molecule organic light-emitting diodes // *Journal of Materials Chemistry*. 2011. V. 21. N 48. P. 19187–19202. <https://doi.org/10.1039/C1JM13417E>
8. Galliou S., Imbaud J., Goryachev M., Bourquin R., Abbé P. Losses in high quality quartz crystal resonators at cryogenic temperatures // *Applied Physics Letters*. 2011. V. 98. N 9. P. 091911. <https://doi.org/10.1063/1.3559611>
9. Buehler W.J., Gilfrich J.V., Wiley R.C. Effect of low-temperature phase changes on the mechanical properties of alloys near composition TiNi // *Journal of Applied Physics*. 1963. V. 34. N 5. P. 1475–1477. <https://doi.org/10.1063/1.1729603>
10. Öberg K.I., Garrod R.T., van Dishoeck E.F., Linnartz H. Formation rates of complex organics in UV irradiated CH₃OH-rich ices. I. Experiments // *Astronomy & Astrophysics*. 2009. V. 504. N 3. P. 891–913. <http://doi.org/10.1051/0004-6361/200912559>
11. Parise B., Castets A., Herbst E., Caux E., Ceccarelli C., Mukhopadhyay I., Tielens A.G.G.M. First detection of triply-deuterated methanol // *Astronomy & Astrophysics*. 2004. V. 416. N 1. P. 159–163. <http://doi.org/10.1051/0004-6361:20034490>
12. Drobyshev A., Aldiyarov A., Sokolov D., Shinbaeva A., Nurmukan A. IR Spectrometry studies of methanol cryovacuum condensates // *Low Temperature Physics*. 2019. V. 45. N 4. P. 441–451. <https://doi.org/10.1063/1.5093525>
13. Aldiyarov A., Aryutkina M., Drobyshev A., Kaikanov M., Kurnosov V. Investigation of dynamic glass transitions and structural transformations in cryovacuum condensates of ethanol // *Low*

- Temperature Physics*, 2009, vol. 35, no. 4, pp. 251–255. <https://doi.org/10.1063/1.3114588>
14. Drobyshev A., Aldiyarov A., Zhmagaliuly D., Kurnosov V., Tokmoldin N. Thermal desorption and IR spectrometric investigation of polyamorphic and polymorphic transformations in cryovacuum condensates of water. *Low Temperature Physics*, 2007, vol. 33, no. 5, pp. 472–480. <https://doi.org/10.1063/1.2737563>
 15. Aldiyarov A., Aryutkina M., Drobyshev A., Kurnosov V. IR spectroscopy of ethanol in nitrogen cryomatrixes with different concentration ratios. *Low Temperature Physics*, 2011, vol. 37, no. 6, pp. 524–531. <https://doi.org/10.1063/1.3622633>
 16. Drobyshev A., Aldiyarov A., Katpaeva K., Korshikov E., Kurnosov V., Sokolov D. Transformation of cryovacuum condensates of ethanol near the glass transition temperature. *Low Temperature Physics*, 2013, vol. 39, no. 8, pp. 714–718. <https://doi.org/10.1063/1.4818634>
 17. Pontoppidan K.M., Fraser H.J., Dartois E., Thi W.-F., van Dishoeck E.F., Boogert A.C.A., d'Hendecourt L., Tielens A.G.G.M., Bisschop S.E. A 3–5 μm VLT spectroscopic survey of embedded young low mass stars I Structure of the CO ice. *Astronomy & Astrophysics*, 2003, vol. 408, no. 3, pp. 981–1007. <https://doi.org/10.1051/0004-6361:20031030>
 18. McCarthy C., Castillo-Rogez J.C. Planetary ices attenuation properties. *The Science of Solar System Ices*. New York, Springer, 2013, pp. 183–225. https://doi.org/10.1007/978-1-4614-3076-6_7
 19. Moore M.H., Hudson R.L. Far-infrared spectra of cosmic-type pure and mixed ices. *Astronomy and Astrophysics Supplement Series*, 1994, vol. 103, pp. 45–56.
 20. *Interstellar Dust: Proceedings of the 135th Symposium of the International Astronomical Union, Held in Santa Clara, California, July 26–30, 1988*. Ed by L.J. Allamandola, A.G.G.M. Tielens. Springer Science & Business Media, 1989, XVI, 526 p. <https://doi.org/10.1007/978-94-009-2462-8>
 21. Baragiola R.A. Water ice on outer solar system surfaces: Basic properties and radiation effects. *Planetary and Space Science*, 2003, vol. 51, no. 14–15, pp. 953–961. <https://doi.org/10.1016/j.pss.2003.05.007>
 22. Ferraro J.R., Sill G., Fink U. Infrared intensity measurements of cryodeposited thin films of NH_3 , NH_4HS , H_2S , and assignments of absorption bands. *Applied Spectroscopy*, 1980, vol. 34, no. 5, pp. 525–533. <https://doi.org/10.1366/0003702804731339>
 23. Warren S.G. Optical constants of carbon dioxide ice. *Applied Optics*, 1986, vol. 25, no. 16, pp. 2650–2674. <https://doi.org/10.1364/AO.25.002650>
 - Temperature Physics*. 2009. V. 35. N 4. P. 251–255. <https://doi.org/10.1063/1.3114588>
 14. Drobyshev A., Aldiyarov A., Zhmagaliuly D., Kurnosov V., Tokmoldin N. Thermal desorption and IR spectrometric investigation of polyamorphic and polymorphic transformations in cryovacuum condensates of water // *Low Temperature Physics*. 2007. V. 33. N 5. P. 472–480. <https://doi.org/10.1063/1.2737563>
 15. Aldiyarov A., Aryutkina M., Drobyshev A., Kurnosov V. IR spectroscopy of ethanol in nitrogen cryomatrixes with different concentration ratios // *Low Temperature Physics*. 2011. V. 37. N 6. P. 524–531. <https://doi.org/10.1063/1.3622633>
 16. Drobyshev A., Aldiyarov A., Katpaeva K., Korshikov E., Kurnosov V., Sokolov D. Transformation of cryovacuum condensates of ethanol near the glass transition temperature // *Low Temperature Physics*. 2013. V. 39. N 8. P. 714–718. <https://doi.org/10.1063/1.4818634>
 17. Pontoppidan K.M., Fraser H.J., Dartois E., Thi W.-F., van Dishoeck E.F., Boogert A.C.A., d'Hendecourt L., Tielens A.G.G.M., Bisschop S.E. A 3–5 μm VLT spectroscopic survey of embedded young low mass stars I Structure of the CO ice // *Astronomy & Astrophysics*. 2003. V. 408. N 3. P. 981–1007. <https://doi.org/10.1051/0004-6361:20031030>
 18. McCarthy C., Castillo-Rogez J.C. Planetary ices attenuation properties // *The Science of Solar System Ices*. New York: Springer, 2013. P. 183–225. https://doi.org/10.1007/978-1-4614-3076-6_7
 19. Moore M.H., Hudson R.L. Far-infrared spectra of cosmic-type pure and mixed ices // *Astronomy and Astrophysics Supplement Series*. 1994. V. 103. P. 45–56.
 20. *Interstellar Dust: Proceedings of the 135th Symposium of the International Astronomical Union, Held in Santa Clara, California, July 26–30, 1988* / ed by L.J. Allamandola, A.G.G.M. Tielens. Springer Science & Business Media, 1989. XVI, 526 p. <https://doi.org/10.1007/978-94-009-2462-8>
 21. Baragiola R.A. Water ice on outer solar system surfaces: Basic properties and radiation effects // *Planetary and Space Science*. 2003. V. 51. N 14–15. P. 953–961. <https://doi.org/10.1016/j.pss.2003.05.007>
 22. Ferraro J.R., Sill G., Fink U. Infrared intensity measurements of cryodeposited thin films of NH_3 , NH_4HS , H_2S , and assignments of absorption bands // *Applied Spectroscopy*. 1980. V. 34. N 5. P. 525–533. <https://doi.org/10.1366/0003702804731339>
 23. Warren S.G. Optical constants of carbon dioxide ice // *Applied Optics*. 1986. V. 25. N 16. P. 2650–2674. <https://doi.org/10.1364/AO.25.002650>

Authors

Oleg Yu. Golikov — Junior Researcher, Doctoral Student, Al-Farabi Kazakh National University, Almaty, 050040, Republic of Kazakhstan, <https://orcid.org/0000-0002-6691-8346>, golikov@physics.kz

Darkhan Yerezhap — PhD, Senior Researcher, Al-Farabi Kazakh National University, Almaty, 050040, Republic of Kazakhstan, <https://orcid.org/0000-0002-2232-2911>, darhan_13@physics.kz

Dmitriy Yu. Sokolov — PhD, Associate Professor, Al-Farabi Kazakh National University, Almaty, 050040, Republic of Kazakhstan, <https://orcid.org/0000-0001-7966-1140>, yasnyisokol@gmail.com

Авторы

Голиков Олег Юрьевич — младший научный сотрудник, докторант, Казахский национальный университет им. аль-Фараби, Алматы, 050040, Республика Казахстан, <https://orcid.org/0000-0002-6691-8346>, golikov@physics.kz

Ережеп Дархан Есейұлы — PhD, старший научный сотрудник, Казахский национальный университет им. аль-Фараби, Алматы, 050040, Республика Казахстан, <https://orcid.org/0000-0002-2232-2911>, darhan_13@physics.kz

Соколов Дмитрий Юрьевич — PhD, ассоциированный профессор, Казахский национальный университет им. аль-Фараби, Алматы, 050040, Республика Казахстан, <https://orcid.org/0000-0001-7966-1140>, yasnyisokol@gmail.com

Received 21.06.2022

Approved after reviewing 11.11.2022

Accepted 15.01.2023

Статья поступила в редакцию 21.06.2022

Одобрена после рецензирования 11.11.2022

Принята к печати 15.01.2023



Работа доступна по лицензии
Creative Commons
«Attribution-NonCommercial»

doi: 10.17586/2226-1494-2023-23-1-68-78

Investigation on impact and wear behavior of Al6061 (SiC + Al₂O₃) and Al7075 (SiC + Al₂O₃) hybrid composites

Rathinavelu Ravichandaran¹✉, Saminathan Selvarasu², Senthilkumar Gopal³,
Ravisankar Ramachandran⁴

^{1,2} Arunai Engineering College, Tiruvannamalai, 606603, Tamil Nadu, India

³ University College of Engineering Arni, Thatchur, 632326, Tamil Nadu, India

⁴ Kalaignar Karunanidhi Government Arts College, Tiruvannamalai, 606603, Tamil Nadu, India

¹ easternravi@gmail.com✉, <https://orcid.org/0000-0001-7574-634X>

² selvarasu2007@gmail.com, <https://orcid.org/0000-0001-5509-3222>

³ gskannauniversity@gmail.com, <https://orcid.org/0000-0001-7685-211X>

⁴ ravisankarphysics@gmail.com, <https://orcid.org/0000-0002-4798-3820>

Abstract

The current study focuses on the properties of dry sliding characteristics and impact strength of two different aluminum alloys that were reinforced with 100 nm sized Silicon carbide (SiC) and Aluminum oxide (Al₂O₃) ceramic particles, for improving the mechanical properties of the final alloy with the mixing materials characteristics. Stir casting method is adopted for fabricating the composites, matrix being Al6061 and Al7075, utilizing three distinct reinforcement ratios. In order to improve the mechanical properties and increase resistance to wear, tear, and shear, SiC and Al₂O₃ are utilized as reinforcing elements. Following the creation of the composite matrices, their physical and mechanical behaviors are examined in accordance with ASTM standards, and a comparison between the hybrid composites made of Al6061 and Al7075 is then completed. Comparison of the obtained samples showed that the Al7075 (12 % SiC + 6 % Al₂O₃) alloy exhibits characteristics with exceptional tribological and mechanical characteristics. The studied alloy can be used in the automotive industry, for example, in the production of pistons, connecting rods, due to the minimum degree of wear and variable thermal expansion coefficient.

Keywords

hybrid composite, Al6061, Al7075, impact strength, wear resistivity, optical microscope analysis

For citation: Ravichandaran R., Selvarasu S., Gopal S., Ramachandran R. Investigation on impact and wear behavior of Al6061 (SiC + Al₂O₃) and Al7075 (SiC + Al₂O₃) hybrid composites. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 2023, vol. 23, no. 1, pp. 68–78. doi: 10.17586/2226-1494-2023-23-1-68-78

УДК 621.763

Исследование ударных свойств и износостойкости гибридных композитов Al6061 (SiC + Al₂O₃) и Al7075 (SiC + Al₂O₃)

Ратинавелу Равичандаран¹✉, Саминатан Сельварасу², Сентилкумар Гопал³,
Рависанкар Рамачандран⁴

^{1,2} Инженерный колледж Арунаи, Тируваннамалай, 606603, Индия

³ Инженерный университетский колледж Арни, Тэтчур, 632326, Индия

⁴ Государственный колледж искусств Калайньяра Карунанидхи, Тируваннамалай, 606603, Индия

¹ easternravi@gmail.com✉, <https://orcid.org/0000-0001-7574-634X>

² selvarasu2007@gmail.com, <https://orcid.org/0000-0001-5509-3222>

³ gskannauniversity@gmail.com, <https://orcid.org/0000-0001-7685-211X>

⁴ ravisankarphysics@gmail.com, <https://orcid.org/0000-0002-4798-3820>

Аннотация

Исследованы характеристики сухого скольжения и ударной вязкости двух алюминиевых сплавов, армированных керамическими частицами карбида кремния (SiC) и оксида алюминия (Al₂O₃) размером 100 нм. Сплавы

© Ravichandaran R., Selvarasu S., Gopal S., Ramachandran R., 2023

применены для улучшения механических свойств и повышения сопротивления износу, разрыву и сдвигу конечного сплава с характеристиками смешиваемых материалов. Для изготовления композитов использован метод литья с перемешиванием. Матрица композитов представляет собой сплавы Al6061 и Al7075 с использованием трех способов армирования. Физические и механические свойства композитных матриц проверены в соответствии со стандартами Американского общества по материалам и их испытаниям. Приведено сравнение гибридных композитов, изготовленных из Al6061 и Al7075. Сравнение полученных образцов показало, что сплав Al7075 (12 % SiC + 6 % Al₂O₃) обладает исключительными трибологическими и механическими характеристиками. Изученный сплав может найти применение в автомобильной промышленности, например, в производстве поршней, шатунов, благодаря минимальной степени износа и варьируемому коэффициенту теплового расширения.

Ключевые слова

гибридный композит, Al6061, Al7075, ударная вязкость, сопротивление износу, оптический микроскопический анализ

Ссылка для цитирования: Равичандаран Р., Селварасу С., Гопал С., Рамачандран Р. Исследование ударных свойств и износостойкости гибридных композитов Al6061 (SiC + Al₂O₃) и Al7075 (SiC + Al₂O₃) // Научно-технический вестник информационных технологий, механики и оптики. 2023. Т. 23, № 1. С. 68–78 (на англ. яз.). doi: 10.17586/2226-1494-2023-23-1-68-78

Introduction

The alloys of aluminum perform a vital function in numerous programs and in particular in car programs due to their awesome mechanical attributes, resistivity in opposition to corrosion, extremely good ductility, low density and outstanding thermal conductivity. But those alloys do now no longer exhibit honest tribological behavior which limits their programs in numerous engineering disciplines. Increasing call for much less dense and determined fabric brought about the improvement of recent substances known as composites. The engineering of blending or extra substances to attain the behavioral houses of all of the blended substances is called as composite substances. These composites are proved to have ameliorated mechanical and materialistic attributes which incorporates resistivity in opposition to wear, tear and hardness, reduced thermal growth with stepped forward fatigue strength [1–5]. Low-cost production techniques with extemporaneous behavioral attributes also make matrix composites the choice of researchers for applications in many fields, including automotive, aerospace, and others [6].

Combined reinforcement plays an important role in improving behavioral attributes. Aluminum is known to be obtained from bauxite, which is much less dense and ductile than steel, making it easier to cast and machine. Formation of a thin Al₂O₃ (Aluminum Oxide) layer makes it very resistant to corrosion and prevents oxidation. Uniform distribution of the reinforcement throughout the alloy significantly increases the stiffness and strength of the alloy [7]. Carbides, oxides, borides, and nitrides are commonly used as reinforcements [8]. Basically, SiC (Silicon Carbide) is seen as a compelling reinforced material due to its novel properties such as very high melting point, excellent thermal conductivity, and extreme machinability. SiC is also chemically compatible with Al alloys, avoiding intermetallic steps and forming strong bonds [9, 10].

According to the works of Alpas [11], reduced particulate size reinforcements will ensure only the protection of the alloy but whereas the efficient and effective of Al matrix is provided only with the increased

particle size. Thus, for low speed applications, Al matrix with minimized reinforcement is opted [12]. To investigate and determine the dry sliding attributes of the Al matrix, this article involves the study of Al alloy reinforced with the SiC and Al₂O₃ with particle size of 100 nm. To further enhance the experimental investigation, comparison between the attributes of the Al6061 (SiC + Al₂O₃) and Al7075 (SiC + Al₂O₃) hybrid composite was done and concluded; to make the comparison more effective, three different reinforcement ratios were considered in the experimental procedures explained as in later sections.

The three samples of two mixtures taken for comparison in this study, wearing behavior of the alloys were concluded from the POD test simulated using WINDUCOM software. From the results it is clear the Al7075 reinforced with 12 % SiC and 6 % Al₂O₃ has good resistivity in comparison with the other composites. Also, it exhibits improvised hardness and good strength. In additional the alloy study under optical microscope also reveals the same.

Materials and method

Chemical composition of Al6061 & Al7075 alloys

For the comparative analysis of different composite matrix, the alloys considered for the current investigation were Al6061 and Al7075. The reinforcing materials selected for the effective and efficient performance of the alloys were SiC [13, 14] and Al₂O₃ with the average thickness ranging for about 100 nm [15]. The chemical distribution of the selected alloys was tabulated in detail in the Table 1.

Fabrication method

A stir casting method was employed to produce Al composite alloys with three different strengthening rates. This manufacturing technology plays an important role in determining parameters such as: melting temperature, stirring speed, preheating temperature of reinforcing material, etc. [16].

The fabrication process associated with the production of the Al composite alloys is as follows. Initially the SiC reinforcement and the Al₂O₃ were pre-heated for about 40 min at the temperature of 600°C in a muffle furnace. This heating improves the equalization of the volumes

Table 1. Chemical distribution of Al6061 and Al7075 composite alloys

Material	Chemical Composition, %								
	Si	Fe	Cu	Mn	Mg	Cr	Zn	Ti	Al
Al6061	0.35	0.50	1.50	0.59	0.90	1.80	0.59	0.20	Bal
Al7075	0.40	0.43	1.90	0.47	1.50	0.38	1.10	0.20	Bal

Table 2. Detailed Composition of Reinforced particles in composite matrix

Material	Sample no.	wt.% of reinforcement		wt.% of base material
		SiC	Al ₂ O ₃	
Al6061	1	0	0	100
	2	4	2	94
	3	8	4	88
	4	12	6	82
Al7075	1	0	0	100
	2	4	2	94
	3	8	4	88
	4	12	6	82

which gets added in the vertex during stirring and allows repelling out the moisture which causes cluster formations. The reinforcement of the SiC in Al₂O₃ is done to improve the characteristics attribute of the composite alloy in terms of interfacial connect and the instructive phase [17, 18]. The impeller is continuously involved on stirring at the rate of 500 rpm for about 10 min. Then the melt consisting of the reinforcing material is then casted into the mould with a radius of about 15 mm and a length of 150 mm. The mould is cylindrical in shape and ensures constant casting. The attributes of the alloys considered for comparison are tabulated and compared against each other in the Table 2. The required samples are machined as per ASTM (American Society for Testing and Materials) Standards.

Then the casted rods were immediately subjected to cooling at the optimum room temperature. Then the rods were knocked out of the casting machine after the duration of 10 min. The stir casting setup used for fabricating the composite material is shown in Fig. 1, along with the cast

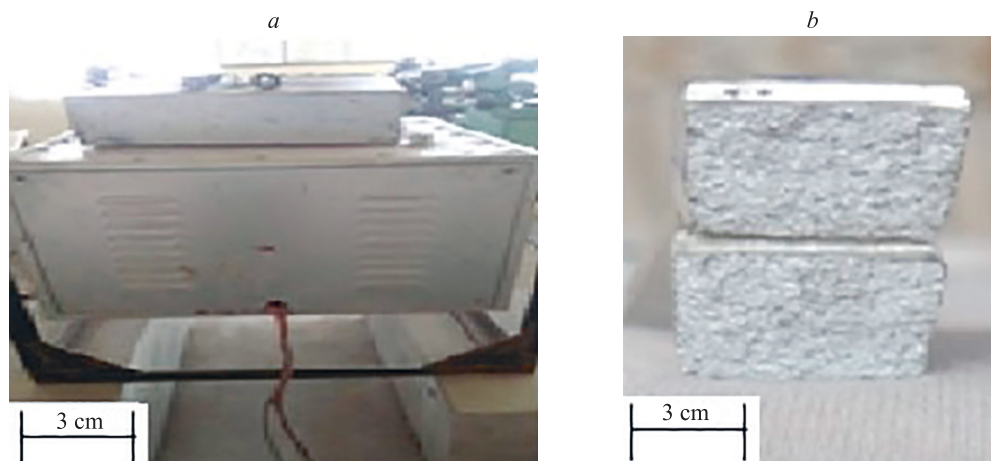
sample. Depending upon the filler ratio, the alloys are differentiated. The ductility level of the reinforced particles was estimated by the brittle testing. The samples fabricated with different filler ratios of each alloys were illustrated in Fig. 1, *b*, as the macrographic picture where the brittle fracture of the samples is significantly visible with normal view to the absence of expanded and fibrous appearance over the surface.

Experimental tests

The investigations for determining the hardness, resistivity against the wear and tear and the impact tests were carried out in agreement to the standards of ASTM.

Impact test

The significant of impact test is analyzed with ASTM E23 standard [19]. The impact behavior is estimated by determining the energy spent to shatter the sample then correlated with the sample cross sectional

Fig. 1. Stir casting setup (*a*) and casted sample (*b*)

dimension, and the standard size adopted for the impact testing in the current investigational procedure which is $10 \times 10 \times 55$ mm.

Hardness test

To conduct the hardness test, the samples were polished to form a smooth surface. The hardness of the samples was determined by the Brinell hardness testing method, as per ASTM E10 standard [20]. This method consists of a ball like structure with the radius of 2.5 mm. A load of 100 kg was subjected on each sample for about 10 secs, and the corresponding readings were recorded. For getting the accurate measuring, the test on each sample was carried out thrice.

Wear test

Pin-on-disc test estimated the wear rate of the composite alloys. The optimal dry laboratory environment was set up for conducting the wear test. All prepared samples were as per the standards of ASTM G99. The experimental set up and the required sample dimension were illustrated in the Fig. 2, *a*. Tribometer estimated the wear and frictional force between the sliding surfaces. The testing sample is placed on the given holder, and then a load is suspended at the sample end [21, 22]. A sliding disc with high carbon high chromium steel was connected across the holder of the Tribometer. The corresponding radius and the thickness of the disc are 100 mm and 8 mm respectively. When the motor starts running, the sliding between the disc and the sample occurs. The motor is stopped after the pre estimated

time of 10 min. The same procedure is repeated for all the samples and the corresponding reading were recorded and tabulated. The required dimension of the pin samples was 6 mm diameter and 25 mm length (Fig. 2, *b*, *c*).

Results and discussion

Impact test results

The results of the instrumented impact test of all HMMC of Al6061 (SiC + Al₂O₃) and Al7075 (SiC + Al₂O₃) alloys at different composition ratio are discussed in this section. The samples before and after impact test were shown in Fig. 3.

From the results it is clearly evident that the impact of both the composite matrices depends upon the dispersion of the particles in the composite matrix [23, 24]. When the stress distribution is uniform, the increase of the impact value of all samples of Al7075 composite alloy is noted. The addition of the reinforcement nanoparticles results in the larger grain boundary with the smaller grain sizes. When the nanoparticles size added to the Al7075 alloy composite increases, the results are obtained as per the expected values with increasing ductility. Thus, the addition of reinforced nanoparticles tends to ameliorate the strength and stiffness and also significantly reduces the thermal expansion.

The comparison of the Al7075 and Al6061 composite alloys is represented in graphical form in the Fig. 4. From

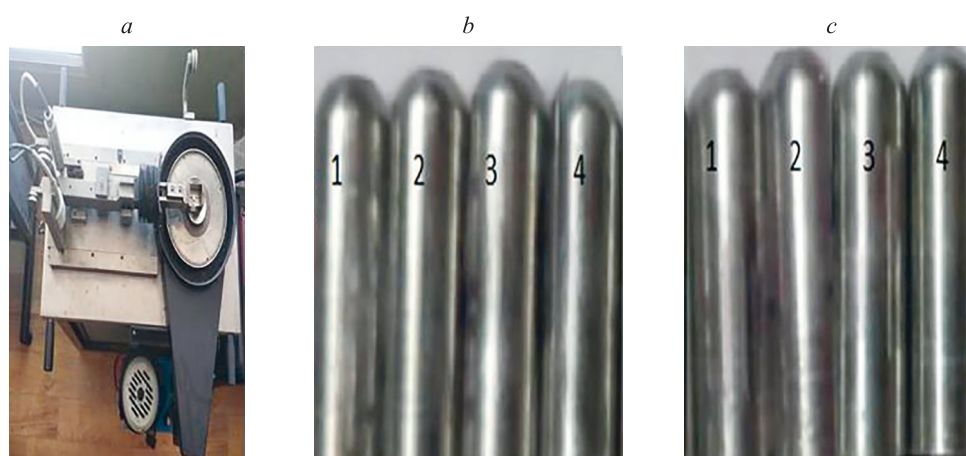


Fig. 2. Wear testing machine (*a*); pin samples: Al6061 (SiC + Al₂O₃) (*b*) and Al7075 (SiC + Al₂O₃) (*c*)

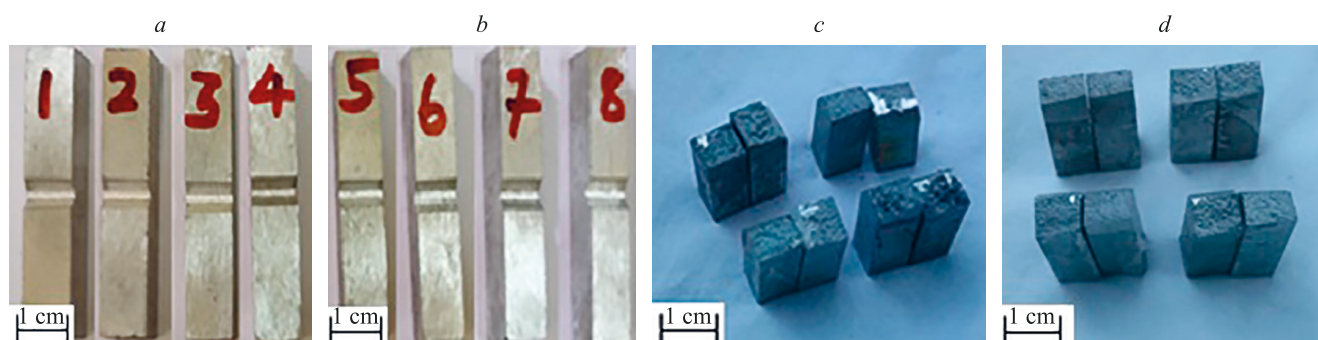


Fig. 3. Samples before test: Al6061 (SiC + Al₂O₃) (*a*); Al7075 (SiC + Al₂O₃) (*b*). Samples after test: Al6061 (SiC + Al₂O₃) (*c*); Al7075 (SiC + Al₂O₃) (*d*)

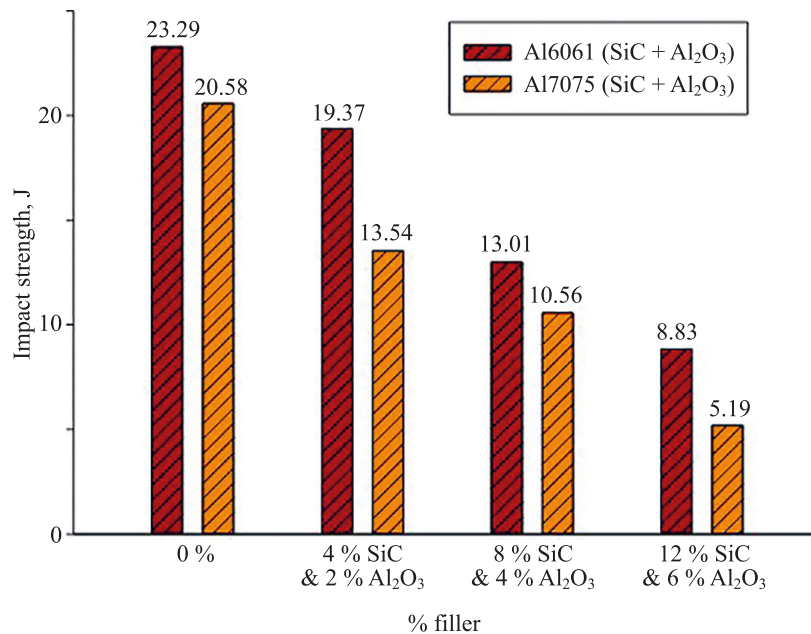


Fig. 4. Graphical representation of the comparative analysis of impact strength

the graph it is observed that the impact strength decreases by increasing the addition of nanoparticles in the alloy matrix. Additionally, for both alloys, the impact strength tends to decline as volume percent. However, it is evident from the results and graph that the alloy Al7075 has lesser impact strength than the other. An additional amazing fact is that the impact strength does not vary with different temperatures [25]. The reducing impact strength in the reinforced alloy composites led to the precipitation of the brittleness in the composite matrix which agrees with the previous studies [26, 27].

Hardness test results

The picture below displays the outcomes of Brinell testing using ASTM E10 standards. A steel ball is

squeezed during this test, after which the ball is taken out. A microscope is used to measure the intended sample diameter.

A graphic illustration of the comparison analysis of the Brinell hardness test data is shown in Fig. 5 where the graphical depiction clearly displays the testing findings for both composite alloys with various filler percentages. The study reveals that the enhanced component ratio is bigger as a result modified mechanical qualities such as toughness and hardness. Upon analyzing the graph, it is evident that the hardness of the Al7075 (12 % SiC+6 % Al₂O₃) is higher to the value of 123 BHN. This clearly signifies that the Al7075 (12 % SiC+6 % Al₂O₃) has better performance characteristics in terms of hardness when compared with

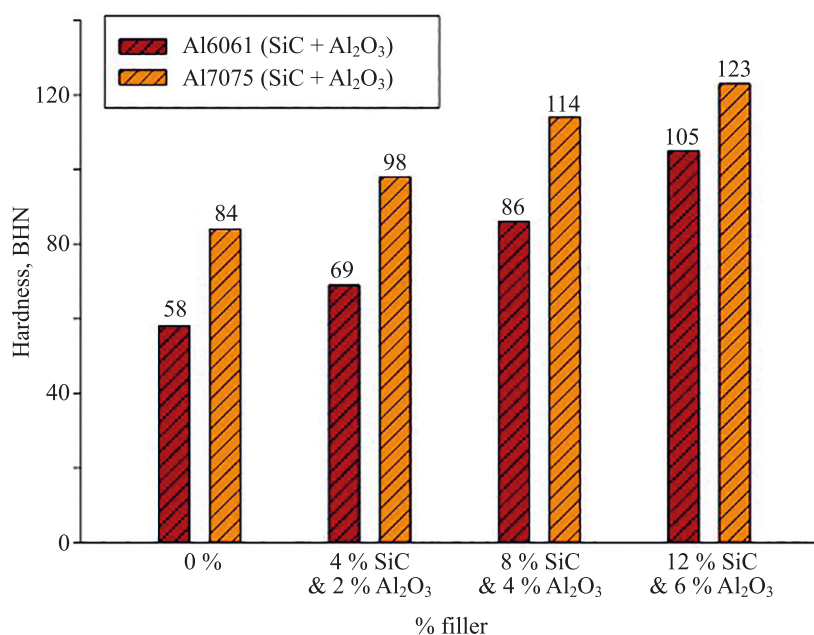


Fig. 5. Comparative analysis of Brinell hardness test

the other alloys. The reinforced material SiC has higher hardness when compared to that of the Al_2O_3 , thus by increasing the composite alloy hardness.

Wear test results

On all three samples of each alloy, the rotary wear test was conducted for approximately 10 min with a load of 30 N spinning at a speed of 500 rpm to evaluate the alloys resistance to wearing. Sample wear rate is identified by weight loss method. Weight is divided for known distance to identify the sliding distance. Different sliding speeds with different normal load were used for wear test. To find out the wear loss, the weights are measured before and after testing the sample. While sliding, friction force of the sample is used to obtain the coefficient of friction of sample and it was plotted against sliding distance.

The ratio of different compositions is kept constant throughout the test procedure. The experimental setup is connected to a DAC that calculates friction forces, and the corresponding coefficients of WINDUCOM software are used to generate graphic data for the results compared the

coefficient of friction force load and rotational speed for both alloy composites tabulated in Table 3.

Wear loss

From the investigation, it was found that with time increasing, the losses due to wear increase indicating an increase in wear loss. With the increasing wear loss, deformation of the Al matrix occurs which could end up in the surface cracking of the alloys. These effects result in the removal of higher material from the substrate as wear loss. The reinforced NPs of SiC and Al_2O_3 also wear out of the surface causing the formation of grooves on the matrix alloy surfaces. From the research it is concluded that the alloy resistivity increased due to SiC nanoparticle reinforcement, as shown in Fig. 6.

Coefficient of friction

The frictional coefficient investigated for the SiC and Al_2O_3 in Al6061 and Al7075 is shown in the Fig. 7. The coefficient of friction tends to decline with the sliding rate and load. The frictional coefficient value of the Al composite matrix is always higher when compared with the

Table 3. Wear test results of Al7075 and Al6061

Material	Reinforcement volume, %		Matrix volume, %	Sliding speed, m/s	Time, min	Wear loss, g	Friction force, N	Coefficient of friction
	SiC	Al_2O_3						
Al6061	0	0	100	1.32	10	0.038	15.53	0.398
	4	2	94	1.32	10	0.027	26.22	0.415
	8	4	88	1.83	10	0.020	32.45	0.456
	12	6	82	2.62	10	0.015	41.37	0.487
Al7075	0	0	100	2.83	10	0.034	13.43	0.343
	4	2	94	2.83	10	0.022	22.51	0.356
	8	4	88	1.62	10	0.016	28.15	0.387
	12	6	82	1.62	10	0.010	34.56	0.405

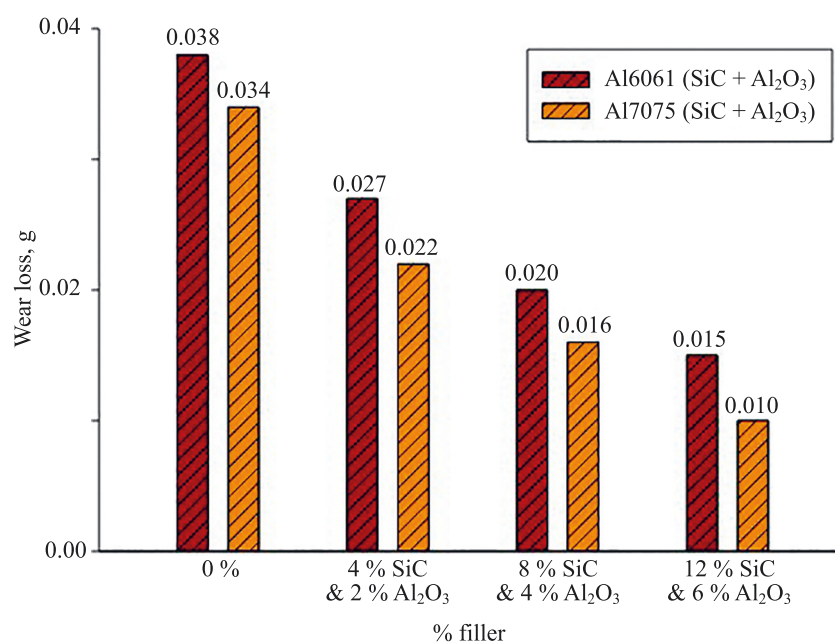


Fig. 6. Wear loss for Al6061 (SiC + Al_2O_3) and Al7075 (SiC + Al_2O_3)

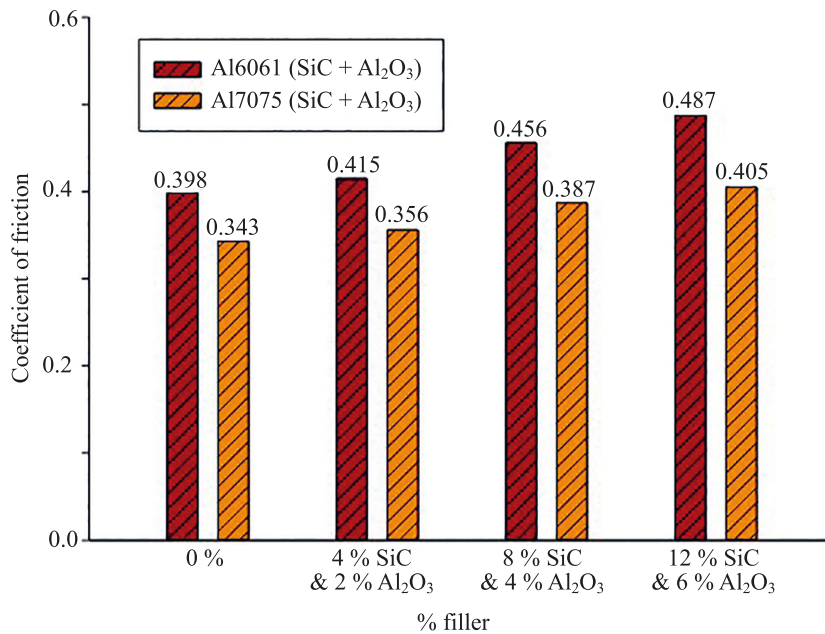


Fig. 7. Coefficient of friction for Al6061 (SiC + Al₂O₃) and Al7075 (SiC + Al₂O₃)

original Al alloy. With the value of the frictional coefficient increase, the wear loss also decreases and, hence, there is less wear loss in the Al composite matrices. This agrees with the conclusions made by other researches. From the graphs it is concluded that the increasing in the frictional coefficient occurs when the reinforcement particles SiC and Al₂O₃ are uniformly dispersed in the Al composite matrix.

Frictional force

From the graphs it is clear that the value of the frictional force depends upon the ratio of the reinforcement materials in the composite matrix. Similar to the results of the frictional coefficient, frictional force also tends to increase with the increasing ratio of the SiC and Al₂O₃, as shown in Fig. 8.

Microstructural analysis of wear test samples

The metallographic process which includes the selection of sample, mounting, grinding, polishing and etching was done in order to subject the samples to the micro structural analysis. Then the constructional architecture of the samples was then microscopically characterized after etching with the NaOH solution. The microscopic studies of the Al7075 and Al6061 with scope ratio of 100× and 500× were illustrated in the Fig. 9.

The microphotographs of the Al7075 (4 % SiC + 2 % Al₂O₃) were shown in the Fig. 9, *a, b* where the distribution of the nanoparticles along the grain margin of the primary solid solution of Al can be seen. The microscopic analysis of the Al7075 (8 % SiC + 4 % Al₂O₃) is illustrated in the

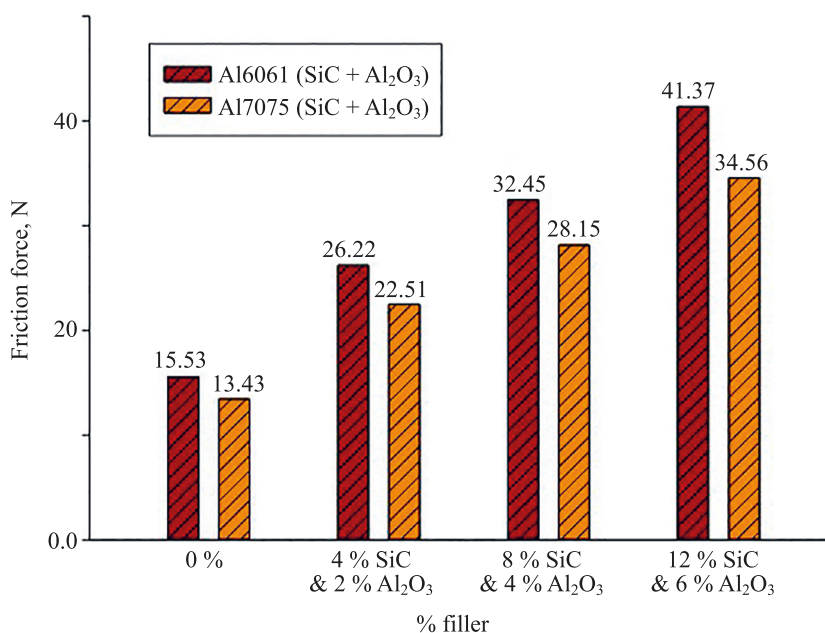


Fig. 8. Frictional force for Al6061 (SiC + Al₂O₃) and Al7075 (SiC + Al₂O₃)

Fig. 9, *c, d*. On these figures the significant feature is the even dispersion of the nanoparticles along the margins. In the Fig. 9, *e, f* the microscopic analysis of Al7075 (12 % SiC + 6 % Al₂O₃) is given. This picture also clearly depicts the even distribution of the SiC and Al₂O₃ throughout the alloy matrices.

Fig. 9, *g, h* illustrates the microphotographs of the alloy Al6061 (4 % SiC + 2 % Al₂O₃). On the pictures the

significant distribution of the SiC and Al₂O₃ is clearly observed. The particles are viewed as the clusters. Fig. 9, *g* presents the isolated distribution of the reinforced nanoparticles along the grain margins. The microscopic illustration of the alloy Al6061 (8 % SiC + 4 % Al₂O₃) was given in the Fig. 9, *i, j* for different magnification. Upon keen observation the fair distribution of the reinforced nanoparticles was clearly evident. The microphotographs

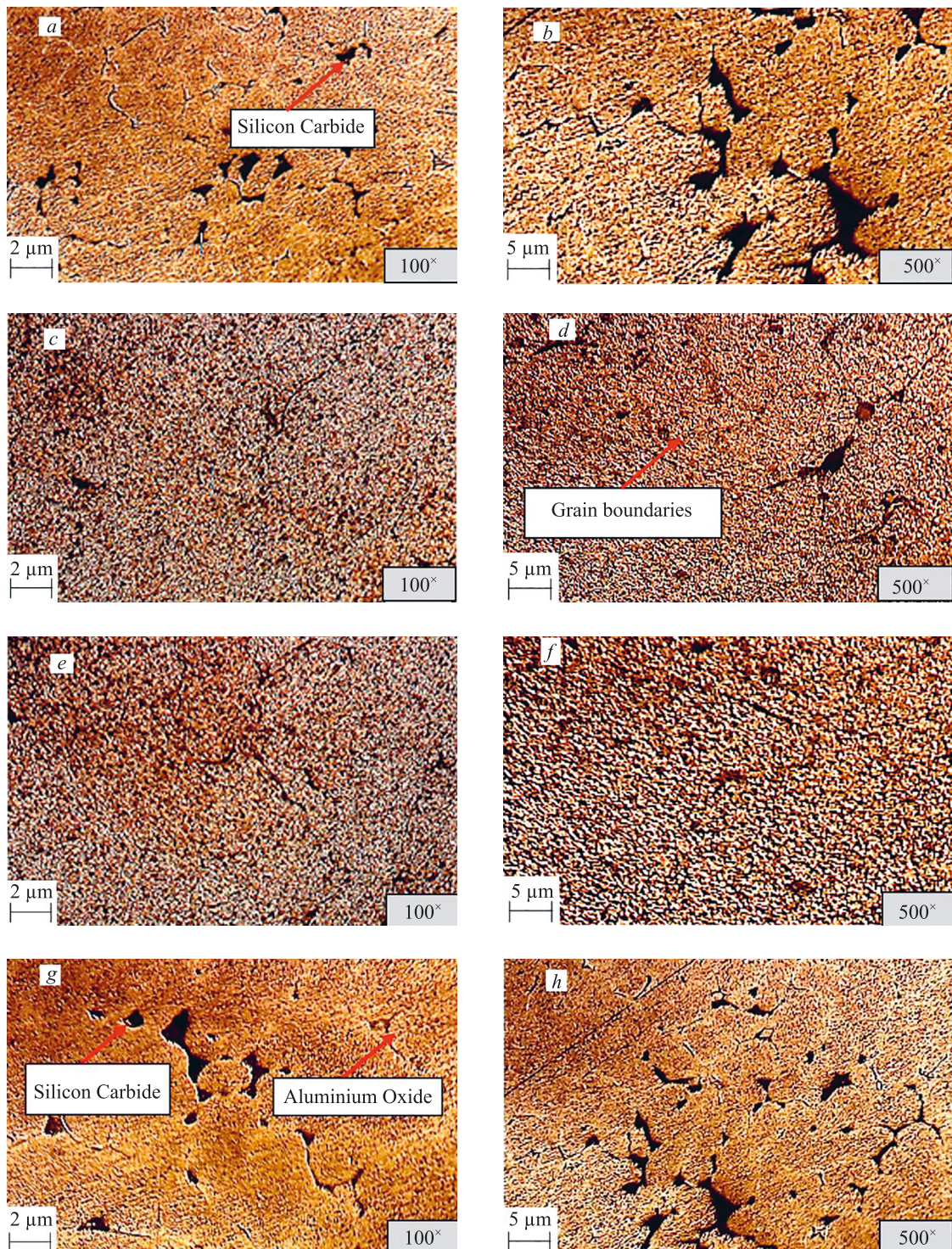


Fig. 9. Microstructure illustration composite alloys with scope ratio of 100 \times and 500 \times : Al7075 (4 % SiC + 2 % Al₂O₃) – 100 \times (a) and 500 \times (b); Al7075 (8 % SiC + 4 % Al₂O₃) – 100 \times (c) and 500 \times (d); Al7075 (12 % SiC + 6 % Al₂O₃) – 100 \times (e) and 500 \times (f); Al6061 (4 % SiC + 2 % Al₂O₃) – 100 \times (g) and 500 \times (h); Al6061 (8 % SiC + 4 % Al₂O₃) – 100 \times (i) and 500 \times (j); Al6061 (12 % SiC + 6 % Al₂O₃) – 100 \times (k) and 500 \times (l)

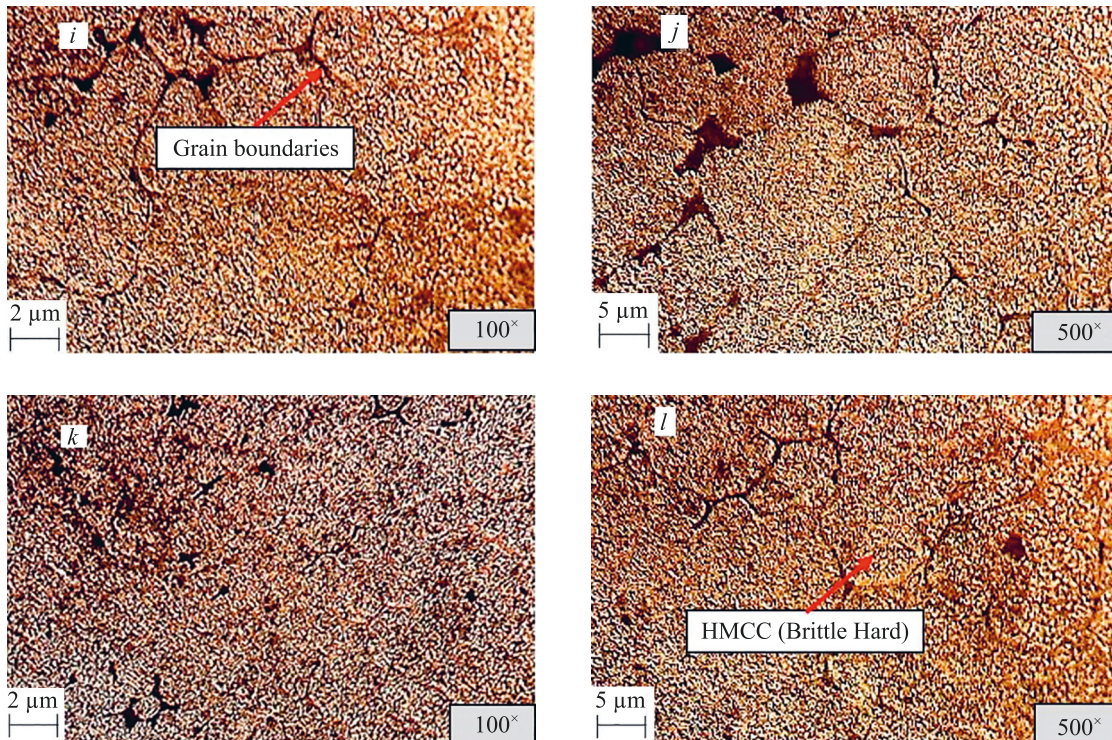


Fig. 9. Continuation

of the alloy Al6061 (12 % SiC + 6 % Al₂O₃) were given in the Fig. 9, *k*, *l*. In these figures the increased concentration of the nanoparticles in the composite matrices can be observed.

Upon investigation of all the samples, it is found that Al7075 (12 % SiC + 6 % Al₂O₃) composite has the minimal wearing rate when compared with the other samples. From the study it was proven that the distribution of the reinforcement materials SiC and Al₂O₃ was uniform throughout the matrix. Thus, it is concluded that with the increase in the filler composition the mechanical attributes and the hardness and wear rate of the matrix is also improved. Also, the thermal expansion was significantly reduced. The fusion of the reinforced particles to the Al alloy progresses the hardness, reduces the impact strength and ameliorates the wear resistance against sliding as concluded in the previous studies.

Conclusion

The considerable conclusions of the current investigation involving the comparison of the Al6061 and Al7075 matrix composite alloys were presented as follows.

The fabrication process (involving the liquid metallurgy methods) is advantageous to produce the composite of Al6061 and Al7075 alloys reinforced with SiC and Al₂O₃ with the filler ratio of about 18 %.

The particulates were dispersed evenly throughout the composite matrix which was confirmed through the microstructural analysis.

Increase in the hardness of the sample with the increasing filler ratio was observed and the corresponding hardness estimated for the Al6061 and Al7075 in the current investigation were 105 and 123 BHN respectively.

The impact test results reveal that the composite alloys have reduced impact strength when compared with that of the base alloys. Among the two alloys under consideration, Al7075 has the highest impact strength.

Generally, the samples of both alloys show improvised resistivity against wear, but Al7075 (12 % SiC + 6 % Al₂O₃), specifically, exhibits notable increase in the resistivity against wear and friction.

From the microscopy analysis of all the samples, Al7075 (12 % SiC + 6 % Al₂O₃) holds minimal wear rate with fair distribution of the reinforced nanoparticles along the grain boundary and throughout the alloy matrix. It is clearly evident that the filler composition of this sample improvises the hardness and rate of wear of the matrix with significant reduction in the thermal expansion.

Thus, if you draw a conclusion after comparing all the samples, Al7075 (12 % SiC + 6 % Al₂O₃) shows an outstanding behavioral attributes with the extraordinary tribological and mechanical specifications.

From the above conclusions, Al7075 (12 % SiC + 6 % Al₂O₃) is considered to be a promising alloy in the applications of automobile industry, like piston manufacturing, connecting rod due to its minimized wear rate and improvised thermal expansion.

References

1. Chawla K.K. *Composite Materials. Science and Engineering*. 2nd ed. New York: Springer Verlag, 1998, 165 p.
2. Chawla N., Shen Y.L. Mechanical behavior of particle reinforced metal matrix composites. *Advanced Engineering Materials*, 2001, vol. 3, no. 6, pp. 357–370.
3. Senthilkumar N., Kalaichelvan K., Elangovan K. Mechanical behaviour of aluminum particulate epoxy composite — experimental study and numerical simulation. *International Journal of Mechanical and Materials Engineering*, 2012, vol. 7, no. 3, pp. 214–221.
4. Srivatsan T.S., Al-Hajri M., Smith C., Petraroli M. The tensile response and fracture behavior of 2009 aluminium alloy metal matrix composite. *Materials Science and Engineering: A*, 2003, vol. 346, no. 1–2, pp. 91–100. [https://doi.org/10.1016/S0921-5093\(02\)00481-1](https://doi.org/10.1016/S0921-5093(02)00481-1)
5. Tan M., Xin Q., Li Z., Zong B.Y. Influence of SiC and Al₂O₃ particulate reinforcements and heat treatments on mechanical properties and damage evolution of Al-2618 metal matrix composites. *Journal of Materials Science*, 2001, vol. 36, no. 8, pp. 2045–2053. <https://doi.org/10.1023/A:1017591117670>
6. Lloyd D.J. Particle reinforced aluminium and magnesium matrix composites. *International Materials Reviews*, 1994, vol. 39, no. 1, pp. 1–23. <https://doi.org/10.1179/imr.1994.39.1.1>
7. Sundaraselvan S., Senthilkumar N. Surface modification of AZ61 magnesium alloy with nano-Al₂O₃ using laser cladding technique: optimization of wear properties through hybrid GRA-PCA. *International Journal of Rapid Manufacturing*, 2019, vol. 8, no. 3, pp. 221. <https://doi.org/10.1504/IJRAPIDM.2019.10020259>
8. Tjong S.C., Wu S.Q., Zhu H.G. Wear behavior of in situ TiB₂-Al₂O₃/Al and TiB₂-Al₂O₃/Al-Cu composites. *Composites Science and Technology*, 1999, vol. 59, no. 9, pp. 1341–1347. [https://doi.org/10.1016/S0266-3538\(98\)00172-9](https://doi.org/10.1016/S0266-3538(98)00172-9)
9. Nair S.V., Tien J.K., Bates R.C. SiC-reinforced aluminium metal matrix composites. *International Metals Reviews*, 1985, vol. 30, no. 1, pp. 275–290. <https://doi.org/10.1179/imtr.1985.30.1.275>
10. Bhojan N., Senthilkumar B., Deepanraj B. Parametric influence of friction stir welding on cast Al6061/20%SiC/2%MoS₂ MMC mechanical properties. *Applied Mechanics and Materials*, 2016, vol. 852, pp. 297–303. <https://doi.org/10.4028/www.scientific.net/AMM.852.297>
11. Alpas A.T., Zhang J. Effect of microstructure (particulate size and volume fraction) and counter face material on the sliding wear resistance of particulate-reinforced. *Metallurgical and Materials Transactions A*, 1994, vol. 25, no. 5, pp. 969–983. <https://doi.org/10.1007/BF02652272>
12. Kwok J.K.M., Lim S.C. High-speed tribological properties of some Al/SiCp composites. I. Frictional and wear-rate characteristics. *Composites Science and Technology*, 1999, vol. 59, no. 1, pp. 55–63. [https://doi.org/10.1016/S0266-3538\(98\)00055-4](https://doi.org/10.1016/S0266-3538(98)00055-4)
13. Friend C.M. Toughness in metal matrix composites. *Materials Science and Technology*, 1989, vol. 5, no. 1, pp. 1–7. <https://doi.org/10.1179/mst.1989.5.1.1>
14. Nardone V.C., Strife J.R., Prew K.M. Microstructurally toughened particulate-reinforced aluminum matrix composites. *Metallurgical Transactions A*, 1991, vol. 22, no. 1, pp. 171–182. <https://doi.org/10.1007/BF03350959>
15. Ozden S., Ekici R., Nair F. Investigation of impact behaviour of aluminium based SiC particle reinforced metal–matrix composites. *Composites Part A: Applied Science and Manufacturing*, 2007, vol. 38, no. 2, pp. 484–494. <https://doi.org/10.1016/j.compositesa.2006.02.026>
16. Ragupathy K., Velmurugan C., Senthilkumar N. Tribological and heat treatment prediction of stir cast Al 6061/SiC/MoS₂ composites using grey relational analysis. *Journal of the Balkan Tribological Association*, 2018, vol. 24, no. 2, pp. 198–217.
17. Saminathan S., Laksmipathy J. Experimental investigation and prediction analysis on Granite/SiC Reinforced Al7050 and Al7075 using hybrid deep neural network based salp swarm optimization. *Silicon*, 2022, vol. 14, no. 11, pp. 5887–5903. <https://doi.org/10.1007/s12633-021-01349-0>
18. Nielson L.E., Landel R.F. *Mechanical Properties of Polymers and Composites*. New York: Marcel Dekker, Inc., 1994.
19. ASTM E23-93a. *Standard test methods for notched bar impact testing of metallic materials*. 1993 Annual book of ASTM Standards, pp. 206–26.

Литература

1. Chawla K.K. *Composite Materials. Science and Engineering* / 2nd ed. New York: Springer Verlag, 1998. 165 p.
2. Chawla N., Shen Y.L. Mechanical behavior of particle reinforced metal matrix composites // *Advanced Engineering Materials*. 2001. V. 3. N 6. P. 357–370.
3. Senthilkumar N., Kalaichelvan K., Elangovan K. Mechanical behaviour of aluminum particulate epoxy composite – experimental study and numerical simulation // *International Journal of Mechanical and Materials Engineering*. 2012. V. 7. N 3. P. 214–221.
4. Srivatsan T.S., Al-Hajri M., Smith C., Petraroli M. The tensile response and fracture behavior of 2009 aluminium alloy metal matrix composite // *Materials Science and Engineering: A*. 2003. V. 346. N 1–2. P. 91–100. [https://doi.org/10.1016/S0921-5093\(02\)00481-1](https://doi.org/10.1016/S0921-5093(02)00481-1)
5. Tan M., Xin Q., Li Z., Zong B.Y. Influence of SiC and Al₂O₃ particulate reinforcements and heat treatments on mechanical properties and damage evolution of Al-2618 metal matrix composites // *Journal of Materials Science*. 2001. V. 36. N 8. P. 2045–2053. <https://doi.org/10.1023/A:1017591117670>
6. Lloyd D.J. Particle reinforced aluminium and magnesium matrix composites // *International Materials Reviews*. 1994. V. 39. N 1. P. 1–23. <https://doi.org/10.1179/imr.1994.39.1.1>
7. Sundaraselvan S., Senthilkumar N. Surface modification of AZ61 magnesium alloy with nano-Al₂O₃ using laser cladding technique: optimization of wear properties through hybrid GRA-PCA // *International Journal of Rapid Manufacturing*. 2019. V. 8. N 3. P. 221. <https://doi.org/10.1504/IJRAPIDM.2019.10020259>
8. Tjong S.C., Wu S.Q., Zhu H.G. Wear behavior of in situ TiB₂-Al₂O₃/Al and TiB₂-Al₂O₃/Al-Cu composites // *Composites Science and Technology*. 1999. V. 59. N 9. P. 1341–1347. [https://doi.org/10.1016/S0266-3538\(98\)00172-9](https://doi.org/10.1016/S0266-3538(98)00172-9)
9. Nair S.V., Tien J.K., Bates R.C. SiC-reinforced aluminium metal matrix composites // *International Materials Reviews*. 1985. V. 30. N 1. P. 275–290. <https://doi.org/10.1179/imtr.1985.30.1.275>
10. Bhojan N., Senthilkumar B., Deepanraj B. Parametric influence of friction stir welding on cast Al6061/20%SiC/2%MoS₂ MMC mechanical properties // *Applied Mechanics and Materials*. 2016. V. 852. P. 297–303. <https://doi.org/10.4028/www.scientific.net/AMM.852.297>
11. Alpas A.T., Zhang J. Effect of microstructure (particulate size and volume fraction) and counter face material on the sliding wear resistance of particulate-reinforced // *Metallurgical and Materials Transactions A*. 1994. V. 25. N 5. P. 969–983. <https://doi.org/10.1007/BF02652272>
12. Kwok J.K.M., Lim S.C. High-speed tribological properties of some Al/SiCp composites. I. Frictional and wear-rate characteristics // *Composites Science and Technology*. 1999. V. 59. N 1. P. 55–63. [https://doi.org/10.1016/S0266-3538\(98\)00055-4](https://doi.org/10.1016/S0266-3538(98)00055-4)
13. Friend C.M. Toughness in metal matrix composites // *Materials Science and Technology*. 1989. V. 5. N 1. P. 1–7. <https://doi.org/10.1179/mst.1989.5.1.1>
14. Nardone V.C., Strife J.R., Prew K.M. Microstructurally toughened particulate-reinforced aluminum matrix composites // *Metallurgical Transactions A*. 1991. V. 22. N 1. P. 171–182. <https://doi.org/10.1007/BF03350959>
15. Ozden S., Ekici R., Nair F. Investigation of impact behaviour of aluminium based SiC particle reinforced metal–matrix composites // *Composites Part A: Applied Science and Manufacturing*. 2007. V. 38. N 2. P. 484–494. <https://doi.org/10.1016/j.compositesa.2006.02.026>
16. Ragupathy K., Velmurugan C., Senthilkumar N. Tribological and heat treatment prediction of stir cast Al 6061/SiC/MoS₂ composites using grey relational analysis // *Journal of the Balkan Tribological Association*. 2018. V. 24. N 2. P. 198–217.
17. Saminathan S., Laksmipathy J. Experimental investigation and prediction analysis on Granite/SiC Reinforced Al7050 and Al7075 using hybrid deep neural network based salp swarm optimization // *Silicon*. 2022. V. 14. N 11. P. 5887–5903. <https://doi.org/10.1007/s12633-021-01349-0>
18. Nielson L.E., Landel R.F. *Mechanical Properties of Polymers and Composites*. New York: Marcel Dekker, Inc., 1994.
19. ASTM E23-93a. *Standard test methods for notched bar impact testing of metallic materials*: 1993 Annual book of ASTM Standards. P. 206–26.
20. Selvakumar V., Muruganandam S., Senthilkumar N. Evaluation of mechanical and tribological behavior of Al–4 %Cu–x %SiC

20. Selvakumar V., Muruganandam S., Senthilkumar N. Evaluation of mechanical and tribological behavior of Al-4 %Cu-x %SiC composites prepared through powder metallurgy technique. *Transactions of the Indian Institute of Metals*, 2017, vol. 70, no. 5, pp. 1305–1315. <https://doi.org/10.1007/s12666-016-0923-7>
21. Thirumalvalavan S., Senthilkumar N. Experimental investigation and optimization of HVOF spray parameters on wear resistance behaviour of Ti-6Al-4V alloy // *Comptes rendus de l'Academie bulgare des Sciences*, 2019, vol. 72, no. 5, pp. 664–673. <https://doi.org/10.7546/crabs.2019.05.15>
22. Gajalakshmi K., Senthilkumar N., Prabu B. Multi-response optimization of dry sliding wear parameters of AA6026 using hybrid gray relational analysis coupled with response surface method. *Measurement and Control*, 2019, vol. 52, no. 5-6, pp. 540–553. <https://doi.org/10.1177/0020294019842603>
23. Bonollo F., Ceschini L., Garagnani G.L. Mechanical and Impact behaviour of (Al₂O₃)p/2014 and (Al₂O₃)p/6061 Al metal matrix composites in the 25–200°C range. *Applied Composite Materials*, 1997, vol. 4, no. 3, pp. 173–185. <https://doi.org/10.1007/BF02481779>
24. Surappa M.K., Sivakumar P. Fracture toughness evaluation of 2040-Al/Al₂O₃ particulate composites by instrumented impact. *Composites Science and Technology*, 1993, vol. 46, no. 3, pp. 287–292. [https://doi.org/10.1016/0266-3538\(93\)90162-A](https://doi.org/10.1016/0266-3538(93)90162-A)
25. Poza P., Llorca J. Fracture toughness and fracture mechanisms of Al–Al₂O₃ composites at cryogenic and elevated temperatures. *Materials Science and Engineering: A*, 1996, vol. 206, no. 2, pp. 183–193. [https://doi.org/10.1016/0921-5093\(95\)09999-9](https://doi.org/10.1016/0921-5093(95)09999-9)
26. Hasson D.F., Hoover S.M., Crowe C.R. Effect of thermal treatment on the mechanical and toughness properties of extruded SiCw/aluminium 6061 metal matrix composite. *Journal of Materials Science*, 1985, vol. 20, no. 11, pp. 4147–4154. <https://doi.org/10.1007/BF00552410>
27. Unsworth J.P., Bandyopadhyay S. Effect of thermal ageing on hardness, tensile and Impact properties of an alumina microsphere-reinforced aluminium metal–matrix composite. *Journal of Materials Science*, 1994, vol. 29, no. 17, pp. 4645–4650. <https://doi.org/10.1007/BF00376291>
- composites prepared through powder metallurgy technique // *Transactions of the Indian Institute of Metals*, 2017, V. 70, N 5. P. 1305–1315. <https://doi.org/10.1007/s12666-016-0923-7>
21. Thirumalvalavan S., Senthilkumar N. Experimental investigation and optimization of HVOF spray parameters on wear resistance behaviour of Ti-6Al-4V alloy // *Comptes rendus de l'Academie bulgare des Sciences*, 2019, V. 72, N 5. P. 664–673. <https://doi.org/10.7546/crabs.2019.05.15>
22. Gajalakshmi K., Senthilkumar N., Prabu B. Multi-response optimization of dry sliding wear parameters of AA6026 using hybrid gray relational analysis coupled with response surface method // *Measurement and Control*, 2019, V. 52, N 5-6. P. 540–553. <https://doi.org/10.1177/0020294019842603>
23. Bonollo F., Ceschini L., Garagnani G.L. Mechanical and Impact behaviour of (Al₂O₃)p/2014 and (Al₂O₃)p/6061 Al metal matrix composites in the 25–200°C range // *Applied Composite Materials*, 1997, V. 4, N 3. P. 173–185. <https://doi.org/10.1007/BF02481779>
24. Surappa M.K., Sivakumar P. Fracture toughness evaluation of 2040-Al/Al₂O₃ particulate composites by instrumented impact // *Composites Science and Technology*, 1993, V. 46, N 3. P. 287–292. [https://doi.org/10.1016/0266-3538\(93\)90162-A](https://doi.org/10.1016/0266-3538(93)90162-A)
25. Poza P., Llorca J. Fracture toughness and fracture mechanisms of Al–Al₂O₃ composites at cryogenic and elevated temperatures // *Materials Science and Engineering: A*, 1996, V. 206, N 2. P. 183–193. [https://doi.org/10.1016/0921-5093\(95\)09999-9](https://doi.org/10.1016/0921-5093(95)09999-9)
26. Hasson D.F., Hoover S.M., Crowe C.R. Effect of thermal treatment on the mechanical and toughness properties of extruded SiCw/aluminium 6061 metal matrix composite // *Journal of Materials Science*, 1985, V. 20, N 11. P. 4147–4154. <https://doi.org/10.1007/BF00552410>
27. Unsworth J.P., Bandyopadhyay S. Effect of thermal ageing on hardness, tensile and Impact properties of an alumina microsphere-reinforced aluminium metal–matrix composite // *Journal of Materials Science*, 1994, V. 29, N 17. P. 4645–4650. <https://doi.org/10.1007/BF00376291>

Authors

Rathinavelu Ravichandaran — PhD, Professor, Arunai Engineering College, Tiruvannamalai, 606603, Tamil Nadu, India, [sc 57486922200](https://orcid.org/0000-0001-7574-634X), <https://orcid.org/0000-0001-7574-634X>, easternravi@gmail.com
Saminathan Selvarasu — PhD, Assistant Professor, Arunai Engineering College, Tiruvannamalai, 606603, Tamil Nadu, India, [sc 57254139100](https://orcid.org/0000-0001-5509-3222), <https://orcid.org/0000-0001-5509-3222>, selvarasu2007@gmail.com
Senthilkumar Gopal — PhD, Professor, University College of Engineering Arni, Thatchur, 632326, Tamil Nadu, India, [sc 40462266500](https://orcid.org/0000-0001-7685-211X), <https://orcid.org/0000-0001-7685-211X>, gskannauniversity@gmail.com
Ravisankar Ramachandran — PhD, Assistant Professor, Kalaingar Karunanidhi Government Arts College, Tiruvannamalai, 606603, Tamil Nadu, India, [sc 14825808700](https://orcid.org/0000-0002-4798-3820), <https://orcid.org/0000-0002-4798-3820>, ravisankarphysics@gmail.com

Авторы

Равичандаран Ратинавелу — PhD, профессор, Инженерный колледж Арунаи, Тируваннамалай, 606603, Индия, [sc 57486922200](https://orcid.org/0000-0001-7574-634X), <https://orcid.org/0000-0001-7574-634X>, easternravi@gmail.com
Сельварасу Саминатан — PhD, доцент, Инженерный колледж Арунаи, Тируваннамалай, 606603, Индия, [sc 57254139100](https://orcid.org/0000-0001-5509-3222), <https://orcid.org/0000-0001-5509-3222>, selvarasu2007@gmail.com
Гопал Сентилкумар — PhD, профессор, Инженерный университетский колледж Арни, Тэтчур, 632326, Индия, [sc 40462266500](https://orcid.org/0000-0001-7685-211X), <https://orcid.org/0000-0001-7685-211X>, gskannauniversity@gmail.com
Рамачандран Рависанкар — PhD, доцент, Государственный колледж искусств Калайньяра Карунанидхи, Тируваннамалай, 606603, Индия, [sc 14825808700](https://orcid.org/0000-0002-4798-3820), <https://orcid.org/0000-0002-4798-3820>, ravisankarphysics@gmail.com

Received 06.10.2022

Approved after reviewing 22.11.2022

Accepted 15.01.2023

Статья поступила в редакцию 06.10.2022

Одобрена после рецензирования 22.11.2022

Принята к печати 15.01.2023



Работа доступна по лицензии
Creative Commons
«Attribution-NonCommercial»

КОМПЬЮТЕРНЫЕ СИСТЕМЫ И ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ COMPUTER SCIENCE

doi: 10.17586/2226-1494-2023-23-1-79-87

УДК 004.932

Вычислительные методы повышения быстродействия дискретного вейвлет-преобразования на базе FPGA

Сергей Владимирович Сай¹✉, Алексей Владимирович Зинкевич²

^{1,2} Тихоокеанский государственный университет, Хабаровск, 680035, Российская Федерация

¹ sai1111@rambler.ru✉, <https://orcid.org/0000-0002-4506-1857>

² zinkevich1985@mail.ru, <https://orcid.org/0000-0003-3231-8569>

Аннотация

Предмет исследования. Рассмотрены вычислительные методы и особенности построения сложно-функционального блока реализации дискретного вейвлет-преобразования Добеши 9/7 в системах цифровой обработки сигналов изображений на базе Field-Programmable Gate Array (FPGA). **Метод.** Предложена математическая модель и алгоритмы реализации методов параллельной и последовательно-конвейерной обработок сигналов для вычисления коэффициентов дискретного биортогонального вейвлета Добеши 9/7 с учетом архитектуры используемой FPGA. В основе модели использованы методы факторизации вейвлет-преобразований с применением лифтинговых схем. В отличие от обычных лифтинговых схем, предлагаемый способ и алгоритмы позволяют повысить быстродействие вычислений дискретного вейвлет-преобразования при упрощении аппаратной реализации. В качестве среды разработки использованы средства автоматизированного проектирования Quartus II и ModelSim. Поведенческая модель написана на языке Verilog HDL. Для апробации использовалась матрица FPGA Altera Cyclone® IV 4CE115. На основе полученной поведенческой модели разработана тестирующая модель и осуществлена симуляция цифровой схемы в среде ModelSim. **Основные результаты.** Предложена формула оценки количества тактов выполнения прямого и обратного дискретного вейвлет-преобразований от числа входных элементов блока и характеристик FPGA. В результате экспериментов получены зависимости количества тактов расчета дискретного вейвлет-преобразования в зависимости от размера стороны квадратного изображения при разных вариантах количества блоков параллельной обработки. Показано, что параллельная работа нескольких независимых модулей позволяет проводить одновременно обработку нескольких входных столбцов (строк) из исходного двумерного массива. Унификация модуля умножителя-сумматора повысила эффективность вычисления и сократила объем занимаемых аппаратных ресурсов. Конвейерная структура дискретного вейвлет-преобразования отличается меньшими аппаратными затратами при реализации блока вычисления и распределения памяти. В результате тестирования цифровой схемы получено, что разработанная структура блока повысила быстродействие дискретного вейвлет-преобразования, а также сократила стоимость системы на кристалле. **Практическая значимость.** Предложенная реализация блока двумерного прямого и обратного вейвлет-преобразований для банка фильтров Добеши 9/7 является законченным модулем и может быть использована как готовый сложно-функциональный блок для последующей разработки систем передачи изображений высокого качества в реальном времени.

Ключевые слова

вейвлет-преобразование, лифтинг-схема, FPGA, сложно-функциональный блок, цифровая обработка сигналов

Благодарности

Исследование выполнено при финансовой поддержке Российского научного фонда в рамках научного проекта № 22-21-00394 «Развитие нейросетевых методов повышения качества передачи цифровых изображений в интеллектуальных видеосистемах».

Ссылка для цитирования: Сай С.В., Зинкевич А.В. Вычислительные методы повышения быстродействия дискретного вейвлет-преобразования на базе FPGA // Научно-технический вестник информационных технологий, механики и оптики. 2023. Т. 23, № 1. С. 79–87. doi: 10.17586/2226-1494-2023-23-1-79-87

Computational methods to increase the speed of FPGA-based discrete wavelet transforms

Sergey V. Sai¹✉, Alexey V. Zinkevich²

^{1,2} Pacific National University, Khabarovsk, 680035, Russian Federation

¹ sai1111@rambler.ru✉, <https://orcid.org/0000-0002-4506-1857>

² zinkevich1985@mail.ru, <https://orcid.org/0000-0003-3231-8569>

Abstract

The article considers the computational methods and features of the construction of a complex functional block for the implementation of the discrete wavelet transform (DWT) Dobeshie 9/7 in digital image signal processing systems based on FPGA. We proposed a mathematical model and algorithms for the implementation of parallel and series-convector methods of signal processing to calculate the coefficients of a discrete bi-orthogonal Dobeshie wavelet 9/7 taking into account the architecture of used FPGA. The model is based on wavelet transform factorization methods using lifting schemes. In contrast to conventional lifting schemes, the proposed method and algorithms can increase the speed of FPGA calculations with simplified hardware implementation. CAD Quartus II and ModelSim are used as a development environment. The behavioral model is written in Verilog HDL. Altera Cyclone® IV 4CE115 was used as FPGA. On the basis of the obtained behavioral model the testing module was developed and the simulation of digital circuit in the ModelSim environment was carried out. The formula for estimating the number of clock cycles of the forward and reverse DWT has been proposed; on its basis the estimate of the number of parallel computations depending on the number of input elements and the characteristics of the FPGA was obtained. As a result of experiments the dependences of the number of cycles for DWT computation depending on the size of the side of a square image with different variants of the number of parallel processing blocks were obtained. It is shown that parallel work of several independent modules gives a possibility to conduct concurrent processing of several input columns (rows) from input 2D array, and unification of the multiplier-summing module allows to increase efficiency of calculations and to reduce volume of occupied hardware resources. Conveyor based DWT structure is characterized by less hardware costs in terms of implementation of the calculator unit and memory allocation. As a result of testing the digital circuit, it was found that the developed block structure can significantly increase the DWT speed as well as reduce the cost of the system on a chip. The proposed realization of the block of two-dimensional forward and reverse wavelet transform for the Dobeshie 9/7 filter bank forms a complete module and can be used as a ready-made complex functional block for further development of high quality image transmission systems in real time.

Keywords

wavelet transform, lifter circuit, FPGA, complex function block, digital signal processing

Acknowledgements

The research supported financially by the Russian Science Foundation within the scientific project no. 22-21-00394 “Development of neural network methods to improve the quality of digital image transmission in intelligent video systems”.

For citation: Sai S.V., Zinkevich A.V. Computational methods to increase the speed of FPGA-based discrete wavelet transforms. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 2023, vol. 23, no. 1, pp. 79–87 (in Russian). doi: 10.17586/2226-1494-2023-23-1-79-87

Введение

Дискретное вейвлет-преобразование (ДВП) является основой стандарта сжатия JPEG2000 [1] который имеет преимущество по сравнению со стандартом JPEG по визуальному качеству изображения при одинаковом коэффициенте компрессии. В системах видеокompresии на основе стандартов H.264 (H.265) используются алгоритмы для компрессии опорных (I) кадров видеоряда.

В работе [2] рассмотрены особенности алгоритма сжатия JPEG2000. Одна из особенностей — выполнение в кодере следующих основных процедур: преобразование цифровых RGB-сигналов исходного кадра изображения в сигналы YVU; выполнение ДВП для каждой из YVU-компонент изображения; квантование коэффициентов трансформации и обнуление малозначущих коэффициентов; энтропийное сжатие. Данные процедуры также используются в алгоритме JPEG, где вместо ДВП применено блочное дискретное косинус-преобразование.

Для двумерного сигнала изображение ДВП реализуется следующим образом. На первой итерации

одномерное ДВП выполняется последовательно по строкам и столбцам, затем формируются четыре матрицы (НН — низкочастотная фильтрация по строкам и столбцам; НВ — низкочастотная фильтрация по строкам и высокочастотная по столбцам; ВН — высокочастотная фильтрация по строкам и низкочастотная по столбцам; ВВ — высокочастотная фильтрация по строкам и столбцам) коэффициентов ДВП с размерами в два раза меньшими ширины и высоты изображения. На второй итерации происходит аналогичное преобразование для НН матрицы и т. д. Обычно достаточно 4–6 итераций. Такой процесс преобразований эквивалентен операции фильтрации сигналов изображения по частотным субдиапазонам для каждой итерации. Таким образом, в процессе итераций образуются масштабированные (уменьшенные) в два раза копии исходного изображения и его дополнение в виде высокочастотных составляющих.

Основные проблемы использования JPEG2000 связаны с большим количеством вычислений, в том числе работы блока для получения коэффициентов двумерного ДВП. Особенно эти проблемы актуальны при об-

работке изображений с высоким разрешением, например, в форматах UHD, 2K или 4K. Реализация ДВП на специализированных интегральных схемах (Application Specific Integrated Circuit, ASIC) или программируемых вентильных матрицах (Field-Programmable Gate Array, FPGA) с применением параллельных вычислений позволяет повысить эффективность и скорость кодирования изображений в реальном масштабе времени.

В работе [3] предложена реализация ДВП путем объединения вейвлет-преобразования и банка фильтров. Добеши применила ДВП к различным изображениям и предложила множество популярных вейвлет-фильтров, включая фильтры 9/7 и 5/3 [4]. В [5] представлена схема лифтинга, и показано, что любое ДВП можно разложить на последовательность простых шагов фильтрации и тем самым ускорить вычислительную способность преобразования.

Для аппаратной реализации вейвлет-преобразований известно несколько архитектур, которые можно разделить на последовательные и параллельные. Часть последовательных архитектур основана на использовании систолических матриц, либо на конвейере, более эффективно реализующем вычислительный процесс [6–17]. Работы [4, 5], использующие факторизацию для выполнения вейвлет-преобразований, и предложенные в них лифтинговые схемы, позволяют значительно упростить аппаратную реализацию ДВП. Чаще всего в подобных работах основное внимание уделяется минимизации арифметической части устройства или используемой памяти. Однако для устройств, имеющих внутреннюю память, размерность арифметического блока не так важна. При этом важно минимизировать объем используемой памяти и повысить быстродействие вычислений, учитывая особенности реализации в конкретной FPGA.

В настоящей работе рассмотрены методы и алгоритмы реализации ДВП на основе биортогонального вейвлета Добеши 9/7 с использованием лифтинговой схемы. Цель работы — сравнительная оценка быстродействия и занимаемых аппаратных ресурсов двух архитектур (параллельной и конвейерной), а также сравнение по количеству вычислительных операций, полученных в [18] при программной реализации ДВП.

Вейвлет-преобразование на основе лифтинговой схемы

На сегодняшний день существуют два подхода к вычислению ДВП, основанных на свертке и лифтинговой схеме. Традиционный подход на основе свертки представляет собой свертку вейвлет-функции с сигналом. Такая функция представляется фильтром с конечной импульсной характеристикой, состоящим из фильтров нижних и верхних частот. Затем применяется операция децимации, цель которой сократить избыточность хранимой информации. В результате получается объем данных коэффициентов ДВП, равный по размерности исходному сигналу. При этом первая половина данных представлена низкочастотными коэффициентами, а вторая половина — высокочастотными.

Второй подход, основанный на лифтинговой схеме, снижает вычислительную сложность в два раза и позволяет экономить память. Схема прямого лифтинг-преобразования реализуется с помощью четырех базовых этапов: разбиения (S), предсказания (P), обновления (U) и масштабирования (K). Функции $P(z)$ и $U(z)$ представляются через z -преобразование. Выходные коэффициенты LP и HP соответствуют низкочастотной и высокочастотной составляющей (рис. 1).

Этап разбиения делит входной сигнал на четные (s) и нечетные (d) элементы:

$$s_i^{(0)} = x_{2i}; d_i^{(0)} = x_{2i+1},$$

где x_{2i} и x_{2i+1} — четные и нечетные элементы входного сигнала.

Значения элементов на каждом этапе преобразования вычисляются по формулам:

$$d_i^{(n)} = d_i^{(n-1)} + \sum_k P_n(k) s_k^{(n-1)}, n \in [1, 2, \dots, N];$$

$$s_i^{(n)} = s_i^{(n-1)} + \sum_k U_n(k) d_k^{(n)}, n \in [1, 2, \dots, N],$$

где i и k — индексы вычисляемых элементов; P_n и U_n — функции предсказания и обновления на n -ом уровне.

Для банка фильтров Добеши 9/7 количество уровней $N = 2$, и используется четыре этапа для вычисления вейвлет-преобразования. Выход последнего этапа предсказания умножается на коэффициент K_1 , а последнего этапа обновления — на K_0 .

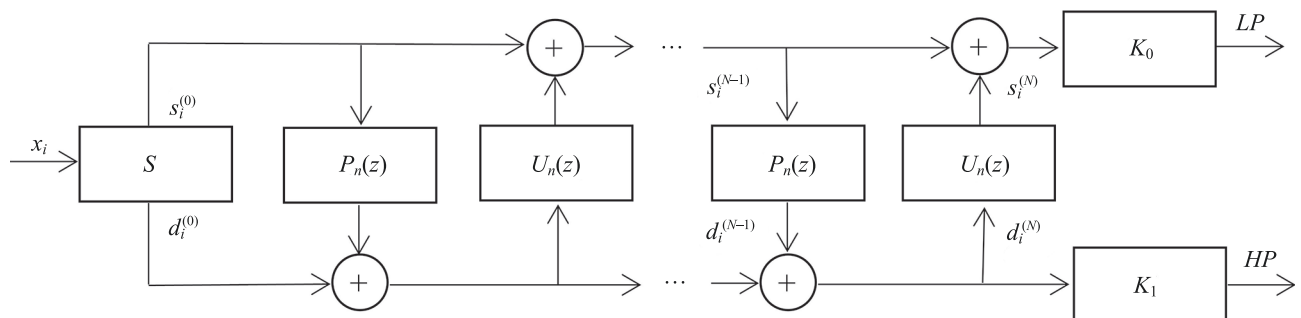


Рис. 1. Лифтинговая схема прямого вейвлет-преобразования

Fig. 1. Lifting scheme of direct wavelet transform

Алгоритм работы лифтинговой схемы прямого вейвлет-преобразования можно представить в виде выражения:

$$\begin{bmatrix} LP(z) \\ HP(z) \end{bmatrix} = \begin{bmatrix} s_m(z) \\ d_m(z) \end{bmatrix} = \tilde{\mathbf{P}}(z) \begin{bmatrix} s(z) \\ d(z) \end{bmatrix}, \quad (1)$$

где

$$\tilde{\mathbf{P}}(z) = \begin{bmatrix} K & 0 \\ 0 & \frac{1}{K} \end{bmatrix} \left\{ \prod_{k=m} \begin{bmatrix} 1 & Q_{2k}(z) \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ Q_{2k-1}(z) & 1 \end{bmatrix} \right\},$$

$LP(z)$ и $HP(z)$ — низкочастотные и высокочастотные коэффициенты сигнала $X(z)$; $Q_{2k}(z)$ и $Q_{2k-1}(z)$ — многочлены Лорана, полученные из факторизации многофазной матрицы, представляющей фильтр по алгоритму Евклида [5, 6]; K и $1/K$ — ненулевые коэффициенты масштабирования.

Применяя факторизацию лифтинговой схемы (1) для банка фильтров Добеши 9/7 получим:

$$\tilde{\mathbf{P}}(z) = \begin{bmatrix} K & 0 \\ 0 & \frac{1}{K} \end{bmatrix} \begin{bmatrix} 1 & d(1+z) \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ c(1+z^{-1}) & 1 \end{bmatrix} \times \\ \times \begin{bmatrix} 1 & b(1+z) \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ a(1+z^{-1}) & 1 \end{bmatrix}, \quad (2)$$

где $a = -1,586134342$; $b = -0,0529801185$; $c = 0,882911076$; $d = 0,443506852$; $K = 1,149604398$.

В выражении (2) и схеме на рис. 1 этап масштабирования представляет собой операцию умножения на ненулевые коэффициенты K и $1/K$. Так как умножители обычно — дорогостоящие логические ресурсы, то желательно минимизировать их количество в схеме. Схему вычисления можно изменить за счет внесения коэффициентов K и $1/K$ следующим образом:

$$\tilde{\mathbf{P}}(z) = \begin{bmatrix} K & dK^2(1+z) \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ \frac{c}{K}(1+z^{-1}) & \frac{1}{K} \end{bmatrix} \times \\ \times \begin{bmatrix} 1 & b(1+z) \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ a(1+z^{-1}) & 1 \end{bmatrix}, \quad (3)$$

где $dK^2 = 0,586134341$; $c/K = 0,7680129595$.

В результате такого преобразования этап масштабирования (K) отдельно не выполняется. Это позволяет унифицировать процесс вычисления для этапов предсказания и обновления, который будет включать три операции умножения и две — сложения.

Схема обратного лифтинг-преобразования реализуется также с помощью четырех базовых этапов и напоминает зеркальное отражение схемы прямого вейвлет-анализа, где этап разбиения (S) заменяется этапом — объединения (M). Получим выражение для вычисления коэффициентов обратного вейвлет-преобразования:

$$\tilde{\mathbf{P}}(z) = \begin{bmatrix} 1 & 0 \\ -a(1+z^{-1}) & 1 \end{bmatrix} \begin{bmatrix} 1 & -b(1+z) \\ 0 & 1 \end{bmatrix} \times \\ \times \begin{bmatrix} 1 & 0 \\ -c(1+z^{-1}) & K \end{bmatrix} \begin{bmatrix} \frac{1}{K} & -dK(1+z) \\ 0 & 1 \end{bmatrix}, \quad (4)$$

где $dK = 0,5098574276$.

Аппаратная реализация ДВП на базе FPGA

Отметим, что микросхемы FPGA известны своими возможностями по цифровой обработке сигналов (Digital Signal Processor, DSP) в формате фиксированной точки. Однако вычисляемые коэффициенты ДВП имеют дробный формат, что требует представления и обработки чисел в формате с плавающей точкой (IEEE 754¹). Использование такого формата усложняет схему и приводит к затратам большего количества ресурсов FPGA. Чтобы этого избежать, необходимо представить коэффициенты целыми значениями, умножив их на 2^k , где k — число двоичных разрядов для сдвига данных влево.

Входными данными изображения являются значения сигналов яркости (Y) или цветности (U , V) в диапазоне от 0 до 255, что соответствует стандартному значению в 8 бит на пиксел. В результате преобразования диапазон будет расширен, а также станет знаковым. Чтобы хранить целую часть промежуточных и выходных коэффициентов потребуется диапазон от минус 1024 до 1024, что будет занимать в дополнительном коде 11 бит. Для хранения дробной части потребуется увеличить число двоичных разрядов, задав общее количество бит, исходя из разрядности типовых умножителей FPGA равной 18, либо кратной 18.

Алгоритм прямого вейвлет-преобразования по выражению (3) для сигнала яркости приведен на рис. 2. «Входная строка» и «Выходная строка» — данные, хранящиеся во внутренней или внешней памяти FPGA. Значения считываются из памяти, проходят все этапы вычислений и записываются обратно. В отдельных блоках памяти (на рис. 2 отмечены одним цветом), хранятся промежуточные результаты четных и нечетных элементов. Обозначения «Е» и «О» являются четными и нечетными позициями, а их индексы соответствуют порядку вычисления. Коэффициенты ($a_1 - a_4$, k_1 , k_2) задаются согласно коэффициентам, указанным в выражении (3), где $a_1 = a$, $a_2 = b$, $a_3 = c/K$, $a_4 = dK^2$, $k_1 = 1/K$, $k_2 = K$. Также имеются единичные коэффициенты на первом этапе предсказания и обновления.

Процесс симметричного расширения — часть основного алгоритма преобразования и обеспечивает вычисление на границах входных данных. Для его эффективной реализации используется внутренняя двухпортовая память, в результате необходимые для расчета операции — чтение и запись, происходят одновременно и независимо по двум шинам. На этапах «Предсказание» и «Обновление» выполняются операции с одинаковым числом умножителей и сумматоров. Поскольку этапы исполняются последовательно, и на каждом происходит три умножения и два сложения, то для расчета используется один модуль. В результате выполнения этапов получим коэффициенты одномерного уровня прямого ДВП. Рассчитанные коэффициенты записываются обратно в исходную строку и заменяют в памяти предыдущие данные. В процессе формирования выходной строки происходит группировка низкоча-

¹ [Электронный ресурс]. Режим доступа: <https://ieeexplore.ieee.org/document/4610935> (дата обращения: 19.12.2022).

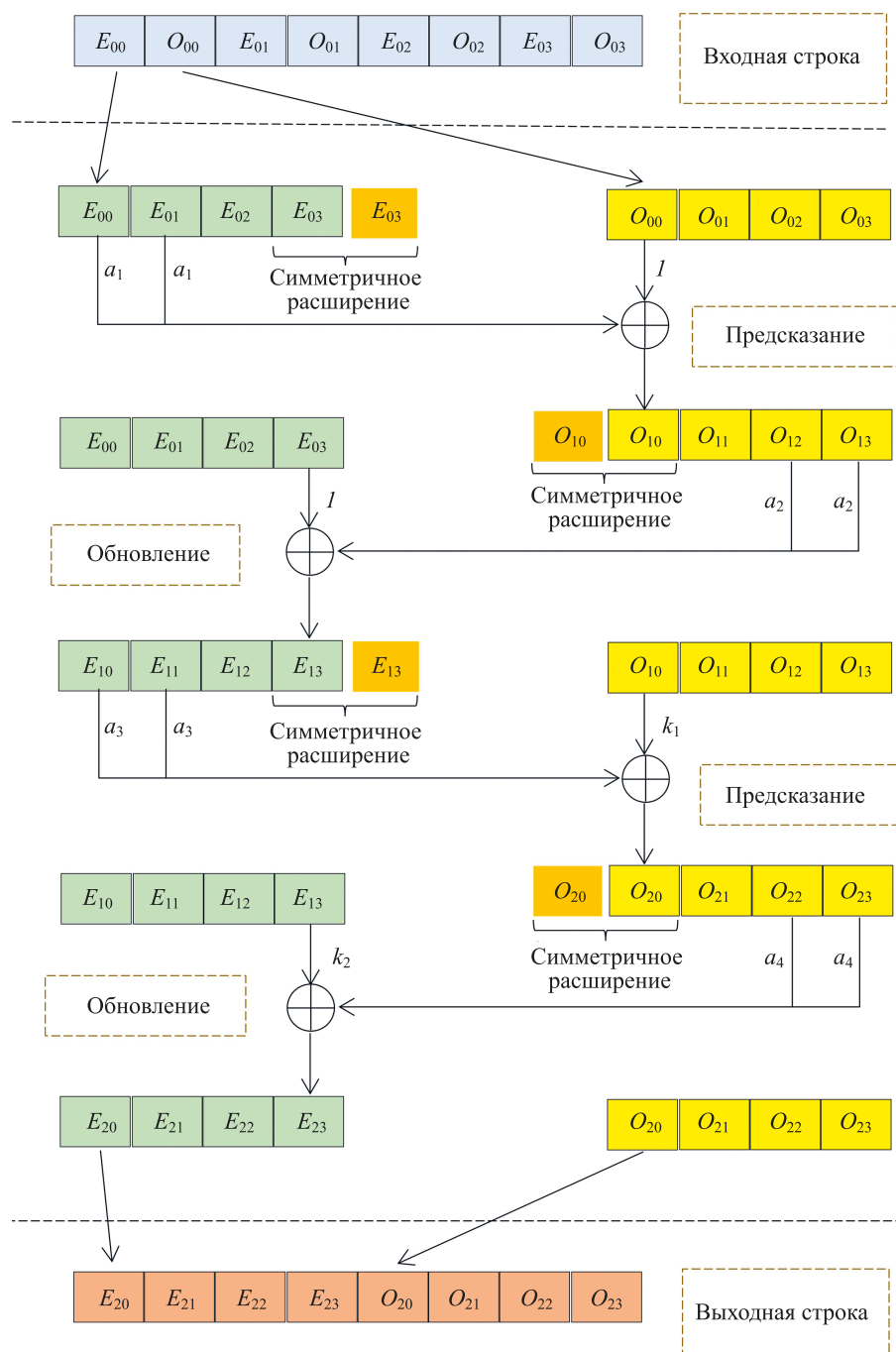


Рис. 2. Алгоритм прямого вейвлет-преобразования

Fig. 2. Direct wavelet transform algorithm

стотных ($E_{20}-E_{23}$) и высокочастотных ($O_{20}-O_{23}$) коэффициентов. После этого одномерное ДВП повторяется для столбцов и получаются итоговые коэффициенты двумерного преобразования.

Схема обратного ДВП может быть получена из выражения (4) и имеет зеркальную структуру схемы прямого вейвлет-преобразования (рис. 2) со следующими значениями коэффициентов: $a_1 = 1,58613$; $a_2 = 0,05298$; $a_3 = -0,88291$; $a_4 = -0,50985$; $k_1 = 1,14960$; $k_2 = 0,86986$.

Для известного количества входных элементов по предложенному алгоритму можно вычислить количе-

ство тактов для одномерного прямого или обратного вейвлет-преобразований:

$$N = 4(w/2 + 1 + p), \quad (5)$$

где w — количество входных элементов; l и p — количество тактов задержки в модуле умножения-сложения (1 такт) и чтения из памяти (1 такт).

Для одного уровня двумерного ДВП количество тактов будет удвоено.

Параллельное и конвейерное вычисления ДВП

Параллельное выполнение ДВП позволяет разделить обработку независимых данных на несколько блоков вычисления. Поскольку процесс двумерного преобразования состоит из последовательного расчета сначала по столбцам, а затем по строкам, то выполнить эти два процесса параллельно невозможно. Таким образом, представляется возможным выполнять параллельно вычисления по строкам или столбцам. В этом случае происходит считывание столбца (строки), его обработка и запись обратно. Во время обработки текущего столбца возможно считывание и обработка следующего и т. д. Если взять количество тактов для расчета одномерного ДВП из выражения (5), то можно рассчитать количество параллельных вычислений по формуле:

$$L = 2(N + w/2 + 1)/w. \quad (6)$$

Так, для столбца (строки) длиной $w = 8$, по выражению (5) получим $N = 4(8/2 + 1 + 1) = 24$ и по выражению (6) — $L = 2(24 + 8/2 + 1)/8 = 7,25$. Отсюда следует, что при использовании семи параллельно работающих вычислительных блоков расчет двумерного ДВП будет выполнен максимально быстро. Однако, если увеличить длину строки, например, $w = 2048$, то $N = 4(2048/2 + 1 + 1) = 4104$ и $L = 2(4104 + 2048/2 + 1)/2048 \approx 5$. Такое изменение связано со временем простоя блока вычисления в моментах, когда осуществляется доступ к памяти.

На рис. 3 приведена зависимость количества тактов для вычисления ДВП в зависимости от размера стороны квадратного изображения при некоторых вариантах количества блоков параллельной обработки.

Выполним оценку корректности рассмотренного алгоритма. Для этого запишем код блока вычисления на языке Verilog HDL, разработаем Testbench (тестирующий модуль) и осуществим симуляцию цифровой схемы в среде ModelSim. Для параллельного расчета ДВП реализован следующий алгоритм. Процесс начинается с чтения из двумерного массива двухпортовой памяти первого столбца. Считывание четного и нечетного элементов столбца происходит одновременно с их записью в память первого блока вычисления. Когда

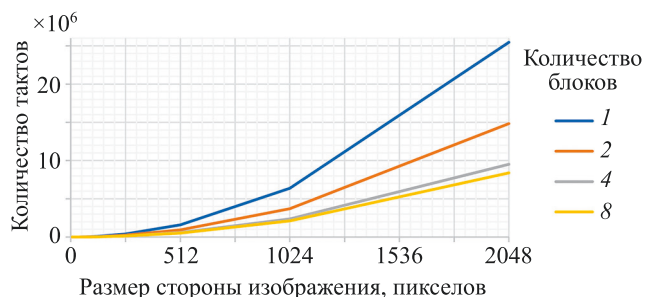


Рис. 3. Зависимость количества тактов для вычисления двумерного дискретного вейвлет-преобразования при различном количестве блоков параллельной обработки

Fig. 3. Dependence of the number of cycles for calculating a two-dimensional DWT with a different number of parallel processing units

данные первого столбца скопированы в память первого блока вычисления, он начинает процесс расчета одномерного ДВП, а элементы из второго столбца двумерного массива аналогично копируются во второй блок вычисления. Когда восьмой блок вычисления получил входные данные, первый уже закончил процесс расчета, и начинается последовательное считывание обратно в двумерный массив. Если все столбцы скопированы, вычисление аналогично повторяется для всех строк двумерного массива. Вычисление последующего уровня происходит тогда, когда получены низкочастотные коэффициенты текущего уровня. При такой организации расчета ДВП занимаемые ресурсы FPGA в основном определяются количеством умножителей и размером памяти для хранения двумерного массива.

Основное ограничение в рассмотренном вычислении — ожидание завершения предыдущего этапа. Так, в формуле (5), умножение на коэффициент 4 показывает, что необходимо выполнить этапы предсказания и обновления и далее повторить их еще раз. При этом промежуточные данные сохраняются в память, что при последующем доступе снова вносят задержку. Заметим (как видно из рис. 2), что для начала вычисления этапа обновления не обязательно ожидать окончания вычисления этапа предсказания, достаточно только первых двух коэффициентов. Аналогично происходит и на последующих этапах — для второго этапа обновления требуется только пара коэффициентов второго этапа предсказания. В этом случае необходимость в памяти для хранения промежуточных строк (столбцов) отсутствует. Исходные данные поступают последовательно на вход блока вычислителя по одному элементу за такт, и через несколько тактов задержки, вносимой конвейером, на выходе последовательно появляются коэффициенты ДВП. Для хранения промежуточных коэффициентов различных этапов используется внутренняя регистровая память с параллельным чтением данных. На FPGA такое устройство можно реализовать с помощью триггеров, входящих в состав типовых конфигурируемых блоков.

Для реализации конвейерной схемы требуется два модуля умножителя-сумматора, аналогичные использовались при параллельном выполнении ДВП (имеется в виду модуль, содержащий три умножителя и два сумматора, используемый для параллельного выполнения ДВП), а также 11 регистров для хранения текущих коэффициентов. С учетом задержки в блоке умножителя-сумматора в один такт задержка конвейера составляет 7 тактов. Это означает, что вычисленные коэффициенты ДВП появляются на выходе схемы спустя 7 тактов. В результате для входной строки длины w вычисление займет $N_k = (w + 7)$ тактов.

В случае вычисления двумерного ДВП чтение и запись двухпортовой памяти происходит одновременно по разным адресам с разницей в 7 тактов. Зависимость количества тактов для вычисления двумерного ДВП практически не отличается от варианта с количеством блоков параллельной обработки равной 8. Однако в этом случае количество затрачиваемых ресурсов FPGA меньше. Так, например, блоков умножителей-сумматоров в четыре раза меньше, а память для хранения

Таблица. Результаты сравнения быстродействия и занимаемых ресурсов архитектур
Table. Results of comparison of performance and resources occupied architectures

Характеристика	Архитектура			
	Программная реализация [18]	Параллельная ($L = 4$)	Параллельная ($L = 8$)	Конвейерная
Количество умножителей	—	24	48	12
Количество тактов вычисления	4 456 448	148 736	131 200	131 079
Занимаемая внутренняя память, бит	—	1 198 080	1 198 080	1 179 648

промежуточных четных и нечетных коэффициентов, суммарно 4 строки (столбца), не используется.

параллельной и конвейерной архитектур для квадратного изображения со стороной 256 пикселей.

Апробация алгоритмов на FPGA

Для апробации разработанных алгоритмов была выбрана FPGA Altera Cyclone® IV 4CE115. Выбранная FPGA имеет следующие характеристики: 114 480 логических элементов; 3888 кбит встроенной блочной памяти; 532 встроенных аппаратных умножителя разрядностью 9 бит; максимальная рабочая частота 250 МГц.

Использование внутренней памяти, в силу ее ограниченного объема, позволяет реализовать ДВП только для квадратных изображений со стороной не более 256 пикселей. Также возможны варианты эффективного параллельного вычисления, когда строки и столбцы исходного изображения будут кратны количеству блоков вычисления, определенных по формуле (6).

В ходе синтеза схемы, для достижения требуемой разрядности коэффициентов, умножители объединены попарно. Таким образом, становится возможным обрабатывать значения 18 битных чисел. При этом первые 11 бит будут хранить целую часть коэффициентов, а остальные 7 бит — дробную часть. Для перехода от чисел с плавающей точкой к числам с фиксированной точкой входные данные необходимо умножить на 2^7 путем сдвига влево на 7 разрядов, тогда все промежуточные и выходные коэффициенты будут располагаться в диапазоне от 1 до 131 072. При переводе выходных коэффициентов обратно необходимо выполнить их сдвиг вправо на 7 разрядов. Один модуль умножения-сложения в такой конфигурации использует шесть 9-битных умножителей для операций с фиксированной точкой.

Средства анализа и синтеза в среде Quartus II позволяют оценить максимальную рабочую частоту блока вычисления. Поскольку в общей схеме самой «медленной» является операция умножения, то она ограничивает максимальную тактовую частоту величиной 164 МГц. При этом задержка между вводом данных в модуль умножения-сложения и получением результата составит три такта.

В таблице приведены результаты сравнения быстродействия и занимаемых ресурсов программной [18],

Заключение

Предложенные реализации параллельной и конвейерной архитектур двумерного прямого и обратного вейвлет-преобразований для банка фильтров Добеши 9/7 являются законченным решением и могут быть использованы как готовый сложно-функциональный блок для последующей разработки системы кодирования и передачи изображений.

Параллельная работа нескольких независимых модулей дает возможность проводить одновременно обработку нескольких входных столбцов (строк) из исходного двумерного массива, а унификация модуля умножителя-сумматора позволяет повысить эффективность вычисления и сократить объем занимаемых аппаратных ресурсов. Конвейерная архитектура дискретного вейвлет-преобразования, ориентированная на синтез в FPGA или в функциональность сложно-функционального блока, отличается меньшими аппаратными затратами при реализации блока вычисления и распределения памяти. В результате можно повысить быстродействие, а также уменьшить стоимость системы на кристалле.

Использование FPGA с небольшим объемом внутренней памяти ограничивает реализацию блока дискретного вейвлет-преобразования для изображений большого размера. При этом обработка может выполняться блоками, согласно стандарту JPEG2000, либо, при использовании внешней памяти и некоторого изменения в модуле, целиком для всего изображения.

Предложенные алгоритмы, реализованные в виде поведенческой модели на языке Verilog HDL, при незначительной корректировке могут быть использованы на базе отечественных аналогов, например, в программируемых логических интегральных схемах производства АО «ВЗПП-С» (г. Воронеж).

К перспективным направлениям исследования относится разработка и внедрение в видеосистему на кристалле дополнительных к блоку дискретного вейвлет-преобразования модулей: нейросетевого классификатора блоков изображения, адаптивного квантования и сжатия.

Литература

References

1. Taubman D., Marcellin M.W. *JPEG2000: Image Compression Fundamentals, Standards and Practice*. Kluwer Academic Publishers, 2002. 774 p.
2. Воробьев В.И., Грибунин В.Г. Теория и практика вейвлет-преобразования. СПб.: Изд-во ВУС, 1999. 208 с.
3. Mallat S. A theory for multiresolution signal decomposition: The wavelet representation // *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 1989. V. 11. N 7. P. 674–693. <https://doi.org/10.1109/34.192463>
4. Cohen A., Daubechies I., Feauveau J.-C. Biorthogonal bases of compactly supported wavelets // *Communications on Pure and Applied Mathematics*. 1992. V. 45. N 5. P. 485–560. <https://doi.org/10.1002/cpa.3160450502>
5. Daubechies I., Sweldens W. Factoring wavelet transforms into lifting steps // *Journal of Fourier Analysis and Applications*. 1998. V. 4. N 3. P. 247–269. <https://doi.org/10.1007/BF02476026>
6. Sameen I., Chang Y.C., Ng M.S., Goi B.-M., Ooi C.-P. A Unified FPGA-based system architecture for 2-D discrete wavelet transform // *Journal of Signal Processing Systems*. 2013. V. 71. N 2. P. 123–142. <https://doi.org/10.1007/s11265-012-0687-1>
7. Rekha K., Ravi K. Design of high speed lifting based DWT using 9/7 wavelet transform for image compression // *Proc. of the International Conference on Recent Advances in Electronics and Communication Technology (ICRAECT)*. 2017. P. 132–137. <https://doi.org/10.1109/ICRAECT.2017.38>
8. Zhong X., Jiang H., Cao H., Yang R. Efficient lifting based CDF9/7 wavelet transform using fixed point // *Proc. of the 3rd International Congress on Image and Signal Processing*. 2010. P. 3094–3097. <https://doi.org/10.1109/CISP.2010.5648075>
9. Dimitroulakos G., Zervas N., Sklavos N., Goutis C. An efficient VLSI implementation for forward and inverse wavelet transform for JPEG2000 // *Proc. of the 14th International Conference on Digital Signal Processing Proceedings (DSP)*. V. 1. 2002. P. 233–236. <https://doi.org/10.1109/ICDSP.2002.1027877>
10. Grzeszczak A., Yeap T., Panchanathan S. VLSI architecture for discrete wavelet transform // *Proc. of the Canadian Conference on Electrical and Computer Engineering*. V. 2. 1994. P. 461–464. <https://doi.org/10.1109/CCECE.1994.405788>
11. Grzeszczak A., Mandal M., Panchanathan S. VLSI implementation of discrete wavelet transform // *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*. 1996. V. 4. N 4. P. 421–433. <https://doi.org/10.1109/92.544407>
12. Swami S., Mulani A. An efficient FPGA implementation of discrete wavelet transform for image compression // *Proc. of the International Conference on Energy, Communication, Data Analytics and Soft Computing (ICECDS)*. 2017. P. 385–3389. <https://doi.org/10.1109/ICECDS.2017.8390088>
13. Nagaraj P., Rajasekaran M., Muneeswaran V., Sudar K.M., Gokul K. VLSI implementation of image compression using TSA optimized discrete wavelet transform techniques // *Proc. of the Third International Conference on Smart Systems and Inventive Technology (ICSSIT)*. 2020. P. 667–670. <https://doi.org/10.1109/ICSSIT48917.2020.9214220>
14. Sakthivel S. Sankar A. Real time watermarking of grayscale images using integer DWT transform // *Proc. of the International Conference on VLSI Systems, Architectures, Technology and Applications (VLSI-SATA)*. 2016. P. 7593056. <https://doi.org/10.1109/VLSI-SATA.2016.7593056>
15. Xiong C., Tian J., Liu J. A fast VLSI architecture for two-dimensional discrete wavelet transform based on lifting scheme [image compression applications] // *Proc. of the 7th International Conference on Solid-State and Integrated Circuits Technology*. V. 3. 2004. P. 1661–1664. <https://doi.org/10.1109/ICSICT.2004.1435150>
16. Xiong C., Tian J., Liu J. Efficient architectures for two-dimensional discrete wavelet transform using lifting scheme // *IEEE Transactions on Image Processing*. 2007. V. 16. N 3. P. 607–614. <https://doi.org/10.1109/TIP.2007.891069>
17. Zhang C., Long Y., Kurdahi F. A hierarchical pipelining architecture and FPGA implementation for lifting-based 2-D DWT // *Journal of Real-Time Image Processing*. 2007. V. 2. N 4. P. 281–291. <https://doi.org/10.1007/s11554-007-0057-6>
18. Сай С.В., Шоберг А.Г. Способ управления качеством передачи мелких структур изображений в стандарте JPEG2000 // *Компьютерная оптика*. 2020. Т. 44. № 3. С. 401–408. <https://doi.org/10.18287/2412-6179-co-616>
1. Taubman D., Marcellin M.W. *JPEG2000: Image Compression Fundamentals, Standards and Practice*. Kluwer Academic Publishers, 2002. 774 p.
2. Vorobev V.I., Gribunin V.G. *Principles and Practices of the Wavelet Transform*. St.Petersburg, 1999, 208 p. (in Russian)
3. Mallat S. A theory for multiresolution signal decomposition: The wavelet representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1989, vol. 11, no. 7, pp. 674–693. <https://doi.org/10.1109/34.192463>
4. Cohen A., Daubechies I., Feauveau J.-C. Biorthogonal bases of compactly supported wavelets. *Communications on Pure and Applied Mathematics*, 1992, vol. 45, no. 5, pp. 485–560. <https://doi.org/10.1002/cpa.3160450502>
5. Daubechies I., Sweldens W. Factoring wavelet transforms into lifting steps. *Journal of Fourier Analysis and Applications*, 1998, vol. 4, no. 3, pp. 247–269. <https://doi.org/10.1007/BF02476026>
6. Sameen I., Chang Y.C., Ng M.S., Goi B.-M., Ooi C.-P. A Unified FPGA-based system architecture for 2-D discrete wavelet transform. *Journal of Signal Processing Systems*, 2013, vol. 71, no. 2, pp. 123–142. <https://doi.org/10.1007/s11265-012-0687-1>
7. Rekha K., Ravi K. Design of high speed lifting based DWT using 9/7 wavelet transform for image compression. *Proc. of the International Conference on Recent Advances in Electronics and Communication Technology (ICRAECT)*, 2017, pp. 132–137. <https://doi.org/10.1109/ICRAECT.2017.38>
8. Zhong X., Jiang H., Cao H., Yang R. Efficient lifting based CDF9/7 wavelet transform using fixed point. *Proc. of the 3rd International Congress on Image and Signal Processing*, 2010, pp. 3094–3097. <https://doi.org/10.1109/CISP.2010.5648075>
9. Dimitroulakos G., Zervas N., Sklavos N., Goutis C. An efficient VLSI implementation for forward and inverse wavelet transform for JPEG2000. *Proc. of the 14th International Conference on Digital Signal Processing Proceedings (DSP)*. V. 1, 2002, pp. 233–236. <https://doi.org/10.1109/ICDSP.2002.1027877>
10. Grzeszczak A., Yeap T., Panchanathan S. VLSI architecture for discrete wavelet transform. *Proc. of the Canadian Conference on Electrical and Computer Engineering*. V. 2, 1994, pp. 461–464. <https://doi.org/10.1109/CCECE.1994.405788>
11. Grzeszczak A., Mandal M., Panchanathan S. VLSI implementation of discrete wavelet transform. *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, 1996, vol. 4, no. 4, pp. 421–433. <https://doi.org/10.1109/92.544407>
12. Swami S., Mulani A. An efficient FPGA implementation of discrete wavelet transform for image compression. *Proc. of the International Conference on Energy, Communication, Data Analytics and Soft Computing (ICECDS)*, 2017, pp. 385–3389. <https://doi.org/10.1109/ICECDS.2017.8390088>
13. Nagaraj P., Rajasekaran M., Muneeswaran V., Sudar K.M., Gokul K. VLSI implementation of image compression using TSA optimized discrete wavelet transform techniques. *Proc. of the Third International Conference on Smart Systems and Inventive Technology (ICSSIT)*, 2020, pp. 667–670. <https://doi.org/10.1109/ICSSIT48917.2020.9214220>
14. Sakthivel S. Sankar A. Real time watermarking of grayscale images using integer DWT transform. *Proc. of the International Conference on VLSI Systems, Architectures, Technology and Applications (VLSI-SATA)*, 2016, pp. 7593056. <https://doi.org/10.1109/VLSI-SATA.2016.7593056>
15. Xiong C., Tian J., Liu J. A fast VLSI architecture for two-dimensional discrete wavelet transform based on lifting scheme [image compression applications]. *Proc. of the 7th International Conference on Solid-State and Integrated Circuits Technology*. V. 3, 2004, pp. 1661–1664. <https://doi.org/10.1109/ICSICT.2004.1435150>
16. Xiong C., Tian J., Liu J. Efficient architectures for two-dimensional discrete wavelet transform using lifting scheme. *IEEE Transactions on Image Processing*, 2007, vol. 16, no. 3, pp. 607–614. <https://doi.org/10.1109/TIP.2007.891069>
17. Zhang C., Long Y., Kurdahi F. A hierarchical pipelining architecture and FPGA implementation for lifting-based 2-D DWT. *Journal of Real-Time Image Processing*, 2007, vol. 2, no. 4, pp. 281–291. <https://doi.org/10.1007/s11554-007-0057-6>
18. Sai S.V., Shoberg A.G. Quality control method of the transmission of fine image details in the JPEG2000. *Computer Optics*, 2020, vol. 44, no. 3, pp. 401–408. (in Russian). <https://doi.org/10.18287/2412-6179-co-616>

Авторы

Сай Сергей Владимирович — доктор технических наук, профессор, заведующий кафедрой, Тихоокеанский государственный университет, Хабаровск, 680035, Российская Федерация, [sc 14042260800](https://orcid.org/0000-0002-4506-1857), <https://orcid.org/0000-0002-4506-1857>, sai1111@rambler.ru

Зинкевич Алексей Владимирович — кандидат технических наук, доцент, Тихоокеанский государственный университет, Хабаровск, 680035, Российская Федерация, <https://orcid.org/0000-0003-3231-8569>, zinkevich1985@mail.ru

Authors

Sergey V. Sai — D.Sc., Professor, Head of Department, Pacific National University, Khabarovsk, 680035, Russian Federation, [sc 14042260800](https://orcid.org/0000-0002-4506-1857), <https://orcid.org/0000-0002-4506-1857>, sai1111@rambler.ru

Alexey V. Zinkevich — PhD, Associate Professor, Pacific National University, Khabarovsk, 680035, Russian Federation, <https://orcid.org/0000-0003-3231-8569>, zinkevich1985@mail.ru

Статья поступила в редакцию 22.08.2022
Одобрена после рецензирования 10.11.2022
Принята к печати 15.01.2023

Received 22.08.2022
Approved after reviewing 10.11.2022
Accepted 15.01.2023



Работа доступна по лицензии
Creative Commons
«Attribution-NonCommercial»

doi: 10.17586/2226-1494-2023-23-1-88-95
УДК 004.89

Диалоговая система на основе устных разговоров с доступом к неструктурированной базе знаний

Сергей Михайлович Маслюхин✉

ООО «ЦРТ-инновации», Санкт-Петербург, 194044, Российская Федерация
Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация
maslyukhin@speechpro.com✉, <https://orcid.org/0000-0002-9054-5252>

Аннотация

Предмет исследования. Представлен подход к построению задачно-ориентированной диалоговой системы (разговорного агента) с доступом к неструктурированной базе знаний на основе устных разговоров с применением аугментации письменной речи, имитирующей результаты распознавания устной речи, комбинирования предсказаний классификаторов, генерации текста, дополненной поиском. **Метод.** Предложенный подход предусматривает аугментацию обучающих данных двумя способами: преобразованием текста в речь и обратно с помощью систем синтеза и распознавания речи; заменой части слов на основе матрицы спутываний системы распознавания речи. Диалоговая система с доступом к неструктурированной базе знаний решает задачу обнаружения высказывания, для которого необходим поиск дополнительной информации в неструктурированной базе знаний. С этой целью выполнено обучение моделей Support Vector Machine, Convolutional Neural Network, Bidirectional Encoder Representations from Transformers и Generative Pre-trained Transformer 2. Лучшие из представленных моделей использованы при формировании предсказания путем взвешенного комбинирования. Осуществлен выбор подходящего текстового фрагмента из базы знаний и генерация обоснованного ответа. Поставленные задачи решены путем адаптации модели генерации текста, аугментированной поиском Retrieval Augmented Generation. **Основные результаты.** Выполнена апробация подхода на данных конкурса 10th Dialogue System Technology Challenge (DSTC10). По всем метрикам, кроме Precision, новый подход значительно превзошел результаты базовых моделей, предложенных организаторами конкурса DSTC10. **Практическая значимость.** Результаты работы могут найти применение при создании чат-бот систем, обеспечивающих автоматическую обработку обращений пользователей на естественном языке на основе неструктурированной базы знаний, например базы ответов на часто задаваемые вопросы.

Ключевые слова

диалоговые системы, разговорные агенты, поиск информации, текстовая аугментация, генерация, аугментированная поиском

Благодарности

Исследование выполнено за счет гранта Российского научного фонда (№ 22-11-00128, <https://rscf.ru/project/22-11-00128/>).

Ссылка для цитирования: Маслюхин С.М. Диалоговая система на основе устных разговоров с доступом к неструктурированной базе знаний // Научно-технический вестник информационных технологий, механики и оптики. 2023. Т. 23, № 1. С. 88–95. doi: 10.17586/2226-1494-2023-23-1-88-95

Dialogue system based on spoken conversations with access to an unstructured knowledge base

Sergei M. Masliukhin✉

STC-innovations Limited, Saint Petersburg, 194044, Russian Federation
ITMO University, Saint Petersburg, 197101, Russian Federation
maslyukhin@speechpro.com✉, <https://orcid.org/0000-0002-9054-5252>

Abstract

This paper describes an approach for constructing a task-oriented dialog system (a conversational agent) with an unstructured knowledge access based on spoken conversations including: written speech augmentation that simulates

the speech recognition results; combination of classifiers; retrieval augmented text generation. The proposed approach provides the training data augmentation in two ways: by converting the original texts into sound waves by a text-to-speech model and then transforming back into texts by an automated speech recognition model; injecting artificially generated errors based on phonetic similarity. A dialogue system with access to the unstructured knowledge base solves the task of detecting a turn, which requires searching for additional information in an unstructured knowledge base. For this purpose, the Support Vector Machine, Convolutional Neural Network, Bidirectional Encoder Representations from Transformers, and Generative Pre-trained Transformer 2 models were trained. The best of the presented models are used in the weighted combination. Next, a suitable text fragment is selected from the knowledge base and a reasonable answer is generated. The tasks are solved by adapting the retrieval augmented text generation model Retrieval Augmented Generation. The proposed method tested on the data from the 10th Dialogue System Technology Challenge. In all metrics, except Precision, the new approach significantly outperformed the results of the basic models proposed by the organizers of the competition. The results of the work can be used to create chat-bot systems that provide automatic processing of user requests in natural language based on an unstructured knowledge access, such as a database of answers to frequently asked questions.

Keywords

dialogue systems, conversational agents, information retrieval, text augmentation, retrieval augmented generation

Acknowledgements

This research is financially supported by the Russian Science Foundation (No. 22-11-00128, <https://rscf.ru/project/22-11-00128/>).

For citation: Masliukhin S.M. Dialogue system based on spoken conversations with access to an unstructured knowledge base. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 2023, vol. 23, no. 1, pp. 88–95 (in Russian). doi: 10.17586/2226-1494-2023-23-1-88-95

Введение

Традиционно задачно-ориентированные диалоговые системы сконцентрированы на предоставлении информации и выполнении действий в соответствии с запросами пользователей, которые могут быть обработаны только с использованием программного интерфейса и обращения к базам данных. Однако в дополнение к запросам, ориентированным на задачи, у пользователей также существуют потребности, которые требуют большей информации, кроме предоставленной из внутренних баз данных. Например, большинство разговорных агентов могут помочь пользователям забронировать отель, ресторан или купить билеты в кино, но они не отвечают на дополнительные возникающие вопросы, например: имеется ли парковка транспортных средств; разрешено ли приводить в зарезервированное место домашних животных или детей; или какова политика отмены бронирования. Для обработки таких запросов обычно нет записи в базе данных. С другой стороны, соответствующая информация уже доступна на веб-страницах в виде раздела ответов на часто задаваемые вопросы и отзывов клиентов для многих из этих сценариев, выходящих за рамки возможностей программного интерфейса. Поскольку современные диалоговые системы не включают эти внешние источники информации в задачно-ориентированное моделирование диалога, пользователям необходимо самим посещать веб-сайты, чтобы узнать любую дополнительную информацию, выходящую за рамки программного интерфейса, что делает диалоговые взаимодействия неэффективными. Данное исследование направлено на поддержку сценариев, в которых диалог не прерывается, когда у пользователей есть запросы, которые выходят за рамки программного интерфейса, но потенциально необходимая информация доступна во внешних источниках.

В работах [1–3] рассмотрены важность и сложность моделирования диалога с использованием внеш-

них источников знаний в открытом домене. В настоящей работе подходы, используемые при создании разговорных агентов в открытом домене, адаптированы под задачно-ориентированное моделирование диалога. В отличие от открытого домена, который предполагает наличие и использование широкого набора знаний о мире, например из википедии, при задачно-ориентированном моделировании диалога система располагает ограниченным пулом знаний, необходимых системе для ответа на вопросы в рамках решаемой задачи. При этом большое значение имеет точность выбора знаний из неструктурированной базы и их правильное использование при генерации ответа. В процессе работы, в первую очередь, система выполняет задачу обнаружения — определяет необходимость обращения к базе знаний, в случае, когда ответ на запрос не может быть получен на основе программного интерфейса. Для решения этой задачи проведено обучение моделей: Support Vector Machine (SVM) [4], Convolutional Neural Network (CNN) [5], Bidirectional Encoder Representations from Transformers (BERT) [6] и Generative Pre-Trained Transformer 2 (GPT-2) [7] — лучшие из которых использованы при формировании предсказания путем взвешенного комбинирования. Далее системой осуществлена задача выбора — поиск в базе знаний текстового фрагмента, содержащего необходимую для ответа информацию. При решении данной задачи проведено сравнение популярных моделей поиска, таких как Dense Passage Retrieval (DPR) [8] и BiEncoder [9]. Последний этап — формирование обоснованного ответа, при котором решена задача генерации. Для этого проведено дообучение генеративной модели Bidirectional and Auto-Regressive Transformers (BART) [10]. Также рассмотрен подход генерации, дополненной поиском Retrieval Augmented Generation (RAG), решающий задачи выбора и генерации в рамках одной модели, обучающейся методом сквозного (end-to-end) обучения.

Другой важный аспект — подготовка системы к работе с устной речью. Применяемые в настоящее время

мя диалоговые системы достигли многообещающих результатов в письменных диалогах, однако их использование непосредственно в устной речи затруднено из-за различий в распределении данных, включающих несоответствие между письменной и устной речью, а также дополнительные шумы из-за ошибок систем распознавания речи. Рассмотрены два способа аугментации обучающих данных для адаптации системы к устной речи: преобразование текста в речь и обратно с помощью систем синтеза и распознавания речи; замена части слов в обучающих данных на слова из матрицы спутываний системы распознавания речи.

Постановка задачи

Для оценки результатов работы предложенного подхода использованы данные второй задачи второго трека международного конкурса 10th Dialogue System Technology Challenge (DSTC10) [11]. В отличие от первого трека конкурса 9th Dialogue System Technology Challenge (DSTC9) [12], система оценена не на письменных диалогах, а на устных. По условиям задачи кросс-доменные разговорные агенты дают ответы на вопросы, которые невозможно сгенерировать на основе программного интерфейса и записей в базе данных, поэтому им приходится извлекать связанные пары вопрос-ответ из неструктурированной базы знаний часто задаваемых вопросов. На основе полученных пар вопрос-ответ агенты генерируют ответ на естественном языке. Конкурс DSTC10 не предусматривал специального обучающего набора. Вместо этого использован обучающий датасет первого трека DSTC9, состоящий из 72 518 письменных диалогов. Этот набор представляет собой расширенную версию набора данных Multi-Domain Wizard-of-Oz dataset (MultiWOZ 2.1) [13–15], в который добавлены запросы, для ответа на которые необходимо обладать дополнительной информацией, содержащейся в текстовых фрагментах. Текстовые фрагменты собраны со страниц часто задаваемых вопросов и состоят из пар вопрос-ответ. Они относительно короткие и охватывают четыре разных домена: отель, ресторан, поезд и такси. Первые два разделены на сущности. На основе текстовых фрагментов сформирована неструктурированная база знаний. В то время как в валидационном наборе данных использованы аналогичные фрагменты и локация, что и в обучающем, в тестовом наборе представлена новая локация — Сан-Франциско и новые текстовые фрагменты, некоторые из которых относятся к новому домену — достопримечательности. Около половины из 4181 диалога тестового набора данных получены путем расширения набора

данных MultiWOZ. Другая половина собрана из разговоров между людьми, посвященных туристическим поездкам в Сан-Франциско. Примерно десятая часть разговоров выполнена в устной форме и имеет аналогичные свойства, что и данные DSTC10. Отметим, что данные теста DSTC9 включают не результаты распознавания, а транскрипты, написанные человеком, без ошибок распознавания. Валидационный набор данных DSTC10 представляет собой те же 263 диалога, что и устная речь в тестовых данных DSTC9, но полученных с помощью системы распознавания речи. Тестовые данные состоят из 1988 дополнительных диалогов, собранных в той же локации. Сравнительные характеристики датасетов представлены в табл. 1.

В качестве системы автоматического распознавания речи при подготовке датасетов использована модель Wav2Vec 2.0 [16], предварительно обученная на 960 ч LibriSpeech [17] и дообученная с использованием 10 % данных из целевого домена. Далее получены топ-10 предсказаний с помощью языковой модели-декодера, построенной на основе KenLM [18], для всех письменных текстов из наборов данных MultiWOZ и DSTC9. Система достигла уровня ошибок в словах 24,09 %, что привело к заметным искажениям в данных. База знаний такая же, как и в тестовом наборе DSTC9. Дополнительная информация о датасетах представлена в [11] с описанием конкурса.

Описание подхода

Предложенный подход включает в себя несколько этапов и решение ряда задач. Первый этап — подготовка данных, направленная на формирование необходимых для обучения и оценки данных, которые затем использованы при решении всех последующих задач. Следующий этап включает в себя непосредственно реализацию задачно-ориентированной диалоговой системы с доступом к неструктурированной базе данных, которая решает задачи: обнаружения высказывания, для которого необходим поиск дополнительной информации в неструктурированной базе знаний; выбора подходящего текстового фрагмента из базы знаний; генерации обоснованного ответа. Общая структура предлагаемого подхода представлена на рисунке.

Подробная информация по каждому этапу представлена ниже.

Задача аугментации данных. На этапе подготовки данных решена задача аугментации обучающих данных для адаптации системы к данным, получаемым в результате распознавания речи.

Таблица 1. Сравнительные характеристики датасетов DSTC9 и DSTC10

Table 1. DSTC9 and DSTC10 comparative characteristics of datasets

Датасет	Набор данных					
	Обучающий		Валидационный		Тестовый	
	Письменная речь	Устная речь	Письменная речь	Устная речь	Письменная речь	Устная речь
DSTC9	72 518	0	9663	0	3918	263
DSTC10	72 518	0	0	263	0	1988



Рисунок. Общая структура предлагаемого подхода
Figure. The general structure of the proposed approach

Популярный способ имитации устной речи — зашумление письменной речи путем последовательного преобразования текстовых данных в речь и затем обратного преобразования речи в текст [19, 20]. В связи с тем, что система синтеза речи генерирует простую для распознавания речь, в рамках эксперимента в аудиосигналы дополнительно добавлены шумы, имитирующие запись в городской среде. Кроме применения английской модели синтеза речи, выполнен эксперимент с моделью для русского языка, поскольку она генерирует речь с сильным русским акцентом, которая сложнее для распознавания системами автоматического распознавания речи (Automated Speech Recognition, ASR).

Другой способ аугментации данных основан на анализе ошибок систем ASR в представленных организаторами конкурса DSTC10 валидационных данных, которые содержат 10 лучших предсказаний системы распознавания речи для каждого высказывания в диалоге. На основе сопоставления гипотез системы распознавания речи составлена карта частых акустических спутываний. Далее в обучающих данных случайным образом заменены слова, для которых есть созвучные из карты спутываний. Таким образом, датасет был размножен и получил версии, содержащие одну, пять и десять копий оригинального датасета с разными искажениями, внесенными в них.

В рамках данной работы исследованы перечисленные способы аугментации и проанализировано влияние

доли аугментированных данных в обучении на качество работы моделей.

Задача обнаружения. Обнаружение запросов, требующих обращения к базе знаний, сводится к задаче бинарной классификации: модель должна определить, требуется дополнительная информация из базы знаний или нет.

Проведем ряд экспериментов по сравнению различных моделей-классификаторов в рамках решаемой задачи: SVM, CNN, BERT и GPT-2. На вход моделей подадим данные текущего запроса без истории диалога, так как эти высказывания могут привести к неверному выбору класса, что было подтверждено в ходе экспериментов. В результате экспериментов выбран ряд моделей для участия в формировании финального предсказания. Веса моделей при комбинировании подобраны с помощью линейной регрессии на валидационных данных.

Задача выбора. Цель задачи выбора состоит в поиске наиболее подходящего документа из базы знаний для заданного диалога. Для каждого текстового фрагмента из базы знаний модель предсказывает, подходит он или нет, и выбирает документ с наивысшей оценкой релевантности. Выполним сравнение моделей BiEncoder и DPR. Они имеют схожую структуру и состоят из двух кодировщиков на основе предобученной модели BERT. Один кодировщик использован для кодирования контекста, а второй кодирует текстовые фрагменты, представляющие собой кандидатов. Для оценки схожести

применено значение скалярного произведения между вектором контекста и матрицей кандидатов. Разница в подходах, кроме незначительных архитектурных различий, заключается в способе негативного семплирования [21]. Для модели BiEncoder дистракторами являются верные ответы для других запросов из того же батча — части данных, обрабатываемой моделью за один шаг обучения. При обучении модели DPR выбраны некоторые количества сложных дистракторов из текстовых фрагментов для той же сущности/домена и простых дистракторов для других сущностей/доменов из базы. Для DPR доступна модель, предобученная на задаче выбора релевантных текстовых фрагментов на данных из википедии¹.

Задача генерации. Задача генерации предполагает поддержание естественного диалога в соответствии с контекстом диалога и извлеченным из базы знаний текстовым фрагментом.

Для решения этой задачи применена модель последовательность-в-последовательность с условной (на основе некоторого контекста) генерацией BART. Модель состоит из кодировщика и декодировщика. Кодировщик формирует векторное представление входной последовательности, содержащее контекст диалога и извлеченный на предыдущем этапе из базы знаний текстовый фрагмент. Декодировщик на основе полученного представления генерирует ответ на естественном языке. Модель обучена на истинно верных фрагментах из базы знаний, так как при ошибке на предыдущем этапе модель в любом случае не может сгенерировать верный ответ из-за недостатка информации.

Генерация, дополненная поиском. Модель RAG объединяет в себе модели DPR и BART и решает сразу обе задачи: выбора подходящих документов и генерации соответствующего ответа. Первые эксперименты проведены с оригинальной моделью RAG². Однако эксперимент оказался неудачным, так как оба компонента модели на данных из нового домена изначально давали очень плохие предсказания, и модель не обучалась. В связи с этим модель DPR была заменена на обученную на целевом домене модель BiEncoder. Модель BART была также дообучена на целевых данных. Затем модель RAG обучена в режиме end-to-end. Таким образом, получен наилучший результат на валидационных данных для подзадач выбора и генерации.

Эксперименты и результаты

Проведены эксперименты по увеличению количества данных за счет аугментации на задаче обнаружения. Для оценки результатов использованы стандартные метрики бинарной классификации: accuracy, precision, recall и F-score³. Зашумление данных путем

последовательного преобразования текстовых данных в речь и затем обратного преобразования речевых данных в текст не дало положительных результатов. Это связано с тем, что синтезированные данные оказались слишком простыми для системы распознавания речи и были распознаны с нулевой ошибкой, в том числе при наложении дополнительных шумов. При использовании русскоязычной модели синтеза речи получен обратный результат. Вывод системы распознавания оказался полностью отличным от оригинального текста.

Аугментация данных на основе карты акустических спутываний системы распознавания речи позволила заметно повысить качество на проверочных данных DSTC10. В ходе экспериментов проведено сравнение результатов при добавлении различного объема аугментированных данных. Каждая порция аугментированных данных содержит такое же количество данных, как и оригинальный датасет, так как получается путем полного прохода алгоритма аугментации по нему. В результате наилучший результат получен при добавлении десяти порций аугментированных данных (табл. 2).

Наилучшие результаты на задаче обнаружения показала модель GPT-2, обученная на аугментированных данных (табл. 3). Принято считать, что модель BERT лучше подходит для задачи классификации текстов. Вероятно, лучшие результаты модели GPT-2 связаны с размерами моделей: GPT-2 более чем в 3 раза больше модели BERT по количеству обучаемых параметров. Для расчета финального предсказания на тестовых данных DSTC10 использовано комбинирование лучших моделей. Для всех возможных комбинаций классификаторов, представленных в табл. 3, подобраны веса предсказаний моделей, участвующие в комбинировании, при помощи линейной регрессии на проверочном наборе данных. Результаты наилучшей комбинации: CNN и GPT-2 с весами 0,33 и 0,72 соответственно — совпадают с предсказаниями лучшей модели. Такой результат может быть связан с малым количеством примеров в проверочной выборке — 104 примера. Также подбор весов моделей при помощи линейной регрессии возможно недостаточно эффективен, и в дальнейшем планируются эксперименты с другими способами агрегации, например путем обучения двухслойного перцептрона.

В задаче выбора проведено сравнение моделей BiEncoder и DPR. Особое внимание было уделено обработке сложных случаев, когда модели предлагалось выбрать верный ответ из набора похожих кандидатов из одного домена. В этом случае была собрана тестовая выборка для отдельной оценки модели на сложных и простых наборах кандидатов. Для оценки результатов использована метрика полноты ранжирования (recall). Полнота показывает долю случаев, когда верный ответ оказался в топ- k предсказаний модели и обозначена — $Rn@k$, где n — общее количество кандидатов, из которых выбирает модель, k — максимальная позиция верного ответа в предсказании модели. При этом предсказание модели считается верным⁴. Подход

¹ [Электронный ресурс]. Режим доступа: <https://github.com/facebookresearch/DPR> (дата обращения: 17.12.2022).

² [Электронный ресурс]. Режим доступа: <https://huggingface.co/facebook/rag-sequence-base> (дата обращения: 17.12.2022).

³ [Электронный ресурс]. Режим доступа: https://scikit-learn.org/stable/modules/model_evaluation.html (дата обращения: 17.12.2022).

⁴ [Электронный ресурс]. Режим доступа: [https://en.wikipedia.org/wiki/Evaluation_measures_\(information_retrieval\)](https://en.wikipedia.org/wiki/Evaluation_measures_(information_retrieval)) (дата обращения: 17.12.2022).

Таблица 2. Результаты сравнения производительности модели GPT-2 на задаче обнаружения при добавлении различного объема аугментированных данных

Table 2. Comparison results of the GPT-2 model performance on the detection task when adding different amounts of augmented data

Данные в обучении	Количество порций аугментированных данных	Accuracy	Precision	Recall	F-score
Оригинальные DSTC9 (обучающие)	—	0,863	0,972	0,673	0,795
Оригинальные DSTC9 (все)	—	0,920	0,977	0,817	0,890
Аугментированные	1	0,932	0,989	0,837	0,906
	5	0,939	0,968	0,875	0,919
	10	0,947	0,959	0,904	0,931

Таблица 3. Результаты экспериментов с моделями классификаторов для задачи обнаружения

Table 3. Results of experiments with classifier models for the detection task

Модель	Accuracy	Precision	Recall	F-score
SVM	0,905	0,976	0,779	0,866
CNN	0,905	0,954	0,798	0,869
BERT	0,928	0,967	0,846	0,903
GPT-2	0,947	0,959	0,904	0,931
Комбинация	0,947	0,959	0,904	0,931

Таблица 4. Результаты сравнения моделей BiEncoder и DPR на задаче выбора

Table 4. Comparison results of BiEncoder and DPR models on the selection task

Модель	Коэффициент скорости обучения	Количество простых негативных примеров	Количество сложных негативных примеров	$R10@1$ для сложных негативных примеров	$R100@1$ для простых негативных примеров
BiEncoder	$1 \cdot 10^{-5}$	—	—	0,79	0,96
DPR	$1 \cdot 10^{-5}$	12	12	0,50	0,43
	$2 \cdot 10^{-6}$	12	12	0,54	0,52
	$2 \cdot 10^{-6}$	1	15	0,52	0,28

Таблица 5. Результаты экспериментов сравнения конвейера BiEncoder и BART с моделью RAG на задаче генерации

Table 5. Comparison results of the BiEncoder and BART pipeline with the RAG model on the generation task

Модель	$R12039@1$	$R12039@5$	BLEU-1	Meteor	Rouge-1
BiEncoder + BART	0,625	0,760	0,132	0,150	0,143
RAG	0,625	0,769	0,139	0,155	0,189

с использованием ответов из батча в качестве дистракторов оказался очень эффективным, и по результатам эксперимента модель BiEncoder значительно превзошла модель DPR на обоих наборах кандидатов (табл. 4). Для модели DPR применение одинакового количества сложных и простых дистракторов в обучении показало наилучшие результаты.

Модель RAG незначительно превосходит модель BiEncoder при выборе подходящих документов (табл. 5). В процессе обучения модель лучше учится обрабатывать сложные случаи, когда несколько кандидатов похожи, при этом верным является только один. Для оценки результатов экспериментов на задаче генерации использованы метрики, основанные на пересечении между верной и предсказанной последовательностями слов: BLEU-1, Meteor и Rouge-1, где 1 — размер

n -граммы, используемый при определении совпадений¹. Модель RAG превосходит модель BART на задаче генерации, так как позволяет эффективно учитывать топ-20 предсказаний модели поиска релевантных текстовых фрагментов, а не топ-1.

Финальный результат получен в два этапа. Вначале получены предсказания для задачи обнаружения с помощью комбинирования моделей классификаторов. Далее для запросов, требующих обращения к базе знаний, вызвана модель RAG, которая извлекла релевантные документы и сгенерировала ответы. По всем метрикам, кроме precision, предложенный подход

¹ [Электронный ресурс]. Режим доступа: <https://blog.paperspace.com/automated-metrics-for-evaluating-generated-text/> (дата обращения: 17.12.2022).

Таблица 6. Результаты сравнения предложенного подхода с базовой моделью на тестовых данных DSTC10

Table 6. Comparison results of the proposed approach with the baseline on DSTC10 test data

Подходы для сравнения	Задача обнаружения			Задача выбора	Задача генерации		
	Precision	Recall	F-score	R12039@1	BLEU-1	Meteor	Rouge-1
Базовая модель	0,897	0,674	0,769	0,495	0,125	0,152	0,136
Предложенный подход	0,888	0,890	0,889	0,572	0,145	0,158	0,178

значительно превосходит результаты базовых моделей, предложенных организаторами конкурса DSTC10 (табл. 6). Выбор модели в сторону наивысшего значения recall, а не precision, сделан осознанно, так как он позволяет пропустить как можно меньше запросов, требующих дополнительной информации из базы знаний.

Заключение

Полученные результаты могут найти широкое применение при создании чат-бот систем, обеспечивающих автоматическую обработку обращений пользователей

в различных сферах жизни. Например, в службах поддержки банков, медицинских и государственных учреждений и т. д., или в системах бронирования билетов, ресторанов, отелей и т. п. Предложенный подход также может быть использован в голосовых роботах, благодаря адаптации моделей к устной речи.

В качестве дальнейшего развития подхода предполагается доработка метода аугментации данных путем последовательного преобразования текстовых данных в речь и затем обратного преобразования речи в текст. Предполагается расширение возможностей поиска не только в базе ответов на часто задаваемые вопросы, но и в базе отзывов пользователей.

Литература

1. Moghe N., Arora S., Banerjee S., Khapra M.M. Towards exploiting background knowledge for building conversation systems // *Proc. of the 2018 Conference on Empirical Methods in Natural Language Processing*. 2018. P. 2322–2332. <https://doi.org/10.18653/v1/D18-1255>
2. Dinan E., Roller S., Shuster K., Fan A., Auli M., Weston J. Wizard of wikipedia: Knowledge-powered conversational agents // *arXiv*. 2019. arXiv:1811.01241. <https://doi.org/10.48550/arXiv.1811.01241>
3. Zhou K., Prabhume S., Black A.W. A dataset for document grounded conversations // *Proc. of the 2018 Conference on Empirical Methods in Natural Language Processing*. 2018. P. 708–713. <https://doi.org/10.18653/v1/D18-1076>
4. Hearst M., Dumais S., Osuna E., Platt J. Scholkopf B. Support vector machines // *IEEE Intelligent Systems and their Applications*. 1998. V. 13. N 4. P. 18–28. <https://doi.org/10.1109/5254.708428>
5. Johnson R., Zhang T. Convolutional neural networks for text categorization: Shallow word-level vs. deep character-level // *ArXiv*. 2016. arXiv:1609.00718. <https://doi.org/10.48550/arXiv.1609.00718>
6. Devlin J., Chang M.-W., Lee K., Toutanova K. BERT: Pre-training of deep bidirectional transformers for language understanding // *Proc. of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. Vol. 1 (Long and Short Papers). 2019. P. 4171–4186. <https://doi.org/10.18653/v1/N19-1423>
7. Radford A., Narasimhan K., Salimans T., Sutskever I. Improving language understanding by generative pre-training: preprint. 2018.
8. Karpukhin V., Oğuz B., Min S., Lewis P., Wu L., Edunov S., Chen D., Yih W.-T. Dense passage retrieval for open-domain question answering // *Proc. of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. 2020. P. 6769–6781. <https://doi.org/10.18653/v1/2020.emnlp-main.550>
9. Humeau S., Shuster K., Lachaux M., Weston J. Poly-encoders: Architectures and pre-training strategies for fast and accurate multi-sentence scoring // *arXiv*. 2020. arXiv:1905.01969. <https://doi.org/10.48550/arXiv.1905.01969>
10. Lewis M., Liu Y., Goyal N., Ghazvininejad M., Mohamed A., Levy O., Stoyanov V., Zettlemoyer L. BART: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension // *Proc. of the 58th Annual Meeting of the Association for Computational Linguistics*. 2020. P. 7871–7880. <https://doi.org/10.18653/v1/2020.acl-main.703>
11. Kim S., Liu Y., Jin D., Papangelis A., Hedayatnia B., Gopalakrishnan K., Hakkani-Tur D. DSTC10 Track Proposal:

References

1. Moghe N., Arora S., Banerjee S., Khapra M.M. Towards exploiting background knowledge for building conversation systems. *Proc. of the 2018 Conference on Empirical Methods in Natural Language Processing*, 2018, pp. 2322–2332. <https://doi.org/10.18653/v1/D18-1255>
2. Dinan E., Roller S., Shuster K., Fan A., Auli M., Weston J. Wizard of wikipedia: Knowledge-powered conversational agents. *arXiv*, 2019, arXiv:1811.01241. <https://doi.org/10.48550/arXiv.1811.01241>
3. Zhou K., Prabhume S., Black A.W. A dataset for document grounded conversations. *Proc. of the 2018 Conference on Empirical Methods in Natural Language Processing*, 2018, pp. 708–713. <https://doi.org/10.18653/v1/D18-1076>
4. Hearst M., Dumais S., Osuna E., Platt J. Scholkopf B. Support vector machines. *IEEE Intelligent Systems and their Applications*, 1998, vol. 13, no. 4, pp. 18–28. <https://doi.org/10.1109/5254.708428>
5. Johnson R., Zhang T. Convolutional neural networks for text categorization: Shallow word-level vs. deep character-level. *ArXiv*, 2016, arXiv:1609.00718. <https://doi.org/10.48550/arXiv.1609.00718>
6. Devlin J., Chang M.-W., Lee K., Toutanova K. BERT: Pre-training of deep bidirectional transformers for language understanding. *Proc. of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. Vol. 1 (Long and Short Papers), 2019, pp. 4171–4186. <https://doi.org/10.18653/v1/N19-1423>
7. Radford A., Narasimhan K., Salimans T., Sutskever I. *Improving language understanding by generative pre-training*. Preprint. 2018.
8. Karpukhin V., Oğuz B., Min S., Lewis P., Wu L., Edunov S., Chen D., Yih W.-T. Dense passage retrieval for open-domain question answering. *Proc. of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2020, pp. 6769–6781. <https://doi.org/10.18653/v1/2020.emnlp-main.550>
9. Humeau S., Shuster K., Lachaux M., Weston J. Poly-encoders: Architectures and pre-training strategies for fast and accurate multi-sentence scoring. *arXiv*, 2020, arXiv:1905.01969. <https://doi.org/10.48550/arXiv.1905.01969>
10. Lewis M., Liu Y., Goyal N., Ghazvininejad M., Mohamed A., Levy O., Stoyanov V., Zettlemoyer L. BART: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension. *Proc. of the 58th Annual Meeting of the Association for Computational Linguistics*, 2020, pp. 7871–7880. <https://doi.org/10.18653/v1/2020.acl-main.703>
11. Kim S., Liu Y., Jin D., Papangelis A., Hedayatnia B., Gopalakrishnan K., Hakkani-Tur D. *DSTC10 Track Proposal*:

- Knowledge-grounded Task-oriented Dialogue Modeling on Spoken Conversations. 2021.
12. Kim S., Eric M., Gopalakrishnan K., Hedayatnia B., Liu Y. Hakkani-Tur D.Z. Beyond domain APIs: task-oriented conversational modeling with unstructured knowledge access // *Proc. of the 21st Annual Meeting of the Special Interest Group on Discourse and Dialogue*. 2020. P. 278–289.
 13. Budzianowski P., Wen T.-H., Tseng B.-H., Casanueva I., Ultes S., Ramadan O., Gašić M. MultiWOZ - A large-scale multi-domain wizard-of-oz dataset for task-oriented dialogue modelling // *Proc. of the 2018 Conference on Empirical Methods in Natural Language Processing*. 2018. P. 5016–5026. <https://doi.org/10.18653/v1/D18-1547>
 14. Eric M., Goel R., Paul S., Sethi A., Agarwal S., Gao S., Kumar A., Goyal A., Ku P., Hakkani-Tür D. Multiwoz 2.1: Multi-domain dialogue state corrections and state tracking baselines // *Proc. of the Twelfth Language Resources and Evaluation Conference*. 2020. P. 422–428.
 15. Zang X., Rastogi A., Sunkara S., Gupta R., Zhang J., Chen J. MultiWOZ 2.2: A dialogue dataset with additional annotation corrections and state tracking baselines // *Proc. of the 2nd Workshop on Natural Language Processing for Conversational AI*. 2020. P. 109–117. <https://doi.org/10.18653/v1/2020.nlp4convai-1.13>
 16. Baevski A., Zhou H., Mohamed A., Auli M. Wav2vec 2.0: a framework for self-supervised learning of speech representations // *Proc. of the 34th International Conference on Neural Information Processing Systems (NIPS'20)*. 2020. P. 12449–12460.
 17. Panayotov V., Chen G., Povey D., Khudanpur S., Librispeech: An ASR corpus based on public domain audio books // *Proc. of the 2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. 2015. P. 5206–5210. <https://doi.org/10.1109/ICASSP.2015.7178964>
 18. Heafield K. KenLM: Faster and Smaller language model queries // *Proc. of the Sixth Workshop on Statistical Machine Translation*. 2011. P. 187–197.
 19. Gopalakrishnan K., Hedayatnia B., Wang L., Liu Y., Hakkani-Tür D. Are neural open-domain dialog systems robust to speech recognition errors in the dialog history? an empirical study // *Proc. Interspeech 2020*. 2020. P. 911–915. <https://doi.org/10.21437/Interspeech.2020-1508>
 20. Wang L., Fazel-Zarandi M., Tiwari A., Matsoukas S., Polymenakos L. Data Augmentation for Training Dialog Models Robust to Speech Recognition Errors // *Proc. of the 2nd Workshop on Natural Language Processing for Conversational AI*. 2020. P. 63–70. <https://doi.org/10.18653/v1/2020.nlp4convai-1.8>
 21. Xu L., Lian J., Zhao W.X., Gong M., Shou L., Jiang D., Xie X., Wen J. Negative sampling for contrastive representation learning: A review // *ArXiv*. 2022. arXiv:2206.00212. <https://doi.org/10.48550/arXiv.2206.00212>

Автор

Маслюхин Сергей Михайлович — ведущий научный сотрудник, ООО «ЦРТ-инновации», Санкт-Петербург, 194044, Российская Федерация; инженер, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, <https://orcid.org/0000-0002-9054-5252>, maslyukhin@speechpro.com

Статья поступила в редакцию 05.10.2022
Одобрена после рецензирования 01.12.2022
Принята к печати 15.01.2023

Author

Sergei M. Masliukhin — Leading Researcher, STC-innovations Limited, Saint Petersburg, 194044, Russian Federation; Engineer, ITMO University, Saint Petersburg, 197101, Russian Federation, <https://orcid.org/0000-0002-9054-5252>, maslyukhin@speechpro.com

Received 05.10.2022
Approved after reviewing 01.12.2022
Accepted 15.01.2023



Работа доступна по лицензии
Creative Commons
«Attribution-NonCommercial»

ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ И КОГНИТИВНЫЕ ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ ARTIFICIAL INTELLIGENCE AND COGNITIVE INFORMATION TECHNOLOGIES

Аннотация раздела

Искусственный интеллект сегодня — одна из наиболее развивающихся областей знаний, направленная на создание алгоритмов воспроизведения когнитивных и творческих способностей человека в условиях неопределенности и неполноты данных. Данный раздел нашего журнала посвящен как методическим основам искусственного интеллекта, так и специфике его применения в составе когнитивных информационных технологий.

Искусственный интеллект на основе машинного обучения имитирует сам процесс работы человека с данными в условиях неопределенности. Как следствие, в разделе представлены работы, раскрывающие аспекты таких задач, как восстановление математических описаний неизвестных физических законов по данным, тематическое моделирование на основе неразмеченных и частично размеченных текстовых данных, а также определение предсказуемости временных рядов как основы для прогнозирования по данным. Отдельно рассмотрены приложения технологий искусственного интеллекта и машинного обучения для создания прикладных социотехнических систем в части выявления механизмов влияния стресса на принятие решений и предсказания покупательской активности.

Результаты, представленные в работах, будут полезны для создания прикладных интеллектуальных решений, как автономных, так и взаимодействующих с человеком. Они могут быть интересны специалистам в области искусственного интеллекта, математического моделирования и машинного обучения, а также отраслевым потребителям в различных областях знаний.

Abstract

Artificial Intelligence today is one of the most dynamic areas of expertise aimed at designing algorithms for recreating human cognitive and creative abilities in conditions of uncertainty and incomplete data. This section of the journal will be devoted to both the methodological foundations of Artificial Intelligence and its application specifics as a part of the area of cognitive information technologies.

Artificial Intelligence based on machine learning mimics the very process of human work with data under uncertainty. As a result, this particular volume covers aspects of such problems as discovery of physical laws from the observation data, topic modeling from unlabeled and partially labeled textual data, and determining the predictability of time series as a basis for forecasting from data. Separately, the applications of Artificial Intelligence and machine learning technologies for creating applied sociotechnical systems are considered in terms of identifying the mechanisms of the influence of stress on decision-making and predicting purchasing activity.

The results presented in this section will be useful for the creation of applied intelligent solutions, both autonomous and interacting with a human. They may be also attractive to both specialists in the field of Artificial Intelligence, mathematical modeling and machine learning, and major consumers in various fields of knowledge.

doi: 10.17586/2226-1494-2023-23-1-97-104

Multiobjective evolutionary discovery of equation-based analytical models for dynamical systems

Mikhail A. Maslyaev¹, Alexander A. Hvatov²✉

^{1,2} ITMO University, Saint Petersburg, 197101, Russian Federation

¹ mikemaslyaev@itmo.ru, <https://orcid.org/0000-0001-5595-0802>

² alex_hvatov@itmo.ru✉, <https://orcid.org/0000-0002-5222-583X>

Abstract

In this article, an approach to modeling dynamical systems in case of unknown governing physical laws has been introduced. The systems of differential equations obtained by means of a data-driven algorithm are taken as the desired models. In this case, the problem of predicting the state of the process is solved by integrating the resulting differential equations. In contrast to classical data-driven approaches to dynamical systems representation, based on the general machine learning methods, the proposed approach is based on the principles, comparable to the analytical equation-based modeling. Models in forms of systems of differential equations, composed as combinations of elementary functions and operation with the structure, were determined by adapted multi-objective evolutionary optimization algorithm. Time-series describing the state of each element of the dynamic system are used as input data for the algorithm. To ensure the correct operation of the algorithm on data characterizing real-world processes, noise reduction mechanisms are introduced in the algorithm. The use of multicriteria optimization, held in the space of complexity and quality criteria for individual equations of the differential equation system, makes it possible to improve the diversity of proposed candidate solutions and, therefore, to improve the convergence of the algorithm to a model that best represents the dynamics of the process. The output of the algorithm is a set of Pareto-optimal solutions of the optimization problem where each individual of the set corresponds to one system of differential equations. In the course of the work, a library of data-driven modeling of dynamic systems based on differential equation systems was created. The behavior of the algorithm was studied on a synthetic validation dataset describing the state of the hunter-prey dynamic system given by the Lotka-Volterra equations. Finally, a toolset based on the solution of the generated equations was integrated into the algorithm for predicting future system states. The method is applicable to data-driven modeling of arbitrary dynamical systems (e.g. hydrometeorological systems) in cases where the processes can be described using differential equations. Models generated by the algorithm can be used as components of more complex composite models, or in an ensemble of methods as an interpretable component.

Keywords

differential equation discovery, evolutionary optimization, multi-objective optimization, differential equations system, symbolic regression

Acknowledgements

This research is financially supported by the Russian Scientific Foundation, Agreement No. 21-71-00128.

For citation: Maslyaev M.A., Hvatov A.A. Multiobjective evolutionary discovery of equation-based analytical models for dynamical systems. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 2023, vol. 23, no. 1, pp. 97–104. doi: 10.17586/2226-1494-2023-23-1-97-104

УДК 004.89

Определение аналитических моделей динамических систем в форме дифференциальных уравнений на основе многокритериальной эволюционной оптимизации

Михаил Александрович Масляев¹, Александр Александрович Хватов²✉

^{1,2} Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация

¹ mikemaslyaev@itmo.ru, <https://orcid.org/0000-0001-5595-0802>

² alex_hvatov@itmo.ru ✉, <https://orcid.org/0000-0002-5222-583X>

Аннотация

Предмет исследования. В работе предложен метод моделирования динамических систем при условии, что управляющие процессом физические законы неизвестны. В качестве искоемых моделей приняты полученные при помощи управляемого данными алгоритма системы дифференциальных уравнений. В результате решается задача прогнозирования состояния процесса при помощи интегрирования результирующих дифференциальных уравнений. В отличие от классических подходов к воспроизведению динамических систем на основе данных, основанных на общих принципах машинного обучения, предложенный алгоритм позволяет сформировать модели процессов, сопоставимые с аналитическими. **Метод.** В качестве модели процесса приняты системы дифференциальных уравнений, представленные через комбинации элементарных функций и операторов, определенные при помощи адаптированного эволюционного алгоритма многокритериальной оптимизации. В качестве входных данных для алгоритма использованы временные ряды, описывающие состояние каждого элемента динамической системы. Для обеспечения работы алгоритма на данных, характеризующих реальные процессы, в алгоритм включены механизмы компенсации шума. Использование многокритериальной оптимизации, проводимой в пространстве критериев сложности и качества отдельных уравнений системы дифференциальных уравнений, позволило улучшить разнообразие предлагаемых кандидатных решений. Также получена высокая сходимость алгоритма, что обеспечило поиск модели, наилучшим образом показывающей динамику процесса. Результатом работы алгоритма является множество Парето-оптимальных решений оптимизационной задачи, каждое из которых соответствует одной системе дифференциальных уравнений.

Основные результаты. В ходе работы создана библиотека управляемого данными моделирования динамических систем на основе систем дифференциальных уравнений. Поведение алгоритма исследовано на синтетическом валидационном наборе данных, описывающем состояние динамической системы «охотник-жертва», заданной уравнениями Лотки–Вольтерра. Предложен интегрированный в алгоритм механизм прогнозирования состояний системы, основанный на решении сформированных уравнений. **Практическая значимость.** Метод применим к управляемому данными моделированию произвольных динамических систем (например, гидрометеорологических) в случаях, когда процессы могут быть описаны при помощи дифференциальных уравнений. Сформированные алгоритмом модели можно использовать в качестве компонент более сложных композитных моделей, или в ансамбле методов как интерпретируемую составляющую.

Ключевые слова

определение дифференциальных уравнений, эволюционная оптимизация, многокритериальная оптимизация, динамические системы, символьная регрессия

Благодарности

Исследование поддержано Российским научным фондом, грант № 21-71-00128.

Ссылка для цитирования: Масляев М.А., Хватов А.А. Определение аналитических моделей динамических систем в форме дифференциальных уравнений на основе многокритериальной эволюционной оптимизации // Научно-технический вестник информационных технологий, механики и оптики. 2023. Т. 23, № 1. С. 97–104 (на англ. яз.). doi: 10.17586/2226-1494-2023-23-1-97-104

Introduction

Systems of Ordinary or Partial Differential Equations (ODEs and PDEs) are powerful tools that can describe complex dynamics of structures involving multiple variables. While many tools can be used for creating mathematical models for processes, such as classical machine learning models, or unconventional ones, like Bayesian networks [1], they tend to have strict limitations to their applications. In cases of many real-world systems, in addition to the issues above, these models are often abstracted from the intrinsic physical principles guiding the system. The classical approach to deriving systems of differential equations necessitates the use of mathematical analysis in combination with an in-depth understanding of the process. The data-driven approach to system discovery involves the creation of an individual differential equation

for each dependent variable that can be measured from a system.

The forms of the discovered models containing systems of differential equations are selected due to the prevalence of differential equation in physical systems. For example, flow of viscous fluid is governed by Navier-Stokes equations that are a system of partial differential equations. Dynamics and interactions between electric and magnetic components of the electromagnetic field are described with Maxwell's equations which are a system of PDE as well. Many simpler systems, such as rotation of the spherical pendulum, can be defined with system of ordinary differential equations.

Apart from the descriptive possibilities, provided by models in forms of systems of differential equations, obtained systems can be solved to predict further states of the process. While the toolkit for the automatic solution of

systems of ordinary/partial differential equations is out of scope of this study, several studies have been conducted towards implementing equation-solving module into the frameworks of differential equations discovery, as in work [2]. With this ability to solve model equations, the system dynamics can be propagated into the future.

Analysis of existing approaches

Creating models for dynamical systems governed by differential equations has recently drawn interest. The first perspective to the task involves developing substitutes for the equations in forms of propagation operators that map the state of the system forward in time like in [3] or [4]. Dynamic Mode Decomposition (DMD) involves approximation of the system dynamics with a finite-dimensional linear operator. While that can be useful for multiple real-world applications where the propagator is linear, many other cases involve non-linear dynamics that cannot be fully explained with DMD approach.

A number of data-driven solutions to the problem of explaining dynamical system with explicitly derived governing equations have been developed. Here, we will inspect methods that are applicable not only for problems of discovering Ordinary Differential Equations (ODE) and systems of ODEs, but also for tasks of partial differential equations discovery. The first problem has sufficient solution in forms of Multilayer Stochastic Models (MSMs) proposed by Kondrashov, Chekroun and Ghil in [5]. However, due to the non-Markovian approach, the approach is not extendable to the problems of partial differential equations.

The earliest advances were made with the symbolic regression [6]. Governing equations are viewed as computational tree graphs where leaves are inputs, and on the other levels various operators are located. The search of the equation can be done with the typical graph-targeted evolutionary optimization algorithms. More contemporary approaches are represented by sparse regression based models developed in many works, including Kaheman et al. in [7] and Berg & Nystrom in [8], and with artificial neural networks (ANN) representation of the dynamical system. While there are multiple approaches to discover differential equations with artificial neural networks, notable ones include PINN [9], PDE-Net, developed by Long et al. [10], and physics-informed neural networks by Raissi et al. [11].

Partial differential equation search with sparse regression uses LASSO operator that is applied to approximate time derivative with a library of candidate terms. That library has to contain all possible equation terms, and the usage of sparsity operator allows selection of only a few active feature terms. The main issues of this approach can be linked with its rigidity: the term library has to be extensive enough to contain all possible terms including all non-linear functions that can be present in equations. While many of the presented approaches can be applied to the systems of differential equations, their possibilities are limited by description of time dynamics of a vector variable, like in paper [12].

The algorithm described in this article is based on the multiobjective evolutionary optimization approach,

where the model obtained is evaluated by several metrics describing quality and complexity of the equations of the system. Thus, the algorithm can provide the parsimonious model that is not overly complex but can sufficiently simulate the dynamics of the process. However, the problem of selecting that parsimonious model from the discovered Pareto frontier is the problem for another study. This paper is dedicated to the problem of discovering the optimal set of candidate equations for the further expert conclusions and applications.

Equation discovery problem

To describe some unstudied process, which involves multiple (n) dependent variables, we desire to derive a system of differential equations. Let us denote these variables in general problem statement as $\mathbf{u} = (u_1(t, \mathbf{x}), u_2(t, \mathbf{x}), \dots, u_n(t, \mathbf{x}))$. They are defined in the spatial domain Ω , represented by coordinates \mathbf{x} , and dependent from time t . In case of a system of ordinary differential equations, the variables can be assumed to be only time-dependent (i.e., $u_1(t), u_2(t), \dots, u_n(t)$).

For the equations search process, the algorithm requires sets of observations arranged on a rectangular grid. For the equation search process, the algorithm demands arrays of calculated derivatives. While in some cases these derivatives can be obtained directly, using measurement techniques, in others they necessitate a preprocessing phase where the derivatives are calculated numerically from the input data variables. While the numerical techniques of derivatives estimation are numerous [13], the most efficient approaches are finite-difference differentiation and analytical differentiation of variable-approximating polynomials. In many cases, additional smoothing is required to reduce magnitudes of noise in the data. Here, the algorithm employs Gaussian smoothing in the spatial domain, or replacement of the initial data fields with their artificial neural network approximation.

$$S(\mathbf{u}) = \begin{cases} L_1(\mathbf{u}) = 0 \\ \dots \\ L_k(\mathbf{u}) = 0 \end{cases},$$

where $S(\mathbf{u})$ is the system of differential equations that involves variables \mathbf{u} comprised of individual equations $L_1(\mathbf{u})$ to $L_k(\mathbf{u})$.

The search for the optimal structures of equations in the system is done with the multi-objective optimization implemented with the Many-Objective Optimization Evolutionary Algorithm Based on Dominance and Decomposition (MOEA/DD), introduced in [14].

The search is performed in the criteria space of complexities $C(L_j'\mathbf{u})$ and modeling errors $Q(L_j'\mathbf{u})$ for each individual equation in a system. Therefore, the problem can be reformulated to

$$\begin{aligned} \text{minimize } \mathbf{F}(S(\mathbf{u})) &= f_1(S(\mathbf{u})), \dots, f_m(S(\mathbf{u})) = \\ &= (C(L_1'\mathbf{u}), Q(L_1'\mathbf{u}), \dots, C(L_n'\mathbf{u}), Q(L_n'\mathbf{u})). \end{aligned}$$

Here the constraints are introduced in the equations construction logic rather than explicitly specified during the optimization problem statement.

The complexity metric $C(L'_j \mathbf{u})$ is defined as a number of “active” tokens in the equation, i.e. ones presented in terms of non-zero coefficients.

The problem of selecting the most appropriate metric for evaluating the properties of process representation for the equation has been studied in work [2]. The best metrics for modeling quality are L_2 norm of matrices of differential operator residuals represented by:

$$Q(L'_j \mathbf{u}) = \|L'_j \mathbf{u}\|_2$$

or the norm of matrices of differences between the input variable fields u_j and the solutions \tilde{u}_j of corresponding equations

$$Q(L'_j \mathbf{u}) = \|u_j - \tilde{u}_j\|_2.$$

Due to the necessity to conduct optimization, having a limited number of candidate solutions, the implemented approach uses concept of domination for the proposed solutions to the problem of searching for systems of equations. It is said that candidate system $S_1(\mathbf{u})$ dominates candidate system $S_2(\mathbf{u})$ if for all optimized criteria f_i : $f_i(S_1(\mathbf{u})) \leq f_i(S_2(\mathbf{u}))$ and for a single criterion f_i : $f_i(S_1(\mathbf{u})) < f_i(S_2(\mathbf{u}))$. A solution is called Pareto-optimal if no other solutions dominate it. The objective of the implemented algorithm is to obtain a set of candidate solutions where each solution is Pareto-optimal. In addition to the Pareto-optimal set, other non-dominated sets can be introduced by induction: n -th non-dominated level is comprised by solution that is not dominated by any solutions, except the ones on the $n-1$ th, or lower levels.

Approach description

In this section, we briefly describe the main diversions of our approach from the original algorithm [14] and case-specific solutions employed during the system of differential equations derivation, such as evolutionary operators. Following the optimization objectives stated in the previous section, the algorithm performs a simultaneous search of system equations and parameters which define the equations structures. The structure of an equation can be decomposed into a set of equation terms and a set of their real-valued coefficients a_i as in:

$$L'_j \mathbf{u} = \sum_i a_i \prod_j t_{ij}.$$

The terms of constructed equations are represented with a tokens product $\prod_j t_{ij}$, $t_{ij} \in \mathbf{T}$, elementary building blocks containing arbitrary user-defined functions. This approach enables the discovery of non-linear equations with compound structures that can be represented as a sum of product terms. During search of differential equations, various derivatives (e.g. $\frac{\partial^n u_i}{\partial x_j^n}$) are included into the pool \mathbf{T} .

Other case-specific functions or external variables can be included as tokens into the token pool to be available for the algorithm during equation search. For example, suppose a study objective is to discover the equation for the temperature dynamics in a medium. In that case, the velocity field of the medium can be considered an external variable.

To create a system that can model the studied process, it is possible to assume that each equation in the system must represent the spatial or temporal dynamics of at least one variable. By describing a variable dynamic, we understand that the equation contains corresponding derivatives of the variable. During the evolutionary search, evolutionary operators affecting the structures of the equations have to preserve the descriptive properties of such terms.

Evolutionary algorithm details

To start the evolutionary optimization, the algorithm has to construct the initial population $\mathbf{P} = S_1(\mathbf{u}), \dots, S_2(\mathbf{u})$ of randomly generated candidate systems of differential equations. As mentioned above, a system equation has to represent a corresponding variable's dynamic. Therefore, during the initialization, a variable is assigned to each equation as its “main” one. Without loss of generality, we can assume that the i -th equation describes i -th variable.

To emphasize the duality of the system discovery, an individual encoding must represent both equations and meta-parameters of the equations. The chromosome of an individual contains computational graphs of the equations as “equation genes” and values of the parameters that define the creation of the equation. Equation graphs take the form of tree graphs, where the leaves are elementary functions stored in tokens, and intermediate nodes are product operators that form equation terms from factor tokens. The graph root comprises the summation operator which combines separate terms into the equation. The scheme of the equation system encoding is presented in Fig. 1.

A regularization tool has to be created to regulate the complexity of the equations proposed by the algorithm. Its main objective is to exclude terms with low significance and explanatory power in the resulting model. Selection of the terms can be made with sparse regression, operating with the LASSO operator:

$$\|\mathbf{F}_k \beta - F_{\text{target},k}\|_2^2 + \lambda \|\beta\|_1 - \min_{\beta}.$$

As the predicted value of the operator, a random equation term representing an “equation variable”, i.e., containing its derivative, is selected. LASSO operator can obtain a vector of term weights β with values of the terms in the left-hand side of the equation, evaluated on the space-time grid, normalized and combined into matrix \mathbf{F}_k , with vector of right-hand part values of $F_{\text{target},k}$. In the operator statement $\|\cdot\|_p$, the i -th norm of the matrix is designated.

The sparsity constant parameter λ determines the penalty of optimized functional with respect to the values of weights in β , prioritizing setting zero coefficients to the less significant predictors. The algorithm can control the equation complexity by regulating the value of the sparsity constant. Higher values of λ promote equations with fewer numbers of terms, while lower values tend to lead to more complex equations. Due to the significance of the sparsity parameter for the equation definition, it is included for each equation in the system in the encoding of the individual.

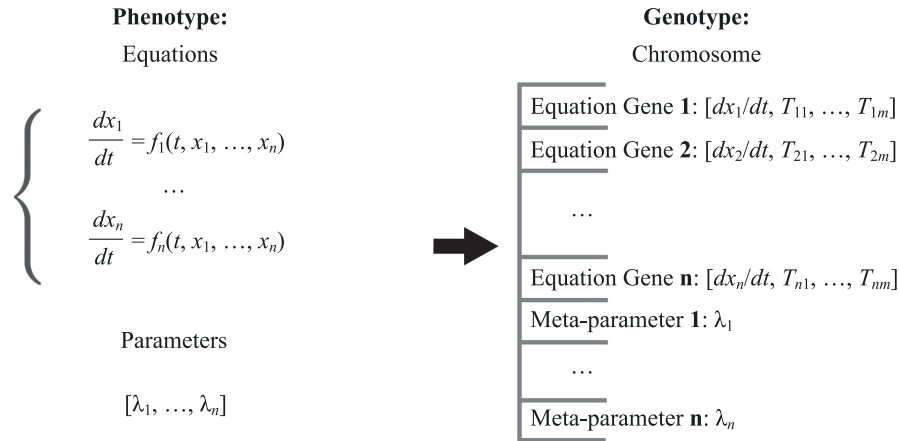


Fig. 1. Scheme of the encoding for a system of ODEs for arbitrary variables x_i with sparsity constants λ_i as meta-parameters. T_{nm} denotes the m -th term of the n -th equation of the system, while f_n is an arbitrary right-hand side function of the n -th equation of the system

The coefficients of the equation are computed with linear regression where active terms from the left-hand part are combined into a matrix of predictors, and values of the term on the right part are used as a predicted value.

Evolutionary operators

The general idea of evolutionary operators affecting the population to obtain the set of optimal systems of equations is borrowed from the single-objective algorithm of equation discovery proposed in [15]. The alterations of an individual equation can be done with operators of mutations and crossover. The operators are applied to individuals of the population following the guidelines presented in the paper and describing the base algorithm [16].

The process of the evolution is held iteratively, for a specified number of iterations and over sectors, defined by the weight vectors introduced into the space of optimization criteria to decompose the problem into smaller sub-problems. As in the original version, the algorithm constructs a set of weight vectors $\mathbf{W} = \mathbf{w}_1, \dots, \mathbf{w}_N$ from a unit simplex, one for each candidate solution in \mathbf{P} .

After the weights are defined, each individual of the population \mathbf{P} is assigned to a random sector of the criteria space. That enables a more even coverage of the search space due to the property that the individuals converge in the directions of weights.

The selection of the individuals for the crossover operators is held in a manner that respects problem decomposition. In the base scenario, the parents are selected from the neighboring sectors to the one associated with the processed weight vector. However, to increase the algorithm exploratory properties, which are vital in the problem of equation construction, with a relatively small probability, the parents are selected from other, non-adjacent sectors. The selected candidates are added to the parent pool, and the crossover is held among them.

The crossover operator affects both systems of equations and corresponding vectors of meta-parameters.

The interactions between equations of the systems comply with variable description requirements. For each modeled variable, the corresponding equations of the parent

systems are affected by crossover. Two main types of operators are used here: term-wise exchange and complete equation swapping.

The first type of equation-level crossover operator involves an exchange of terms between parent equations. All initial terms of the equation are divided into three groups. First group includes terms present in the same form in both parents. Second group includes terms present in both parents, but in this case the parameters of their tokens are different. Third group contains unique ones between parents terms. The first group is not affected by crossover at all. The crossover between parents in the second group is parametric-only: the same tokens exchange the parameter values from a specified proportion.

After the creation of offspring individuals, they are affected by mutation operators. Their purpose is two-fold. They are increasing the exploratory properties of the algorithm and preventing the generation of repeating individuals which is mandatory for the implemented multi-objective optimization approach. The main idea of the mutation operator is the random change of a term into a new, unique one. The first type of operator changes a factor representing a token into a new, randomly generated one; or changes token parameters (e.g. frequency of a sine token) with an increment taken from normal distribution $N(0, \sigma)$ with pre-defined variance σ^2 . The second type involves a replacement of a term with a newly generated one. When the offspring creation procedures are conducted, the Pareto levels are updated with respect to the newly created solution. The population update algorithm considers the decomposition of the problems with a set of weight vectors and their domination. The approach in which the evolutionary operators are applied during the search is presented in Fig. 2.

Validation

Several validation experiments have been conducted to assess the proposed approach performance in discovering systems of equations that govern the dynamical system. The most demonstrative approach to check the behavior of our algorithm employs synthetic data obtained from the solution of known equations.

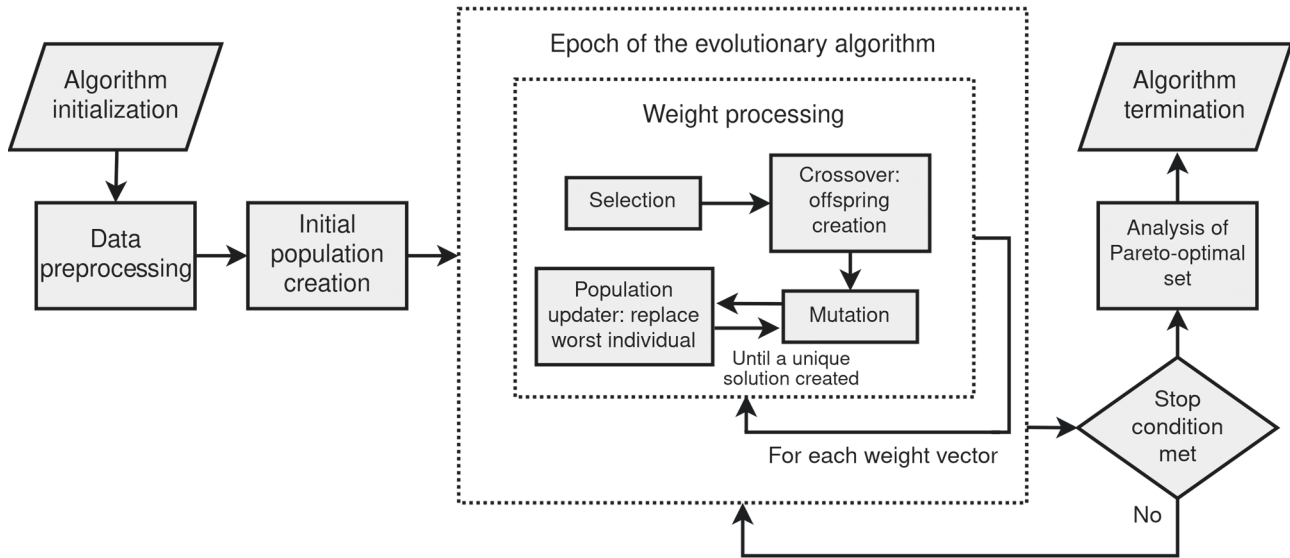


Fig. 2. Generalized scheme of the main search sequence of the algorithm

A hunter-prey model described by Lotka-Volterra equations:

$$S(\mathbf{u}) = \begin{cases} \dot{u} = \alpha u - \beta uv \\ \dot{v} = \delta uv - \gamma u \end{cases}$$

was selected as the dynamical system to be described. The model represents simplified dynamics of two species: $u = u(t)$ depicts “prey”, while $v = v(t)$ represents the “hunter” species. Usual time-derivative convention $\dot{u} = \frac{du}{dt}$ is implemented. Constants α , β , δ , and γ determine the dynamics of the system.

The solutions of the equations were numerically obtained using Runge-Kutta methods. The solutions for $u(t)$ and $v(t)$ are demonstrated in Fig. 3.

The Pareto-optimal set of equations obtained from the algorithm typically has forms similar to the one presented in Fig. 4. Here, the algorithm output is reformulated with

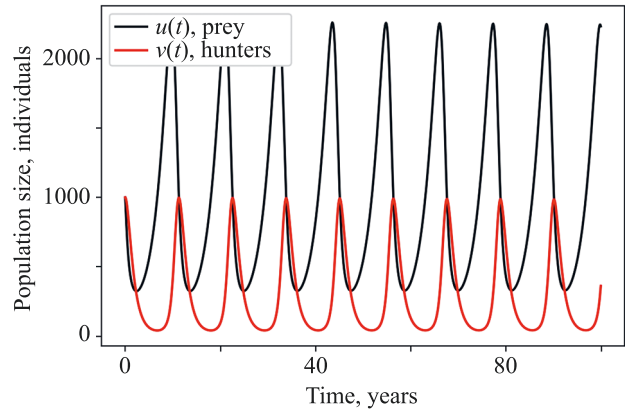


Fig. 3. Visualization of the solution of Lotka-Volterra equations

the combination of optimized metrics: instead of evaluating complexity or approximation errors of individual equations

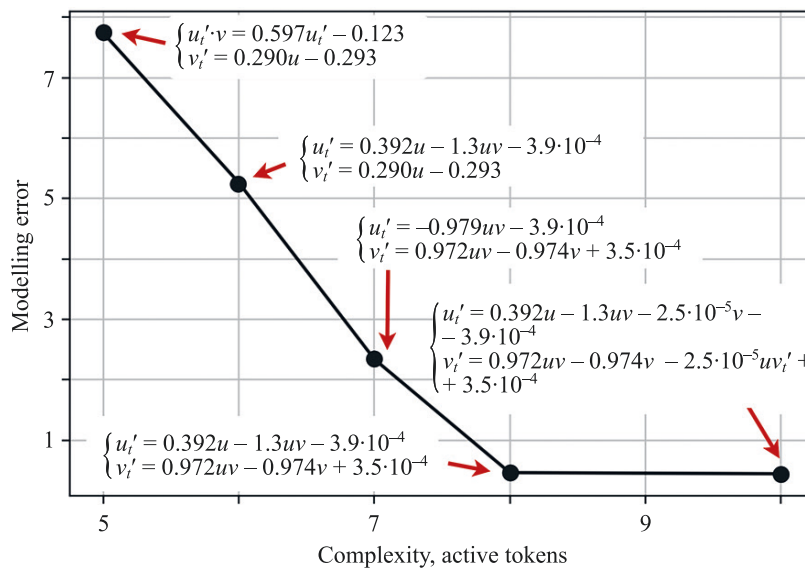


Fig. 4. Pareto frontier of systems of equations obtained by the algorithm

they are viewed for the system integrally. The allowable interval for complexity controlling parameters λ used in sparsity operators is between 10^{-8} and 10^{-2} . Next, an additional family of trigonometric tokens was introduced into the pool to create a diversity of created terms.

Ten independent runs with ten multi-objective optimization evolutionary algorithm iterations and ten more with 25 iterations were performed, and the obtained Pareto-optimal sets were compared. Due to the relatively simple structure of the initial system of equations, a successful convergence to the similar (in terms of obtaining sets with similar equations) was achieved in every case. While this test cannot be considered a comprehensive study of the algorithm properties, it can be viewed as proof that the algorithm can operate and discover the equations.

Conclusion

In this article, we proposed a robust extension of the single differential equation discovery approach to the problems of creating models for systems of differential equations. The multi-objective approach enables the

creation of a diverse set of models. With the analysis of complexity-quality tradeoff, an expert should be able to select the parsimonious model for the process description. The approach has high levels of versatility that are untypical and novel among equation discovery algorithms. It can obtain both ordinary and partial differential equations with arbitrary structures.

The main drawback of the developed approach has high computational cost which can be especially noticeable in multidimensional data (i.e., systems of partial differential equations) or data with high noise levels where high numbers of iterations are required for the algorithm convergence. Therefore, improving the algorithm computational performance can be the priority for further development. Also, developing sufficient tools for using the derived equations for the process state prediction is another goal of the next research.

The numerical solution data and the Python code that partially reproduce the experiments are available at the GitHub repository of the ITMO University¹.

¹ Available at: <https://github.com/ITMO-NSS-team/EPDE> (accessed: 07.12.2022).

References

1. Bubnova A.V., Deeva I., Kalyuzhnaya A.V. MiXBN: library for learning Bayesian networks from mixed data. *Procedia Computer Science*, 2021, vol. 193, pp. 494–503. <https://doi.org/10.1016/j.procs.2021.10.051>
2. Maslyaev M., Hvatov A. Solver-based fitness function for the data-driven evolutionary discovery of partial differential equations. *Proc. of the 2022 IEEE Congress on Evolutionary Computation (CEC)*, 2022. <https://doi.org/10.1109/cec55065.2022.9870370>
3. Brunton S.L., Brunton B.W., Proctor J.L., Kaiser E., Kutz J.N. Chaos as an intermittently forced linear system. *Nature Communications*, 2017, vol. 8, no. 1, pp. 19. <https://doi.org/10.1038/s41467-017-00030-8>
4. Schmid P.J., Sesterhenn J. Dynamic mode decomposition of numerical and experimental data. *Proc. of the 61st Annual Meeting of the APS Division of Fluid Dynamics*. American Physical Society, November 2008.
5. Kondrashov D., Chekroun M.D., Ghil M. Data-driven non-Markovian closure models. *Physica D: Nonlinear Phenomena*, 2015, vol. 297, pp. 33–55. <https://doi.org/10.1016/j.physd.2014.12.005>
6. Schmidt M., Lipson H. Distilling free-form natural laws from experimental data. *Science*, 2009, vol. 324, no. 5923, pp. 81–85. <https://doi.org/10.1126/science.1165893>
7. Kaheman K., Kutz J.N., Brunton S.L. SINDy-PI: a robust algorithm for parallel implicit sparse identification of nonlinear dynamics. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 2020, vol. 476, no. 2242, pp. 20200279. <https://doi.org/10.1098/rspa.2020.0279>
8. Berg J., Nyström K. Data-driven discovery of PDEs in complex datasets. *Journal of Computational Physics*, 2019, vol. 384, pp. 239–252. <https://doi.org/10.1016/j.jcp.2019.01.036>
9. Han G., Zahr M.J., Wang J.-X. Physics-informed graph neural Galerkin networks: A unified framework for solving PDE-governed forward and inverse problems. *Computer Methods in Applied Mechanics and Engineering*, 2022, vol. 390, pp. 114502. <https://doi.org/10.1016/j.cma.2021.114502>
10. Long Z., Lu Y., Ma X., Dong B. PDE-Net: learning PDEs from data. *Proceedings of Machine Learning Research*, 2018, vol. 80, pp. 3208–3216.
11. Raissi M. Deep hidden physics models: Deep learning of nonlinear partial differential equations. *Journal of Machine Learning Research*, 2018, vol. 19, pp. 1–24.
12. Zhang J., Ma W. Data-driven discovery of governing equations for fluid dynamics based on molecular simulation. *Journal of Fluid*

Литература

1. Bubnova A.V., Deeva I., Kalyuzhnaya A.V. MiXBN: library for learning Bayesian networks from mixed data // *Procedia Computer Science*. 2021. V. 193. P. 494–503. <https://doi.org/10.1016/j.procs.2021.10.051>
2. Maslyaev M., Hvatov A. Solver-based fitness function for the data-driven evolutionary discovery of partial differential equations // *Proc. of the 2022 IEEE Congress on Evolutionary Computation (CEC)*. 2022. <https://doi.org/10.1109/cec55065.2022.9870370>
3. Brunton S.L., Brunton B.W., Proctor J.L., Kaiser E., Kutz J.N. Chaos as an intermittently forced linear system // *Nature Communications*. 2017. V. 8. N 1. P. 19. <https://doi.org/10.1038/s41467-017-00030-8>
4. Schmid P.J., Sesterhenn J. Dynamic mode decomposition of numerical and experimental data // *Proc. of the 61st Annual Meeting of the APS Division of Fluid Dynamics*. American Physical Society, November 2008.
5. Kondrashov D., Chekroun M.D., Ghil M. Data-driven non-Markovian closure models // *Physica D: Nonlinear Phenomena*. 2015. V. 297. P. 33–55. <https://doi.org/10.1016/j.physd.2014.12.005>
6. Schmidt M., Lipson H. Distilling free-form natural laws from experimental data // *Science*. 2009. V. 324. N 5923. P. 81–85. <https://doi.org/10.1126/science.1165893>
7. Kaheman K., Kutz J.N., Brunton S.L. SINDy-PI: a robust algorithm for parallel implicit sparse identification of nonlinear dynamics // *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*. 2020. V. 476. N 2242. P. 20200279. <https://doi.org/10.1098/rspa.2020.0279>
8. Berg J., Nyström K. Data-driven discovery of PDEs in complex datasets // *Journal of Computational Physics*. 2019. V. 384. P. 239–252. <https://doi.org/10.1016/j.jcp.2019.01.036>
9. Han G., Zahr M.J., Wang J.-X. Physics-informed graph neural Galerkin networks: A unified framework for solving PDE-governed forward and inverse problems // *Computer Methods in Applied Mechanics and Engineering*. 2022. V. 390. P. 114502. <https://doi.org/10.1016/j.cma.2021.114502>
10. Long Z., Lu Y., Ma X., Dong B. PDE-Net: learning PDEs from data // *Proceedings of Machine Learning Research*. 2018. V. 80. P. 3208–3216.
11. Raissi M. Deep hidden physics models: Deep learning of nonlinear partial differential equations // *Journal of Machine Learning Research*. 2018. V. 19. P. 1–24.
12. Zhang J., Ma W. Data-driven discovery of governing equations for fluid dynamics based on molecular simulation // *Journal of Fluid Mechanics*. 2020. V. 892. P. A5. <https://doi.org/10.1017/jfm.2020.184>

- Mechanics*, 2020, vol. 892, pp. A5. <https://doi.org/10.1017/jfm.2020.184>
13. Van Breugel F., Kutz J.N., Brunton B.W. Numerical differentiation of noisy data: A unifying multi-objective optimization framework // *IEEE Access*, 2020, vol. 8, pp. 196865–196877. <https://doi.org/10.1109/access.2020.3034077>
 14. Maslyaev M., Hvatov A., Kalyuzhnaya A. Partial differential equations discovery with EPDE framework: Application for real and synthetic data. *Journal of Computational Science*, 2021, vol. 53, pp. 101345. <https://doi.org/10.1016/j.jocs.2021.101345>
 15. Li K., Deb K., Zhang Q., Kwong S. An evolutionary many-objective optimization algorithm based on dominance and decomposition. *IEEE Transactions on Evolutionary Computation*, 2015, vol. 19, no. 5, pp. 694–716. <https://doi.org/10.1109/TEVC.2014.2373386>
 16. Das I., Dennis J.E. Normal-boundary intersection: A new method for generating the Pareto surface in nonlinear multicriteria optimization problems. *SIAM Journal on Optimization*, 1998, vol. 8, no. 3, pp. 631–657. <https://doi.org/10.1137/s1052623496307510>
 13. Van Breugel F., Kutz J.N., Brunton B.W. Numerical differentiation of noisy data: A unifying multi-objective optimization framework // *IEEE Access*, 2020, V. 8, P. 196865–196877. <https://doi.org/10.1109/access.2020.3034077>
 14. Maslyaev M., Hvatov A., Kalyuzhnaya A. Partial differential equations discovery with EPDE framework: Application for real and synthetic data // *Journal of Computational Science*, 2021, V. 53, P. 101345. <https://doi.org/10.1016/j.jocs.2021.101345>
 15. Li K., Deb K., Zhang Q., Kwong S. An evolutionary many-objective optimization algorithm based on dominance and decomposition // *IEEE Transactions on Evolutionary Computation*, 2015, V. 19, N 5, P. 694–716. <https://doi.org/10.1109/TEVC.2014.2373386>
 16. Das I., Dennis J.E. Normal-boundary intersection: A new method for generating the Pareto surface in nonlinear multicriteria optimization problems // *SIAM Journal on Optimization*, 1998, V. 8, N 3, P. 631–657. <https://doi.org/10.1137/s1052623496307510>

Authors

Mikhail A. Maslyaev — Junior Researcher, ITMO University, Saint Petersburg, 197101, Russian Federation, <https://orcid.org/0000-0001-5595-0802>, mikemaslyaev@itmo.ru

Alexander A. Hvatov — PhD (Physics & Mathematics), Head of Laboratory, ITMO University, Saint Petersburg, 197101, Russian Federation, <https://orcid.org/0000-0002-5222-583X>, alex_hvatov@itmo.ru

Received 11.10.2022

Approved after reviewing 07.12.2022

Accepted 15.01.2023

Авторы

Масляев Михаил Александрович — младший научный сотрудник, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, <https://orcid.org/0000-0001-5595-0802>, mikemaslyaev@itmo.ru

Хватов Александр Александрович — кандидат физико-математических наук, заведующий лабораторией, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, <https://orcid.org/0000-0002-5222-583X>, alex_hvatov@itmo.ru

Статья поступила в редакцию 11.10.2022

Одобрена после рецензирования 07.12.2022

Принята к печати 15.01.2023



Работа доступна по лицензии
Creative Commons
«Attribution-NonCommercial»

doi: 10.17586/2226-1494-2023-23-1-105-111

Probabilistic criteria for time-series predictability estimation

Anton N. Kovantsev✉

ITMO University, Saint Petersburg, 197101, Russian Federation

ankovantsev@itmo.ru✉, <https://orcid.org/0000-0003-1765-7001>

Abstract

Assessing the time series predictability is necessary for forecasting models validating, for classifying series to optimize the choice of the model and its parameters, and for analyzing the results. The difficulties in assessing predictability occur due to large heteroscedasticity of errors obtained when predicting several series of different nature and characteristics. In this work, the internal predictability of predictive modeling objects is investigated. Using the example of time series forecasting, we explore the possibility of quantifying internal predictability in terms of the probability (frequency) of obtaining a forecast with an error greater than some certain level. We also try to determine the relationship of such a measure with the characteristics of the time series themselves. The idea of the proposed method is to estimate the internal predictability by the probability of an error exceeding a predetermined threshold value. The studies were carried out on data from open sources containing more than seven thousand time series of stock market prices. We compare the probability of errors which exceed the allowable value (miss probabilities) for the same series on different forecasting models. We show that these probabilities differ insignificantly for different forecasting models with the same series, and hence, the probability can be a measure of predictability. We also show the relationship of the miss probability values with entropy, the Hurst exponent, and other characteristics of the series according to which the predictability can be estimated. It has been established that the resulting measure makes it possible to compare the predictability of time series with pronounced heteroscedasticity of forecast errors and when using different models. The measure is related to the characteristics of the time series and is interpretable. The results can be generalized to any objects of predictive modeling and forecasting quality scores. It can be useful to developers of predictive modeling algorithms, machine learning specialists in solving practical problems of forecasting.

Keywords

intrinsic predictability, forecasting error, misprediction

Acknowledgements

This research is supported by The Russian Science Foundation, Agreement No. 17-71-30029.

For citation: Kovantsev A.N. Probabilistic criteria for time-series predictability estimation. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 2023, vol. 23, no. 1, pp. 105–111. doi: 10.17586/2226-1494-2023-23-1-105-111

УДК 519.246.2

Вероятностный критерий оценки предсказуемости временных рядов

Антон Николаевич Кованцев✉

Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация

ankovantsev@itmo.ru✉, <https://orcid.org/0000-0003-1765-7001>

Аннотация

Предмет исследования. Задача оценки предсказуемости временных рядов возникает при валидации моделей прогнозирования, при классификации рядов с целью оптимизации выбора модели и ее параметров, при анализе результатов. Большая гетероскедастичность ошибок, получаемых при прогнозировании нескольких различных по природе и характеристикам рядов, часто приводит к затруднениям при оценке предсказуемости. В работе исследована внутренняя предсказуемость объектов предсказательного моделирования. На примере прогнозирования временных рядов определена возможность количественной оценки внутренней

предсказуемости по вероятности (частоте) получения прогноза с ошибкой, больше заданного уровня, и связь такой меры с характеристиками самих временных рядов. **Метод.** Суть предлагаемого метода состоит в оценивании внутренней предсказуемости по вероятности возникновения ошибки, большей заранее заданного порогового значения. **Основной результат.** Исследования выполнены на данных из открытых источников, содержащих более 7000 временных рядов биржевых котировок. Проведено сопоставление полученных значений вероятности возникновения ошибок, превосходящих допустимое значение (вероятностей промаха) для одних и тех же рядов на различных моделях прогнозирования. Показано, что при использовании моделей с одним и тем же рядом эти вероятности отличаются незначительно и могут служить мерой предсказуемости. Выявлена связь полученных значений вероятности с энтропией, показателем Хёрста и иными характеристиками рядов, по которым оценивается предсказуемость. Установлено, что полученная мера позволяет сравнивать предсказуемость временных рядов при выраженной гетероскедастичности ошибок прогнозирования и при применении разных моделей. Мера связана с характеристиками временного ряда и интерпретируема. **Практическая значимость.** Полученные результаты могут быть обобщены на любые объекты предсказательного моделирования и меры оценки качества прогноза. Результаты исследования будут полезны разработчикам алгоритмов предсказательного моделирования и специалистам по машинному обучению, при решении практических задач прогнозирования.

Ключевые слова

внутренняя предсказуемость, ошибка прогнозирования, вероятность промаха

Благодарности

Работа выполнена при поддержке Российского научного фонда, соглашение № 17-71-30029.

Ссылка для цитирования: Кованцев А.Н. Вероятностный критерий оценки предсказуемости временных рядов // Научно-технический вестник информационных технологий, механики и оптики. 2023. Т. 23, № 1. С. 105–111 (на англ. яз.). doi: 10.17586/2226-1494-2023-23-1-105-111

Introduction

In almost every case when we deal with time series forecasting, we need a meaningful and understandable measure of predictability to evaluate the results. In other words, we want to know whether our model achieved the best possible quality or not. Hence, some measure of intrinsic predictability is necessary which could explain how likely a misprediction is, or what the range of errors might occur, or what variance of the errors is expected for the certain series. And of course, it would be very nice if we could get all these aspects before the model is constructed, fitted, and running. So, we're going to find the way to answer these questions by time series analysis. For a rather big dataset we can calculate different features whose connection to predictability was approved by related works, and build some regression that hypothetically can connect these features to misprediction probability or errors variance. Obviously, we can't avoid involving some forecasting model which must produce these probabilities and variances, but for this research we presume that this experience with one model can be generalized to the class or classes of models. At least, for the future work we plan to use several models of different classes in ensemble which will approximate the real metrics of intrinsic predictability more accurately.

Related works

The earliest mention of the idea to distinguish realized and intrinsic predictability, which we could find, was proclaimed by Edward N. Lorenz in [1] that refers to a 1996 paper. The philosophic discussion on the predictability issues in various senses is going on till nowadays. For instance, Stefan Rummens [2] argues with Victor Gijsbers [3]. Meanwhile, this discussion is as interesting and entertaining as it is far from practical use and everyday needs. At the same time not so many authors attempted to find some quantity measure for time-series intrinsic

predictability. The idea to match some time series features or their combination to intrinsic predictability was discussed in [4] where predictability was quantified with permutation entropy, and in [5] where several features including Hurst exponent and Kolmogorov-Sinai entropy were used for series clustering according to their predictability. The similar approach base on transforming a time series to graph is proposed in [6]. All these methods are based on forecasting errors estimation, and they don't consider the fact that the series of bad predictability can perform rather good forecasting quality. Having this fact in mind, the authors of [7] state that intrinsic predictability of chaotic systems might be high, but the realized predictability is expected to be low and difficult to improve substantially. In [8] the rank-based nonlinear predictability score was adapted to time series sampled from time-continuous flows and performed a higher sensitivity for deterministic structure in noisy signals.

One more approach is represented in [9] where intrinsic predictability is estimated by wavelet entropy energy measure after time series wavelet transformation. Besides the forecasting error, they also use Nash–Sutcliffe efficiency for quality estimation.

We failed to find any research where statistical or probabilistic were used for matching of the predictability measure, so off we go.

Real-world data

In our experiments we used the open dataset called Huge Stock Market Dataset from Kaggle which contains historical daily prices and volumes of all U.S. stocks and ETFs. There are more than 7,000 time series mostly of fractal nature [10] with a significant part of the random-walking process. Nevertheless, we expected that this set includes the series of different intrinsic predictability and would be sufficient for the aim of our research.

We dropped the short series with less than 730 observations and shrunk the rest up to last 730 points.

So, our series are neither too short to make the predictive model to do its best, nor too long to satisfy memory and computation time requirements. The number of series after such preprocessing was reduced to 5,121. Every data file in the collection contains the series of opening, high, low, closing prices and trading volume per day. As all the prices are rather close to each other, we take the closing price for the target series.

We use the set of the time series features the same as was used in [5]: Kolmogorov-Sinai entropy, Hurst exponent, embedding dimension, noise measure and random-walk detection. Permutation entropy was explored too, as many researchers usually mention in connection to predictability.

The Kolmogorov-Sinai entropy (h_μ^{KS}) can be calculated based on the entropy rates of finite partitions of the state space of the series [11]. For each of the n finite partitions $\xi = \{C_1, C_2, \dots, C_n\}$ of the state space $M = \bigcup_{i=1}^n C_i$ with dynamics given by a measurable transformation T and defined probability measure $\mu(T^{-1}C_{i_1} \cap \dots \cap T^{-n}C_{i_n})$, the entropy rate is

$$h_\mu(T, \xi) = -\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i_1, \dots, i_n} \mu(T^{-1}C_{i_1} \cap \dots \cap T^{-n}C_{i_n}) \times \ln \mu(T^{-1}C_{i_1} \cap \dots \cap T^{-n}C_{i_n}).$$

We take the supremum of the entropy rate over all finite partitions:

$$h_\mu^{KS}(T) = \sup_{\xi} h_\mu(T, \xi).$$

Hurst exponent is used as a measure of long-term memory of time series. According to [12], we estimate it by the re-scaled range:

$$R(\tau) = \max_{1 \leq t \leq \tau} \sum_{j=1}^{\tau} (x_j - \bar{x}_\tau) - \min_{1 \leq t \leq \tau} \sum_{j=1}^{\tau} (x_j - \bar{x}_\tau)$$

and standard deviation:

$$S(\tau) = \sqrt{\frac{1}{\tau} \sum_{t=1}^{\tau} (x_t - \bar{x}_\tau)^2},$$

where $\tau \in [3, N]$ are the time steps for a discrete time series, $x(t)$ is the series value on step t .

Hurst exponent can be calculated as:

$$H(\tau) = \lim_{\tau \rightarrow \infty} \frac{\log \frac{R(\tau)}{S(\tau)}}{\log(\alpha\tau)},$$

where $\alpha \approx 0.5$ is a Hirst's empirically found constant.

These re-scaled range $R(\tau)$ and standard deviation $S(\tau)$ are used to represent the R/S -trajectory $RS(\tau) = \frac{R(\tau)}{S(\tau)}$ that helps to estimate the *memory depth* of the time series.

Embedding dimension is a measure of the dimensionality of the space occupied by a set of random points, often referred to as a type of fractal dimension. It's less noisy when only a small number of points is available and is often in agreement with other calculations of

dimension. It can be calculated by means of the correlative integral for time series of finite length:

$$C(r) = \sum_{i=1}^m \sum_{j=i+1}^m \frac{\theta(r - \rho(i, j))}{m(m-1)},$$

where $\rho_k(i, j) = \sqrt{\sum_{l=1}^k (x_{i-k+l} - x_{j-k+l})^2}$; $\theta(x)$ — Heaviside function; r — characteristic phase space cell size.

The value of embedding dimension is the slope of the logarithmic graph of the correlation integral [11] and it can be evaluated as the following limit:

$$d_k = \lim_{r \rightarrow 0} \lim_{m \rightarrow \infty} \frac{\ln C(r)}{\ln r}.$$

Noise measure feature is based on the idea of Robert M. May, firstly published in 1976 [13], to compare the standard deviation of a time-series with the standard deviation of its first-order differences.

$$F_N = 1 - \sqrt{\frac{N \sum_{i=1}^{N-1} (x'_i - \bar{x}')^2}{(N-1) \sum_{i=1}^{N-1} (x_i - \bar{x})^2}},$$

where $x'_i = x_{i+1} - x_i$ is the first-order differences; \bar{x} , \bar{x}' are mean values of the initial time series and corresponding differenced series. This measure can be used to find a random walk process for which the first-order differences should be noise.

The idea is to detect high-frequency noise that increases the comparative deviation of value change on each time step; so, the higher value of this measure means the lower noise influence in time series.

Forecasting and realized predictability

In order to collect information about forecasting quality statistics, we launched the Extreme Gradient Boosting (XGB) forecasting model for each time series of the data set. Hyperparameters were chosen empirically for the best performance on 30-days horizon for most of the series. Thus, we got 20 estimators and maximal depth equal to 8. Every single time series is forecasted 107 times. We take 365-days observations as the training period and 30-days forecast as a test, then we repeat the same with a 3-day time shift. The stock market time series are chaotic enough to eliminate the effect of dependencies in the sequential forecasting experiments. The procedure is like a rolling window. So, for each of the series, we get 107 values of Mean Absolute Percentage Error (MAPE) and calculate the mean value and standard deviation for these error series, and for the part of errors which is greater than 10 %, we consider that an error of less than 10 % is sufficient for the good forecast quality. This part helps to estimate the misprediction probability.

Of course, this method of predictability estimation is inseparably connected with the forecasting model. This connection could be broken if we show that the same metrics for some other models of different kinds are either close or at least correlated with those for our basic model.

Local approximation (LA) and Maximal Similarity (MS) were applied in the same conditions as the alternate models. The first one was designed by prof. Alexander Loskutov [11], and it is based on neighborhood of points of system trajectory in the state space; the other one was invented by Irina Chuchueva [14], and it uses the rescaled patterns of a certain length in the series to make a forecast. Both models are described in detail in [5] and [6]. In Fig. 1, *a* we can see that squared misprediction probability for MS correlates with that for XGB with coefficient of 0.94. As the values distribution is obviously far from normal, and it will be shown further, we use the Spearman's rank coefficient to estimate the correlations. In the case of LA there is no need to deal with squared values. The correlation is 0.92, the scatter plot is presented in Fig. 1, *b*.

So, we can conclude that at least for three types of forecasting models the scores of misprediction probability are correlated and this measure could be considered model independent.

Regression model for intrinsic predictability measuring

As we have got a set of time series features and the values which we consider to be the measure of predictability, we can try to find some dependence between them. The scatter matrix for the features shown in Fig. 2 looks not very promising. Pairwise correlation coefficients between the features and scores are collected in Table. They are not great as well, but nevertheless we try to build a regression model. This picture also confirms that, as we noted earlier, the value distributions are not normal. Besides, it illustrates the large heteroscedasticity of errors. For instance, the more noise, the wider is standard deviation range.

Presumably, models of linear regression would fail with such sort of predictors. We use an Extra-Trees regressor which fits a number of randomized decision trees on various sub-samples of the dataset and uses averaging to improve the predictive accuracy and control over-fitting. Its

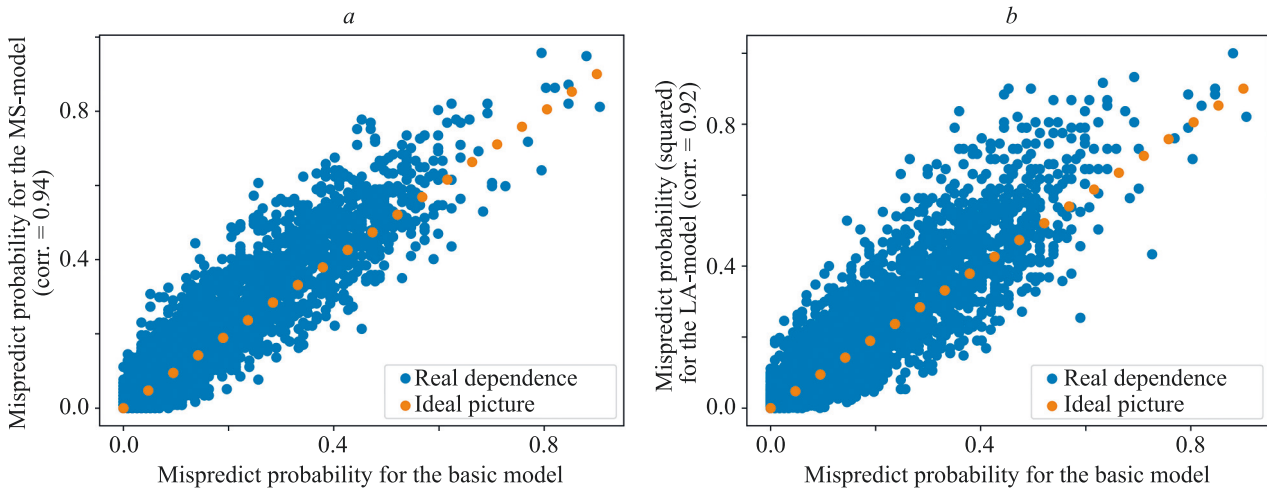


Fig. 1. Misprediction probability for MS (*a*) and LA (*b*) models regarding to XGB model

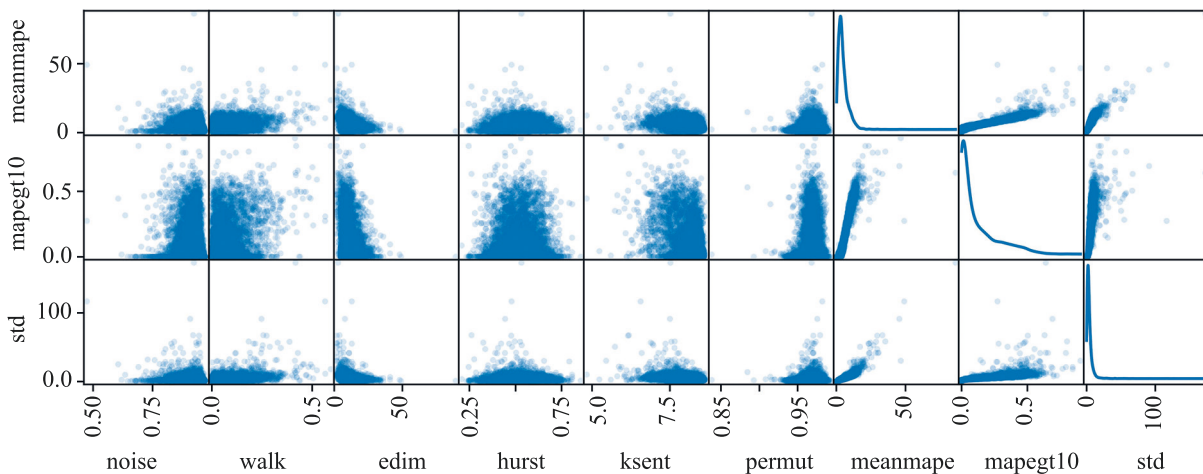


Fig. 2. Scatter matrix for features and scores.

Features: noise — noise measure; walk — random walk measure; edim — embedding dimension; hurst — Hurst exponent; kscent — Kolmogorov-Sinai entropy; permut — permutation entropy. Scores: meanmap — mean MAPE value for all samples; mapect10 — the part of samples with MAPE>10 %, std — standard deviation

Table. Spearman correlation coefficients for features and scores

	Noise measure	Random walk	Embedding dimension	Hurst exponent	Kolmogorov-Sinai entropy	Permutation entropy
Mean MAPE value	-0.12	0.320	-0.17	0.02	-0.370	0.13
Part of samples with MAPE >10 %	-0.11	0.290	-0.18	0.04	-0.383	0.11
Standard deviation	-0.12	0.339	-0.16	-0.01	-0.330	0.09

hyper-parameters were chosen empirically: 128 estimators with minimal samples split of 16 and min samples leaf of 2. We use coefficient of determination (R^2) for regression scoring. This regressor performance score was $R^2 = 0.60$ on the train set, $R^2 = 0.32$ and $R^2 = 0.24$ on 50-times splitting cross-validation. The Mean Absolute Error is 0.12 for 1706 test samples. The error distribution analysis shows that most of errors are in the range of ± 0.2 and there is a bias about +0.07.

We also tried Random Forest Regression model, Gradient Boosting, and based not on random trees model of K -Neighbors Regression. None of them performs better regression quality.

Experiment results and discussion

As we can see in the previous section, the regressor performance is rather poor, so, our experiments are aimed to find out whether it's possible to get some use in spite of that. Table shows that correlations between the features and responses are weak which explains the low score of the regression.

The further feature analysis by means of Shapley values shows that the impact of the features on the result could be opposite for different series which is reflected as the mess of the red and blue dots on the bee-swarm plot in Fig. 3. All the features are actually important for Extra-Trees regression. Kolmogorov-Sinai entropy is most important with relative importance score of 0.33, the least one is

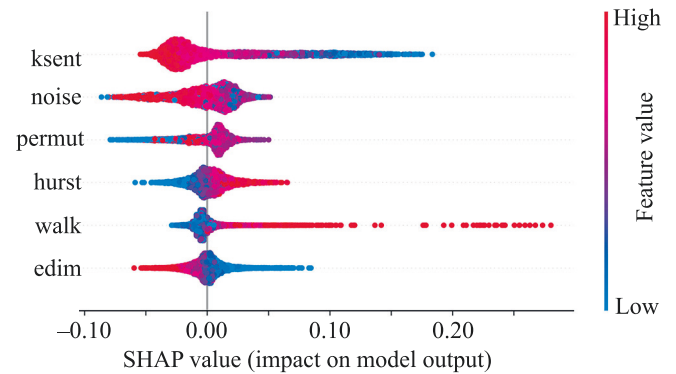


Fig. 3. Features impact analysis by Shapley values (Designations are the same as those in Fig. 2)

Hurst exponent with score of 0.08. We also can notice that Kolmogorov-Sinai entropy is much more correlated with misprediction probability, and its importance is more than those for usually used permutation entropy. The mess of Shapley values at least for our dataset is not that big too.

In the last item of our research agenda, we have got the misprediction probability values for the real model and estimated it by the regressor. The distribution of these values is illustrated in Fig. 4, *a* as a histogram which shows that the regressor tends to overestimate the probability due to its bias, and the range of real values is wider.

To get a more detailed picture of the regressor performance, we can see the scatter plot in Fig. 4, *b* with

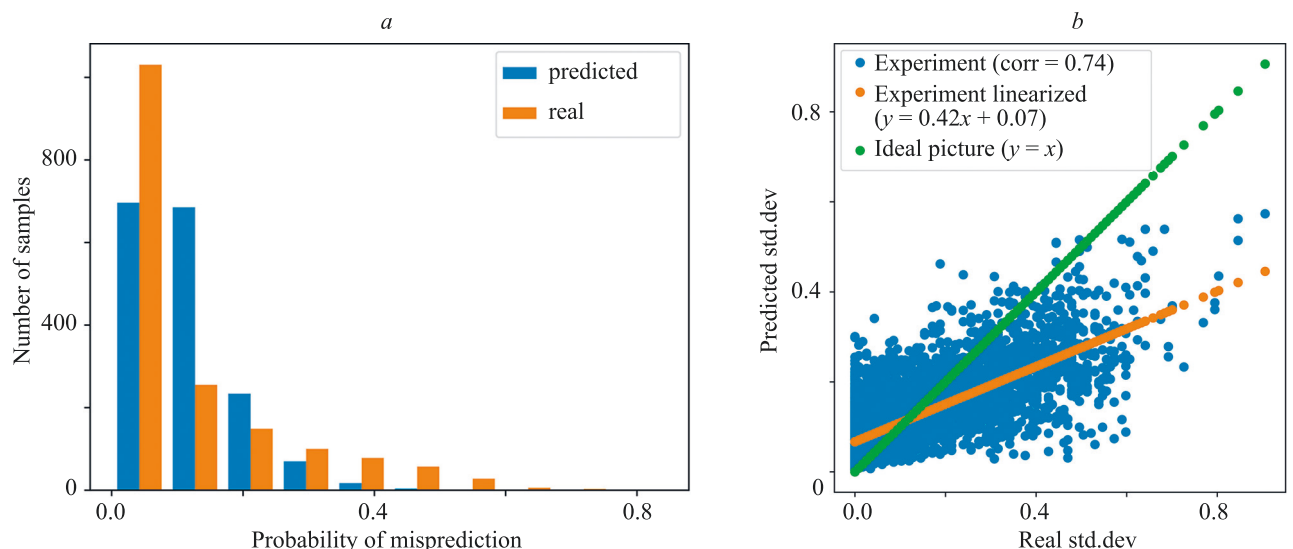


Fig. 4. Distribution of real and estimated misprediction probability as a histogram (*a*) and as a scatter plot with the least squares approximation (*b*)

the least squares approximation and ideal line shown. The correlation and distribution of the real and predicted values are quite suitable for time series predictability estimation. We suppose that for some not very precise real-world tasks that estimation would be enough.

Future work

For our future work we plan to increase the number and variety of our experiments on predictability. Of course, it's necessary to track the dependence of misprediction probability on the forecasting horizon. The time series of different shapes and nature should be tested as well. It would be better to get more different models for experiments and maybe to take the best quality from among all models for the regressor training.

Besides, as we said above, the other score could be useful for time-series predictability measuring. It is forecasting errors standard deviation. This score also worth to be explored. To continue the predictability exploring, it could be useful to take into account not only errors themselves but the shape of their distribution.

Conclusion

Our probabilistic method of intrinsic predictability evaluation is an attempt to pay attention to the numerous noticed fact that the bad forecasting quality not obligatory can occur for the series which features indicate bad

predictability. Our regressor helps to estimate the probability of poor-quality forecast despite the rather big errors of the regressor.

The proposed method of predictability evaluation looks very much like the simple confidence interval calculation. But there is some difference. First, making up the confidence interval we first define the confidence level and then calculate the thresholds for predicted values. Here we approach from the opposite side. We settle the thresholds and count how often the shooter hits it. Then, for the statistical calculation of the confidence interval the normally distributed value is strongly desired. In our case we were never tied by such a requirement. The confidence interval is used mostly for continuous values and doesn't work with classification and ranking models. Our approach can be applied to any quality score for almost every type of model. We may set a threshold on F-score or accuracy and count the amount of hits.

Besides, the time series predictability is usually estimated by one of the series features like entropy or Hurst exponent or something else. We try to get use of a set of several features which turns out to be rather helpful in such estimation.

The benefit of proposed approach to time series predictability scoring is in direct evaluation of misprediction probability in quite understandable terms which suits for all sorts of time series and depends only on the forecasting horizon.

References

1. Lorenz E.N. Predictability — a problem partly solved. *Predictability of Weather and Climate*. Cambridge University Press, 2006, pp. 40–58. <https://doi.org/10.1017/CBO9780511617652.004>
2. Rummens S. The roots of the paradox of predictability: A reply to gijsbers. *Erkenntnis*, 2022. <https://doi.org/10.1007/s10670-022-00617-8>
3. Gijsbers V. The paradox of predictability. *Erkenntnis*, 2021. <https://doi.org/10.1007/s10670-020-00369-3>
4. Pennekamp F., Iles A.C., Garland J., Brennan G., Brose U., Gaedke U., Jacob U., Kratina P., Matthews B., Munch S., Novak M., Palamara G.M., Rall B.C., Rosenbaum B., Tabi A., Ward C., Williams R., Ye H., Petchey O.L. The intrinsic predictability of ecological time series and its potential to guide forecasting. *Ecological Monographs*, 2019, vol. 89, no. 2, pp. e01359. <https://doi.org/10.1002/ecm.1359>
5. Kovantsev A., Gladilin P. Analysis of multivariate time series predictability based on their features. *Proc. of the IEEE International Conference on Data Mining Workshops (ICDMW)*, 2020, pp. 348–355. <https://doi.org/10.1109/icdmw51313.2020.00055>
6. Kovantsev A., Chunaev P., Bochenina K. Evaluating time series predictability via transition graph analysis. *Proc. of the IEEE International Conference on Data Mining Workshops (ICDMW)*, 2021. <https://doi.org/10.1109/ICDMW53433.2021.00135>
7. Beckage B., Gross L.J., Kauffman S. The limits to prediction in ecological systems. *Ecosphere*, 2011, vol. 2, no. 11, pp. 1–12. <https://doi.org/10.1890/ES11-00211.1>
8. Naro D., Rummel C., Schindler K., Andrzejak R.G. Detecting determinism with improved sensitivity in time series: Rank-based nonlinear predictability score. *Physical Review E*, 2014, vol. 90, no. 3, pp. 032913. <https://doi.org/10.1103/PhysRevE.90.032913>
9. Guntu R.K., Yeditha P.K., Rathinasamy M., Perc M., Marwan N., Kurths J., Agarwal A. Wavelet entropy-based evaluation of intrinsic predictability of time series. *Chaos*, 2020, vol. 30, no. 3, pp. 033117. <https://doi.org/10.1063/1.5145005>

Литература

1. Lorenz E.N. Predictability - a problem partly solved // *Predictability of Weather and Climate*. Cambridge University Press, 2006. P. 40–58. <https://doi.org/10.1017/CBO9780511617652.004>
2. Rummens S. The roots of the paradox of predictability: A reply to gijsbers // *Erkenntnis*. 2022. <https://doi.org/10.1007/s10670-022-00617-8>
3. Gijsbers V. The paradox of predictability // *Erkenntnis*. 2021. <https://doi.org/10.1007/s10670-020-00369-3>
4. Pennekamp F., Iles A.C., Garland J., Brennan G., Brose U., Gaedke U., Jacob U., Kratina P., Matthews B., Munch S., Novak M., Palamara G.M., Rall B.C., Rosenbaum B., Tabi A., Ward C., Williams R., Ye H., Petchey O.L. The intrinsic predictability of ecological time series and its potential to guide forecasting // *Ecological Monographs*. 2019. V. 89. N 2. P. e01359. <https://doi.org/10.1002/ecm.1359>
5. Kovantsev A., Gladilin P. Analysis of multivariate time series predictability based on their features // *Proc. of the IEEE International Conference on Data Mining Workshops (ICDMW)*. 2020. P. 348–355. <https://doi.org/10.1109/icdmw51313.2020.00055>
6. Kovantsev A., Chunaev P., Bochenina K. Evaluating time series predictability via transition graph analysis // *Proc. of the IEEE International Conference on Data Mining Workshops (ICDMW)*. 2021. <https://doi.org/10.1109/ICDMW53433.2021.00135>
7. Beckage B., Gross L.J., Kauffman S. The limits to prediction in ecological systems // *Ecosphere*. 2011. V. 2. N 11. P. 1–12. <https://doi.org/10.1890/ES11-00211.1>
8. Naro D., Rummel C., Schindler K., Andrzejak R.G. Detecting determinism with improved sensitivity in time series: Rank-based nonlinear predictability score // *Physical Review E*. 2014. V. 90. N 3. P. 032913. <https://doi.org/10.1103/PhysRevE.90.032913>
9. Guntu R.K., Yeditha P.K., Rathinasamy M., Perc M., Marwan N., Kurths J., Agarwal A. Wavelet entropy-based evaluation of intrinsic predictability of time series // *Chaos*. 2020. V. 30. N 3. P. 033117. <https://doi.org/10.1063/1.5145005>

10. Mandelbrot B.B., Hudson R.L. *The (Mis)behaviour of Markets: A Fractal View of Risk, Ruin and Reward*. Basic Books, 2004.
11. Loskutov Y., Kotlyarov O.L., Istomin I.A., Zhuravlev D.I. Problems of nonlinear dynamics: III. Local methods of time series forecasting. *Moscow University Physics Bulletin*, 2002, vol. 57, no. 6.
12. Mitra S.K. Is Hurst exponent value useful in forecasting financial time series? *Asian Social Science*, 2012, vol. 8, no. 8, pp. 111–120. <https://doi.org/10.5539/ass.v8n8p111>
13. May R.M. Simple mathematical models with very complicated dynamics. *The Theory of Chaotic Attractors*, 2004, pp. 85–93. https://doi.org/10.1007/978-0-387-21830-4_7
14. Chuchueva I.A. The time series extrapolation model based on maximum likeness set. *Information technologies. Information Technologies*, 2010, no. 12, pp. 43–47. (in Russian)

Author

Anton N. Kovantsev — Engineer, ITMO University, Saint Petersburg, 197101, Russian Federation,  57222088145, <https://orcid.org/0000-0003-1765-7001>, ankovantcev@itmo.ru


Received 13.10.2022

Approved after reviewing 07.12.2022

Accepted 22.01.2023

10. Mandelbrot B.B., Hudson R.L. *The (Mis)behaviour of Markets: A Fractal View of Risk, Ruin and Reward*. Basic Books, 2004.
11. Loskutov Y., Kotlyarov O.L., Istomin I.A., Zhuravlev D.I. Problems of nonlinear dynamics: III. Local methods of time series forecasting // *Moscow University Physics Bulletin*. 2002. V. 57. N 6.
12. Mitra S.K. Is Hurst exponent value useful in forecasting financial time series? // *Asian Social Science*. 2012. V. 8. N 8. P. 111–120. <https://doi.org/10.5539/ass.v8n8p111>
13. May R.M. Simple mathematical models with very complicated dynamics // *The Theory of Chaotic Attractors*. 2004. P. 85–93. https://doi.org/10.1007/978-0-387-21830-4_7
14. Чучуева И.А. Модель экстраполяции временных рядов по выборке максимального подобия // *Информационные технологии*. 2010. № 12. С. 43–47.

Автор

Кованцев Антон Николаевич — инженер, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация,  57222088145, <https://orcid.org/0000-0003-1765-7001>, ankovantcev@itmo.ru

Статья поступила в редакцию 13.10.2022

Одобрена после рецензирования 07.12.2022

Принята к печати 22.01.2023



Работа доступна по лицензии
Creative Commons
«Attribution-NonCommercial»

doi: 10.17586/2226-1494-2023-23-1-112-120

УДК 004.912

Программный фреймворк для оптимизации гиперпараметров тематических моделей с аддитивной регуляризацией

Мария Андреевна Ходорченко¹✉, Николай Алексеевич Бутаков²,
Денис Александрович Насонов³, Михаил Юрьевич Фирулик⁴

^{1,2,3} Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация

⁴ ООО «Оператор Газпром ИД», Санкт-Петербург, 191028, Российская Федерация

¹ mariyaxod@yandex.ru✉, <https://orcid.org/0000-0001-5446-5311>

² alipoov.nb@gmail.com, <https://orcid.org/0000-0002-2705-1313>

³ denis.nasonov@gmail.com, <https://orcid.org/0000-0002-2740-0173>

⁴ firulik@gmail.com, <https://orcid.org/0000-0001-7114-1052>

Аннотация

Предмет исследования. Обработка неструктурированных данных, таких как тексты на естественном языке, является одной из актуальных задач при разработке интеллектуальных продуктов. В свою очередь, тематическое моделирование как метод работы с неразмеченными и частично размеченными текстовыми данными активно используется для анализа корпусов документов и создания векторных представлений. В связи с этим особенно важно обучение качественных тематических моделей за короткое время, что возможно с помощью предложенного фреймворка. **Метод.** Разработанный фреймворк реализует эволюционный подход к оптимизации гиперпараметров моделей с аддитивной регуляризацией и высокими результатами по метрикам качества (когерентность, NPMI). Для уменьшения вычислительного времени представлен режим работы с суррогатными моделями, который обеспечивает ускорение вычислений до 1,8 раз без потери качества. **Основные результаты.** Эффективность фреймворка продемонстрирована на трех наборах данных с разными статистическими характеристиками. Получены результаты, превосходящие аналогичные решения в среднем на 20 % по когерентности и 5 % по качеству классификации для двух из трех наборов. Создана распределенная версия фреймворка для проведения экспериментальных исследований тематических моделей. **Практическая значимость.** Полученный фреймворк может быть использован пользователями без специальных знаний в области тематического моделирования, благодаря выстроенному пайплайну работы с данными. Результаты работы могут применяться исследователями для проведения анализа тематических моделей и расширения функционала.

Ключевые слова

AutoML фреймворк, тематическое моделирование, неструктурированные данные, аддитивная регуляризация, эволюционный подход, суррогатные модели

Благодарности

Исследование выполнено при финансовой поддержке Российского научного фонда, проект № 20-11-20270.

Ссылка для цитирования: Ходорченко М.А., Бутаков Н.А., Насонов Д.А., Фирулик М.Ю. Программный фреймворк для оптимизации гиперпараметров тематических моделей с аддитивной регуляризацией // Научно-технический вестник информационных технологий, механики и оптики. 2023. Т. 23, № 1. С. 112–120. doi: 10.17586/2226-1494-2023-23-1-112-120

Software framework for hyperparameters optimization of models with additive regularization

Maria A. Khodorchenko^{1✉}, Nikolay A. Butakov², Denis A. Nasonov³, Mikhail Yu. Firulik⁴

^{1,2,3} ITMO University, Saint Petersburg, 197101, Russian Federation

⁴ LLC "Operator Gazprom ID", Saint Petersburg, 191028, Russian Federation

¹ mariyaxod@yandex.ru✉, <https://orcid.org/0000-0001-5446-5311>

² alipoov.nb@gmail.com, <https://orcid.org/0000-0002-2705-1313>

³ denis.nasonov@gmail.com, <https://orcid.org/0000-0002-2740-0173>

⁴ firulik@gmail.com, <https://orcid.org/0000-0001-7114-1052>

Abstract

The processing of unstructured data, such as natural language texts, is one of the urgent tasks in the development of intelligent products. In turn, topic modeling as a method of working with unmarked and partially marked text data is a natural choice for analyzing document bodies and creating vector representations. In this regard, it is especially important to train high-quality thematic models in a short time which is possible with the help of the proposed framework. The developed framework implements an evolutionary approach to optimizing hyperparameters of models with additive regularization and high results on quality metrics (coherence, NPMI). To reduce the computational time, a mode of working with surrogate models is presented which provides acceleration of calculations up to 1.8 times without loss of quality. The effectiveness of the framework is demonstrated on three datasets with different statistical characteristics. The results obtained exceed similar solutions by an average of 20 % in coherence and 5 % in classification quality for two of the three datasets. A distributed version of the framework has been developed for conducting experimental studies of topic models. The developed framework can be used by users without special knowledge in the field of topic modeling due to the default data processing pipeline. The results of the work can be used by researchers to analyze topic models and expand functionality.

Keywords

AutoML framework, topic modeling, unstructured data, additive regularization, evolutionary approach, surrogate models

Acknowledgements

This research is financially supported by The Russian Science Foundation, Agreement No. 20-11-20270.

For citation: Khodorchenko M.A., Butakov N.A., Nasonov D.A., Firulik M.Yu. Software framework for hyperparameters optimization of models with additive regularization. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 2023, vol. 23, no. 1, pp. 112–120 (in Russian). doi: 10.17586/2226-1494-2023-23-1-112-120

Введение

В настоящее время ведется активная разработка и внедрение интеллектуальных продуктов в разнообразные бизнес-процессы и клиентские услуги. При этом большую популярность получает работа с неструктурированными данными, такими как тексты на естественном языке (документы, отзывы, комментарии и т. д.). Для их качественной обработки необходимо проведение разведочного анализа, основным инструментом которого, для текстовых данных, является такой метод обучения без учителя как тематическое моделирование. Использование этого метода позволяет получать скрытые компоненты (под ними понимаются темы), и их представленность в корпусе документов. Таким образом, текст превращается в интерпретируемое представление, которое можно использовать для решения актуальных проблем, таких как построение профилей интересов пользователей [1, 2], извлечение сентимента [3], разделение данных на тематические и структурно близкие подмножества [4], обогащение контекстуализированных представлений документов [5] и другие.

Существующие методы тематического моделирования можно отнести к нескольким большим семействам: на основе матричных разложений (неотрицательная матричная факторизация (Non-negative Matrix Factorization, NMF) [6]), классические вероятностные (например, вероятностный латентный семантический анализ (Probabilistic Latent Semantic Analysis, PLSA) [7] и латентное распределение Дирихле (Latent Dirichlet

Allocation, LDA) [8]), с аддитивной регуляризацией (аддитивная регуляризация тематических моделей (Additive Regularization of Topic Models, ARTM) [9]) и нейронные модели [10–12]. Долгое время наиболее предпочтительным методом, показывавшим высокие результаты качества, являлось LDA, но его использование на корпусах документов, обладавших спецификой (например, наличие терминологии или малое количество слов в документе) затруднялось необходимостью модификации аппарата обучения модели [13, 14], что требовало от разработчика соответствующих знаний. В свою очередь, развивающиеся нейронные модели демонстрируют высокую степень приспособленности под заданную оценку качества, но при этом, в случае работы с контекстуализированными представлениями, являются вычислительно затратными. К оптимальным и перспективным вариантам можно отнести подход с аддитивной регуляризацией, который обобщает понятие тематического моделирования и обеспечивает простую оптимизацию с использованием алгоритма максимизации ожидания (expectation-maximization, EM-алгоритм).

При работе с моделями, имеющими гиперпараметры, возникает необходимость в их настройке. Существуют различные программные решения для автоматической оптимизации тематических моделей различных семейств. В genism [15] реализованы «классические методы» тематического моделирования, такие как LDA, и возможность настройки их параметров. Библиотека OCTIS [16] предоставляет возможность

настройки ряда тематических моделей, включая нейронные, с помощью байесовской оптимизации, что обеспечивает достаточно эффективную и быструю работу. К недостаткам можно отнести отсутствие специфичных алгоритмов оптимизации под разные семейства моделей. TopicNet работает с семейством моделей с аддитивной регуляризацией посредством жадного алгоритма настройки. При этом автоматическая настройка осложняется отсутствием единых метрик измерения качества [17].

В данной работе представлен фреймворк для настройки моделей с аддитивной регуляризацией, позволяющий за ограниченное время получать тематические модели высокого качества сразу по ряду существующих метрик.

Основные технологии

Рассмотрим базовые технологии, которые лежат в основе разработанного AutoML фреймворка для задачи настройки моделей тематического моделирования.

BigARTM [18] — библиотека для обучения моделей с аддитивной регуляризацией. Обеспечивает быструю подготовку и инференс из готовых моделей. К недостаткам библиотеки относятся ее сложность для рядового пользователя, в том числе необходимость понимания основных механизмов настройки моделей с аддитивной регуляризацией в связи с отсутствием автоматической оптимизации гиперпараметров.

Эволюционный подход к оптимизации гиперпараметров моделей с аддитивной регуляризацией [19] включает в себя способ представления гиперпараметров в виде стратегии обучения, генетический алгоритм настройки и оценку качества. С целью получения более стабильной работы фреймворка внесены изменения в эволюционный подход.

В инициализацию индивидов добавлены базовые модели без сильной регуляризации, что отображено на рис. 1, *a*, где итерации обучения n_2 – n_4 заменены нулями, это позволяет улучшить скорость сходимости модели за счет коррекции направления оптимизации при инициализации сильно разреженных моделей. Остальные гиперпараметры, такие как количество

тем — B_n , декорреляция фоновых и основных тем матрицы Φ — D_Φ^B , D_Φ^S , сглаживание фоновых тем — P_Φ^B , P_Φ^S и разреживание основных тем P_Φ^S , P_Φ^S , сэмплировались в соответствии с правилами, определенными в [19].

Процедура мутации реализована в два шага (рис. 1, *b*). На шаге 1 происходит определение, будет ли произведена мутация индивида с вероятностью m_α , затем каждый элемент из соответствующей категории (которым соответствуют цвета стрелок) мутирует с вероятностью m_β , т. е. заменяется на случайный элемент из категории. На шаге 2 происходит замена параметров мутации с вероятностью m_β на случайное значение из равномерного распределения.

Приведем модифицированную функцию приспособленности, которая обеспечивает разреженность матрицы темы-документы в диапазоне 0,2–0,8, что способствует получению более разнообразных результатов:

$$\alpha(\text{mean}(\text{coh}_{50}) + \min(\text{coh}_{50})),$$

$$\alpha = \begin{cases} 1, & \text{если } 0,2 \leq Sp_\theta \leq 0,8 \\ 0,7, & \text{иначе} \end{cases},$$

где θ — матрица распределения вероятностей тем над документами; Sp_θ — разреженность матрицы θ .

Суррогатное моделирование для ускорения оптимизации [20]. Фреймворк реализует модуль работы с суррогатными моделями для уменьшения времени поиска решения для больших корпусов данных. В качестве наилучшей суррогатной модели функции фита использована модель на основе гауссовских процессов.

Описание разработанного фреймворка

При разработке фреймворка приняты во внимание следующие правила.

— **Простота использования.** Предлагаемый базовый пайплайн не требует дополнительной настройки со стороны пользователя, так как в нем уже определены все необходимые шаги и заданы гиперпараметры, обеспечивающие хорошее качество «в среднем» на наборах данных с различными ста-

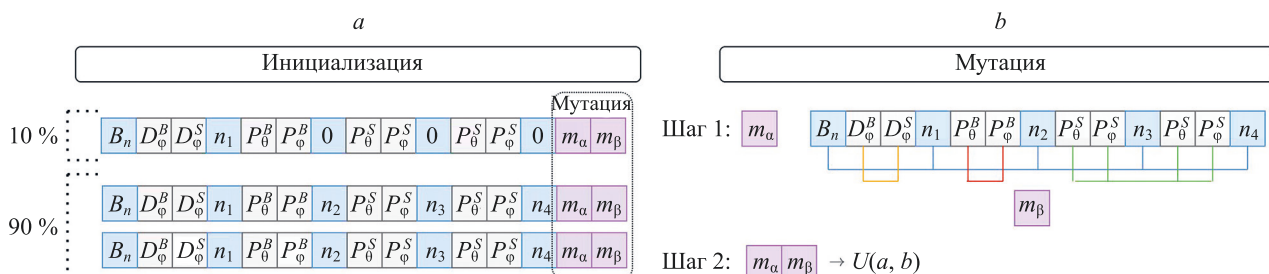


Рис. 1. Инициализация гиперпараметров, где обозначены 10 % базовых индивидов и 90 % индивидов со случайной инициализацией (*a*) и процесс мутации (*b*).

Ячейки хромосом обозначены: оптимизируемые целочисленные (синим цветом) и действительные (серым цветом) гиперпараметры; параметры мутации, изменяющиеся в рамках индивида (фиолетовым цветом)

Fig. 1. Hyperparameters initialization where 10 % of base individuals and 90 % of randomly initialized individuals (*a*) and mutation procedure (*b*) are provided. Blue color indicates integer hyperparameters, white — real ones, violet — mutation parameters that is self-learned with individual development

тистическими характеристиками. Задается только желаемое количество тем.

- *Реализация полного пайплайна, включая предобработку данных.* Часто библиотеки и фреймворки опускают процедуру подготовки данных, оставляя ее за пользователем.
- *Возможность расширения и реализации собственных методов.* В первую очередь предполагается модификация оценок качества.
- *Обеспечение скорости обработки текстов.* Фреймворк предполагает реализацию использования суррогатных моделей, которые позволяют повысить скорость оптимизации до 1,8 раз, существенно не влияя на качество [20]. Также исследователям будет предоставлена возможность проведения быстрых экспериментов при использовании распределенной версии фреймворка.

Фреймворк содержит набор модулей, реализующих общий пайплайн: предобработки данных (очистка, лемматизация, подготовка); оценки качества; оптимизации, содержащей реализации эволюционных алгоритмов; суррогатных моделей. При необходимости возможно расширение имеющихся модулей либо реализация новых.

Общий принцип работы фреймворка (рис. 2) можно описать следующим образом. На вход поступает корпус текстовых данных и производится его предобработка с целью очистки и нормализации текстовых данных, а

также подготовки данных в формате, требуемом для работы библиотеки BigARTM. Затем происходит определение метрики оценки качества целевой тематической модели (ТМ), которую можно либо оставить предложенной по умолчанию, либо выбрать из существующих. Задаются общие настройки для процедуры поиска оптимальных гиперпараметров (Гиперп), реализуемой библиотекой, такие как используемый эволюционный алгоритм, размер, число итераций и т. п. После этого начинается работа алгоритма поиска гиперпараметров, в течение которого происходят итерации отчета метрик качества для каждого индивида в популяции, представляющего набор параметров и стратегию их применения.

В случае, если используется оригинальная версия предложенного метода (т. е. без суррогатов) каждый набор параметров проходит через процедуру построения тематической модели по задаваемой стратегии (рис. 2, правая часть «Обучение тематической модели») путем обучения тематической модели с первым набором параметров в течение некоторого количества итераций, затем смены на второй набор в стратегии и продолжении обучения ранее созданной модели с новым набором, и т. д. до получения конечной модели. Отметим, что таких сменяющих друг друга наборов параметров может быть несколько или стратегия может состоять только из одного набора (так как значение количества итераций обучения может быть равно нулю).

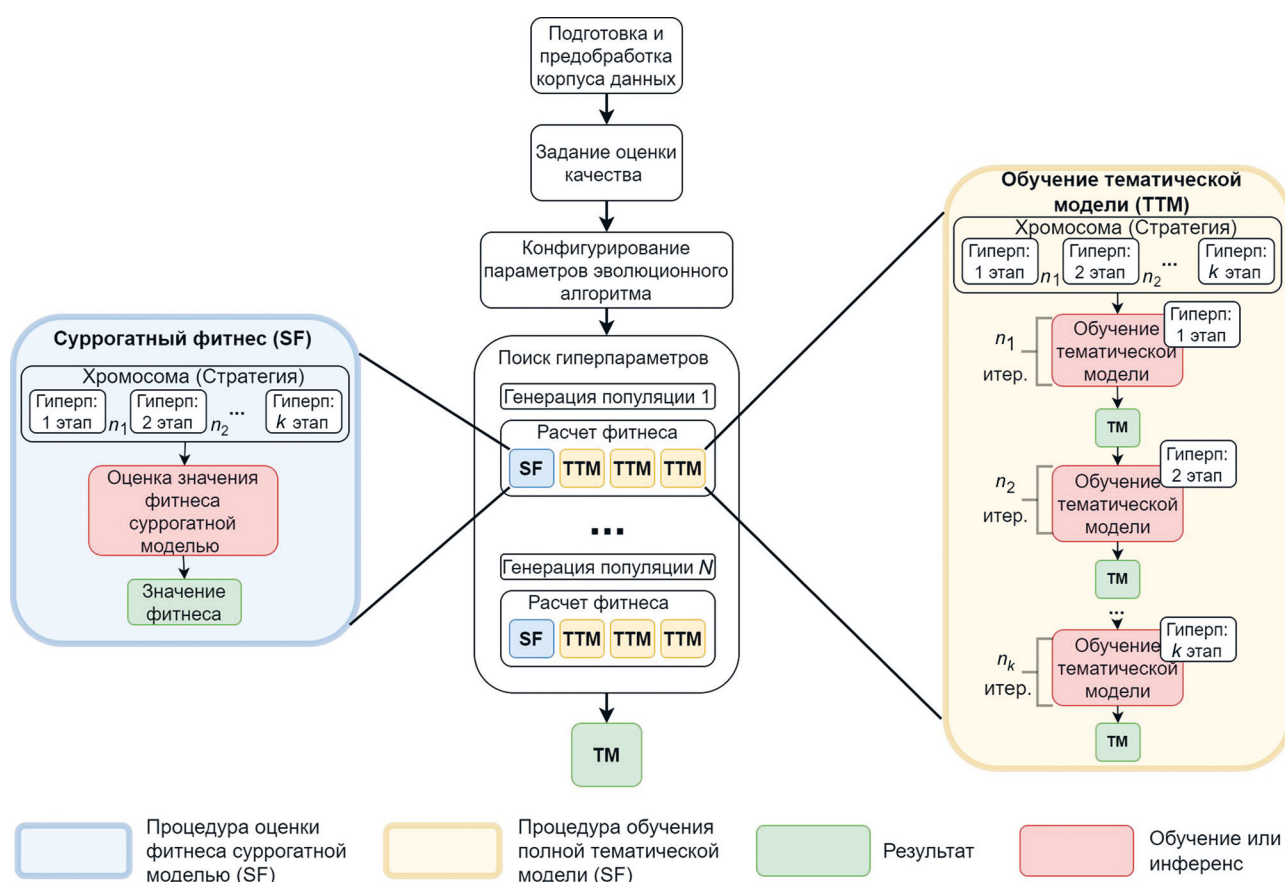


Рис. 2. Общая схема работы фреймворка

Fig. 2. The base scheme of the proposed framework

При использовании модификации метода с применением суррогатов, для некоторого подмножества индивидов в популяции вместо непосредственного вычисления метрики, описанного выше, производится ее прогнозирование с помощью суррогатной модели. Прогнозирование позволяет получить оценку метрики, используя только сами значения гиперпараметров и порядок в стратегии.

По окончании процесса поиска гиперпараметров, эволюционный алгоритм выдает наилучший найденный набор гиперпараметров и стратегию его применения. Фреймворк, в свою очередь, обучает итоговую тематическую модель и возвращает ее пользователю вместе с набором гиперпараметров. При необходимости пользователь может расширить функционал и добавить собственные методы расчета.

Эффективность использования разработанного фреймворка можно повысить за счет ускорения построения тематических моделей. Заметим, что популяционные подходы можно успешно распараллелить на этапе вычисления фитнеса для отдельных особей. В связи с этим предложено распределенное расширение фреймворка за счет включения возможности расчета популяции особей на наборе вычислительных узлов (рис. 3).

Принцип работы распределенной версии имеет несколько значимых отличий от базового варианта.

В данном случае расчеты производятся на наборе вычислительных узлов, где один из них ведущий, а остальные — расчетные.

Подготовленные данные на первом шаге обработки (Подготовка и предобработка корпуса данных) сохраняются в хранилище (Storage), доступном на всех узлах вычислительного кластера для последующего использования при расчетах метрики качества. На этапе поиска гиперпараметров текущая популяция эволюционного алгоритма рассылается на вычислительные узлы для подсчета фитнеса с помощью построения тематических моделей согласно стратегии и гиперпараметрам каждого из индивидов. Результаты оценки метрики качества затем передаются обратно на ведущий узел. Обученные модели, их метрики и логи могут сохраняться для всех индивидов или выборочно в хранилище результатов (MLFlow и Storage).

Если используется модификация метода с суррогатными моделями, прогнозирование метрики качества для фитнеса производится локально, без пересылки на удаленные узлы, так как данная процедура не занимает много времени из-за простоты суррогатных моделей. Однако в будущем такой порядок работы может быть изменен, если в этом появится необходимость ввиду роста вычислительной сложности суррогатных моделей.

Основные компоненты распределенного фреймворка (рис. 4): autotmlib-distr – интеграционная прослойка,

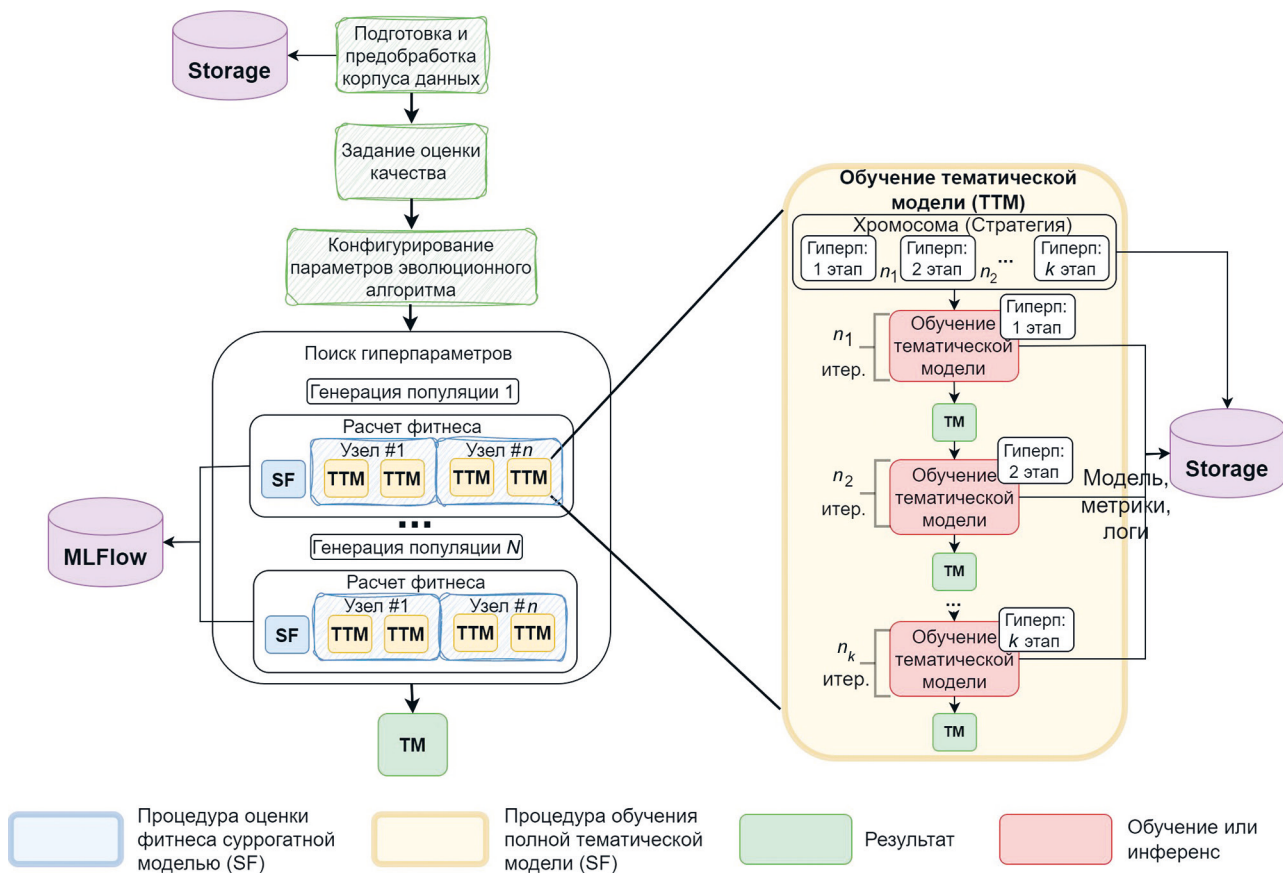


Рис. 3. Схема работы распределенной версии фреймворка.

Шаги предобработки и конфигурирования выполняются на ведущем узле, вычисление фитнеса — на расчетных узлах

Fig. 3. Workflow of the distributed framework version. The preprocessing and configuration steps are performed on the master node; the fitness calculation is performed on the calculation nodes

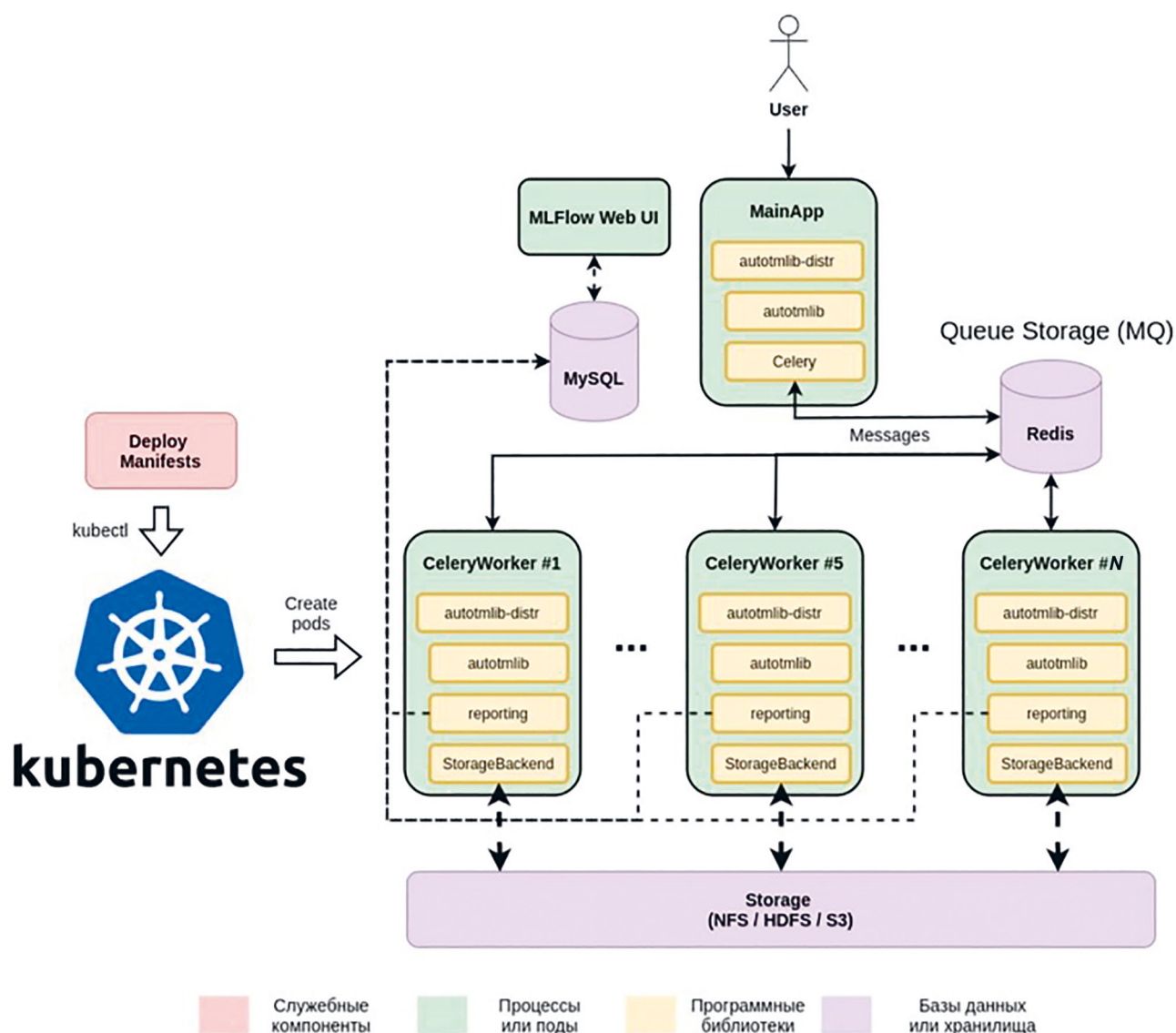


Рис. 4. Схема общей архитектуры фреймворка

Fig. 4. Scheme of the framework architecture

содержащая основной код фреймворка и представляющая вспомогательный материал репортинга и сохранения данных; Celery — брокер задач на обучении экземпляров тематических моделей для подбора гиперпараметров; Queue Storage — хранилище очереди задач на основе базы данных Redis для брокера Celery; MLFlow — система регистрации результатов расчетов и процессов оптимизации гиперпараметров; Storage — долговременное хранилище для промежуточных и конечных тематических моделей; менеджер ресурсов Kubernetes, обеспечивающий выделение ресурсов для обработчиков (CeleryWorker) брокера Celery; модуль манифестов Deploy manifests, отвечающий за развертывание и конфигурирование всех остальных компонентов фреймворка в кластерных средах под управлением Kubernetes.

Включение в архитектуру фреймворка компонентов Storage и MLFlow позволило удобно и эффективно провести регистрацию результатов вычислений фитнеса: построенные модели; полученные наборы тема-

тик; использованный для построения набор гиперпараметров и стратегию их применения (как хромосому); метрики качества; итоговое значение фитнеса; логи обучения тематической модели, и другую служебную информацию.

Таким образом, разработанный фреймворк содержит все необходимые компоненты для проведения оптимизации гиперпараметров тематических моделей, включая предобработку данных. Предложен распределенный вариант библиотеки для проведения экспериментальных исследований.

Исследование эффективности фреймворка

Для сравнения с существующими решениями определим основной критерий, по которому выполним сравнение фреймворков и библиотек — достижение высокого качества за ограниченное время оптимизации.

Для экспериментов выберем три набора данных с различными характеристиками: датасет Lentar, с

новостными постами, собранными за 20 лет; Amazon еда — отзывы о продукции, реализованные Amazon и 20 newsgroups — классический датасет с 20 категориями постов на разные темы. У каждого из датасетов отобран сэмпл в 10 000 документов.

Сравнение произведено с библиотеками genism (настройка алгоритма LDA), Octis (оптимизация гиперпараметров контекстуализированной модели (Contextualized TM), а именно, использовались веса предобученной модели RoBERTa для русского и английского языков) и topicNet (использовался предлагаемый базовый пайплайн). Для сравнения выбрана базовая версия фреймворка с использованием суррогатного моделирования фитнеса.

Когерентность (Coh) [21] и нормализованная мера попарной взаимной информации (NPMI) [22] вычислены для 20 наиболее вероятных токенов в каждой теме. Для расчета качества полученных моделей использованы оценки качества тем, имеющих доказанную корреляцию с человеческим восприятием (когерентность и NPMI). Вычислено качество классификации при использовании полученных представлений доку-

ментов (для датасета Amazon метки классов отсутствуют, поэтому качество классификации не замерялось). Для оценки использован метод k -ближайших соседей с 5 соседями и рассчитана средняя взвешенная $f1$ -мера на 5 фолдах (Cls (f1)).

Из таблицы видно, что разработанный фреймворк показал лучшие результаты даже при ограничении времени оптимизации. При этом наблюдается соблюдение баланса между качеством тем и эффективностью применения эмбедингов для прикладной задачи классификации. Отметим, что алгоритмы TopicNet и Gensim за выделенное в 3 мин время сходятся к стабильному решению, в то время как предложенный фреймворк еще может повысить качество до 10 % при обучении на большем количестве итераций.

На рис. 5 представлен пример визуализации с помощью метода t-SNE (стохастическое вложение соседей с t -распределением) получаемых тематических кластеров при обучении на 25 темах, где отображено их соответствие с метками классов датасета. Заметно, что существуют как тематически обособленные кластеры, так и близкие друг к другу темы.

Таблица. Полученное разными фреймворками качество при среднем времени оптимизации 3 мин. Результаты представлены в безразмерных единицах

Table. Quality obtained by different frameworks with average optimization time (3 minutes). The results are presented in dimensionless units

Фреймворк	Датасет							
	20 newsgroups (10 тем)			Lentaru (50 тем)			Amazon еда (25 тем)	
	Coh	NPMI	Cls (f1)	Coh	NPMI	Cls (f1)	Coh	NPMI
TopicNet	−2,50	0,03	0,05	−4,93	0,02	0,39	−2,34	−0,10
Gensim (LDA)	−3,20	0,01	0,36	−6,66	−0,05	0,48	−0,21	−0,09
Octis (Contextualized TM)	−2,65	−0,05	0,32	−4,56	0,06	0,47	−1,23	−0,23
Разработанный фреймворк	−2,30	0,01	0,37	−3,77	0,05	0,49	−4,95	−0,09

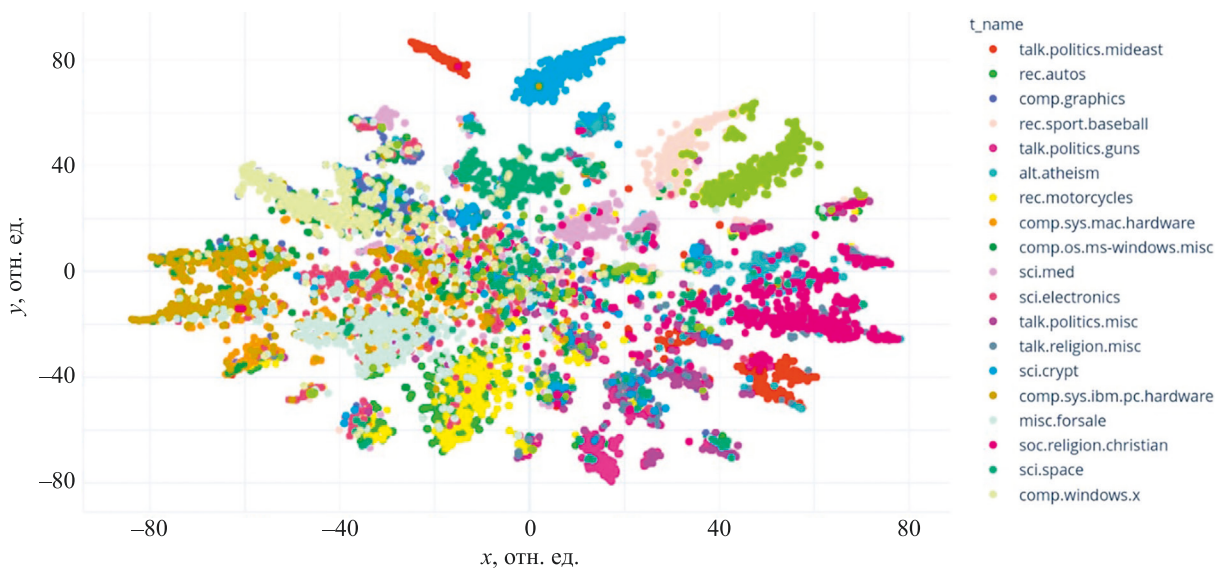


Рис. 5. Кластеризация представлений текстов датасета 20 newsgroups с помощью t-SNE

Fig. 5. Embeddings clustering for 20 newsgroups dataset with t-SNE

Заключение

Предложенный фреймворк автоматического подбора гиперпараметров для моделей с аддитивной регуляризацией позволил получить высококачественные модели за малое количество времени, что продемонстрировано результатами сравнения с аналогичными решениями. Благодаря этим особенностям он очень удобен для решения задач разведочного анализа, экономит время специалистов и позволяет проводить эксперименталь-

ные исследования в режиме распределенной обработки. Полученные эмбединги текстов можно использовать для решения разнообразных прикладных задач.

В дальнейшем планируется расширение фреймворка с помощью включения в него поддержки работы с различными модальностями и разработки схемы обработки данных с частичной разметкой во время обучения модели. Также возможно ускорение работы фреймворка за счет параллелизации на шаге предобработки данных.

Литература

1. Khanthaapha P., Pipanmaekaporn L., Kamonsantiroj S. Topic-based user profile model for POI recommendations // *Proc. of the 2nd International Conference on Intelligent Systems, Metaheuristics Swarm Intelligence*. 2018. P. 143–147. <https://doi.org/10.1145/3206185.3206203>
2. Peña F.J., O'Reilly-Morgan D., Tragos E.Z., Hurley N., Duriakova E., Smyth B., Lawlor A. Combining rating and review data by initializing latent factor models with topic models for Top-N recommendation // *Proc. of the 14th ACM Conference on Recommender Systems*. 2020. P. 438–443. <https://doi.org/10.1145/3383313.3412207>
3. Sokhin T., Butakov N. Semi-automatic sentiment analysis based on topic modeling // *Procedia Computer Science*. 2018. V. 136. P. 284–292. <https://doi.org/10.1016/j.procs.2018.08.286>
4. Nevezhin E., Butakov N., Khodorchenko M., Petrov M., Nasonov D. Topic-driven ensemble for online advertising generation // *Proc. of the 28th International Conference on Computational Linguistics*. 2020. P. 2273–2283. <https://doi.org/10.18653/v1/2020.coling-main.206>
5. Zamiralov A., Khodorchenko M., Nasonov D. Detection of housing and utility problems in districts through social media texts // *Procedia Computer Science*. 2020. V. 178. P. 213–223. <https://doi.org/10.1016/j.procs.2020.11.023>
6. Shi T., Kang K., Choo J., Reddy C.K. Short-text topic modeling via non-negative matrix factorization enriched with local word-context correlations // *Proc. of the World Wide Web Conference (WWW 2018)*. 2018. P. 1105–1114. <https://doi.org/10.1145/3178876.3186009>
7. Hofmann T. Probabilistic latent semantic indexing // *Proc. of the 22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '99)*. 1999. P. 50–57. <https://doi.org/10.1145/312624.312649>
8. Blei D.M., Ng A.Y., Jordan M.I. Latent Dirichlet allocation // *Journal of Machine Learning Research*. 2003. V. 3. P. 993–1022.
9. Vorontsov K., Potapenko A., Plavin A. Additive regularization of topic models for topic selection and sparse factorization // *Lecture Notes in Computer Science*. 2015. V. 9047. P. 193–202. https://doi.org/10.1007/978-3-319-17091-6_14
10. Card D., Tan C., Smith N.A. Neural models for documents with metadata // *Proc. of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. 2018. P. 2031–2040. <https://doi.org/10.18653/v1/p18-1189>
11. Cao Z., Li S., Liu Y., Li W., Ji H. A novel neural topic model and its supervised extension // *Proceedings of the AAAI Conference on Artificial Intelligence*. 2015. V. 29. N 1. P. 2210–2216. <https://doi.org/10.1609/aaai.v29i1.9499>
12. Bianchi F., Terragni S., Hovy D. Pre-training is a hot topic: Contextualized document embeddings improve topic coherence // *Proc. of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*. 2021. P. 759–766. <https://doi.org/10.18653/v1/2021.acl-short.96>
13. Ye J., Jing X., Li J. Sentiment Analysis Using Modified LDA // *Lecture Notes in Electrical Engineering*. 2018. V. 473. P. 205–212. https://doi.org/10.1007/978-981-10-7521-6_25
14. Bodrunova S., Koltsov S., Koltsova O., Nikolenko S., Shimorina A. Interval semi-supervised LDA: Classifying needles in a haystack // *Lecture Notes in Computer Science*. 2013. V. 8265. P. 265–274. https://doi.org/10.1007/978-3-642-45114-0_21

References

1. Khanthaapha P., Pipanmaekaporn L., Kamonsantiroj S. Topic-based user profile model for POI recommendations. *Proc. of the 2nd International Conference on Intelligent Systems, Metaheuristics Swarm Intelligence*, 2018, pp. 143–147. <https://doi.org/10.1145/3206185.3206203>
2. Peña F.J., O'Reilly-Morgan D., Tragos E.Z., Hurley N., Duriakova E., Smyth B., Lawlor A. Combining rating and review data by initializing latent factor models with topic models for top-n recommendation. *Proc. of the 14th ACM Conference on Recommender Systems*, 2020, pp. 438–443. <https://doi.org/10.1145/3383313.3412207>
3. Sokhin T., Butakov N. Semi-automatic sentiment analysis based on topic modeling. *Procedia Computer Science*, 2018, vol. 136, pp. 284–292. <https://doi.org/10.1016/j.procs.2018.08.286>
4. Nevezhin E., Butakov N., Khodorchenko M., Petrov M., Nasonov D. Topic-driven ensemble for online advertising generation. *Proc. of the 28th International Conference on Computational Linguistics*, 2020, pp. 2273–2283. <https://doi.org/10.18653/v1/2020.coling-main.206>
5. Zamiralov A., Khodorchenko M., Nasonov D. Detection of housing and utility problems in districts through social media texts. *Procedia Computer Science*, 2020, vol. 178, pp. 213–223. <https://doi.org/10.1016/j.procs.2020.11.023>
6. Shi T., Kang K., Choo J., Reddy C.K. Short-text topic modeling via non-negative matrix factorization enriched with local word-context correlations. *Proc. of the World Wide Web Conference (WWW 2018)*, 2018, pp. 1105–1114. <https://doi.org/10.1145/3178876.3186009>
7. Hofmann T. Probabilistic latent semantic indexing. *Proc. of the 22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '99)*, 1999, pp. 50–57. <https://doi.org/10.1145/312624.312649>
8. Blei D.M., Ng A.Y., Jordan M.I. Latent Dirichlet allocation. *Journal of Machine Learning Research*, 2003, vol. 3, pp. 993–1022.
9. Vorontsov K., Potapenko A., Plavin A. Additive regularization of topic models for topic selection and sparse factorization. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2015, vol. 9047, pp. 193–202. https://doi.org/10.1007/978-3-319-17091-6_14
10. Card D., Tan C., Smith N.A. Neural models for documents with metadata. *Proc. of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 2018, pp. 2031–2040. <https://doi.org/10.18653/v1/p18-1189>
11. Cao Z., Li S., Liu Y., Li W., Ji H. A novel neural topic model and its supervised extension. *Proceedings of the AAAI Conference on Artificial Intelligence*, 2015, vol. 29, no. 1, pp. 2210–2216. <https://doi.org/10.1609/aaai.v29i1.9499>
12. Bianchi F., Terragni S., Hovy D. Pre-training is a hot topic: Contextualized document embeddings improve topic coherence. *Proc. of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*, 2021, pp. 759–766. <https://doi.org/10.18653/v1/2021.acl-short.96>
13. Ye J., Jing X., Li J. Sentiment Analysis Using Modified LDA. *Lecture Notes in Electrical Engineering*, 2018, vol. 473, pp. 205–212. https://doi.org/10.1007/978-981-10-7521-6_25
14. Bodrunova S., Koltsov S., Koltsova O., Nikolenko S., Shimorina A. Interval semi-supervised LDA: Classifying needles in a haystack. *Lecture Notes in Computer Science*, 2013, vol. 8265, pp. 265–274. https://doi.org/10.1007/978-3-642-45114-0_21

15. Řehůřek R., Sojka P. Software framework for topic modelling with large corpora // *Proc. of the LREC 2010 Workshop on New Challenges for NLP*. 2010. P. 45–50.
16. Terragni S., Fersini E., Galuzzi B.G., Tropeano P., Candelieri A. OCTIS: Comparing and optimizing topic models is simple! // *Proc. of the 16th Conference of the European Chapter of the Association for Computational Linguistics: System Demonstrations*. 2021. P. 263–270. <https://doi.org/10.18653/v1/2021.eacl-demos.31>
17. Khodorchenko M., Butakov N. Developing an approach for lifestyle identification based on explicit and implicit features from social media // *Procedia Computer Science*. 2018. V. 136. P. 236–245. <https://doi.org/10.1016/j.procs.2018.08.262>
18. Vorontsov K., Frei O., Apishev M., Romov P., Dudarenko M. BigARTM: Open source library for regularized multimodal topic modeling of large collections // *Communications in Computer and Information Science*. 2015. V. 542. P. 370–381. https://doi.org/10.1007/978-3-319-26123-2_36
19. Khodorchenko M., Teryoshkin S., Sokhin T., Butakov N. Optimization of learning strategies for ARTM-based topic models // *Lecture Notes in Computer Science*. 2020. V. 12344. P. 284–296. https://doi.org/10.1007/978-3-030-61705-9_24
20. Khodorchenko M., Butakov N., Sokhin T., Teryoshkin S. Surrogate-based optimization of learning strategies for additively regularized topic models // *Logic Journal of the IGPL*. 2022. <https://doi.org/10.1093/jigpal/jzac019>
21. Röder M., Both A., Hinneburg A. Exploring the space of topic coherence measures // *Proc. of the Eighth ACM International Conference on Web Search and Data Mining (WSDM'15)*. 2015. P. 399–408. <https://doi.org/10.1145/2684822.2685324>
22. Newman D., Noh Y., Talley E., Karimi S., Baldwin T. Evaluating topic models for digital libraries // *Proc. of the 10th Annual Joint Conference on Digital Libraries (JCDL'10)*. 2010. P. 215–224. <https://doi.org/10.1145/1816123.1816156>
15. Řehůřek R., Sojka P. Software framework for topic modelling with large corpora. *Proc. of the LREC 2010 Workshop on New Challenges for NLP*, 2010, pp. 45–50.
16. Terragni S., Fersini E., Galuzzi B.G., Tropeano P., Candelieri A. OCTIS: Comparing and optimizing topic models is simple! *Proc. of the 16th Conference of the European Chapter of the Association for Computational Linguistics: System Demonstrations*, 2021, pp. 263–270. <https://doi.org/10.18653/v1/2021.eacl-demos.31>
17. Khodorchenko M., Butakov N. Developing an approach for lifestyle identification based on explicit and implicit features from social media. *Procedia Computer Science*, 2018, vol. 136, pp. 236–245. <https://doi.org/10.1016/j.procs.2018.08.262>
18. Vorontsov K., Frei O., Apishev M., Romov P., Dudarenko M. BigARTM: Open source library for regularized multimodal topic modeling of large collections. *Communications in Computer and Information Science*, 2015, vol. 542, pp. 370–381. https://doi.org/10.1007/978-3-319-26123-2_36
19. Khodorchenko M., Teryoshkin S., Sokhin T., Butakov N. Optimization of learning strategies for ARTM-based topic models. *Lecture Notes in Computer Science*, 2020, vol. 12344, pp. 284–296. https://doi.org/10.1007/978-3-030-61705-9_24
20. Khodorchenko M., Butakov N., Sokhin T., Teryoshkin S. Surrogate-based optimization of learning strategies for additively regularized topic models. *Logic Journal of the IGPL*, 2022. <https://doi.org/10.1093/jigpal/jzac019>
21. Röder M., Both A., Hinneburg A. Exploring the space of topic coherence measures. *Proc. of the Eighth ACM International Conference on Web Search and Data Mining (WSDM'15)*, 2015, pp. 399–408. <https://doi.org/10.1145/2684822.2685324>
22. Newman D., Noh Y., Talley E., Karimi S., Baldwin T. Evaluating topic models for digital libraries. *Proc. of the 10th Annual Joint Conference on Digital Libraries (JCDL'10)*, 2010, pp. 215–224. <https://doi.org/10.1145/1816123.1816156>

Авторы

Ходорченко Мария Андреевна — младший научный сотрудник, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, [sc 57207458742](https://orcid.org/0000-0001-5446-5311), <https://orcid.org/0000-0001-5446-5311>, mariyaxod@yandex.ru

Бутаков Николай Алексеевич — кандидат технических наук, старший научный сотрудник, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, [sc 56218218400](https://orcid.org/0000-0002-2705-1313), <https://orcid.org/0000-0002-2705-1313>, alipoov.nb@gmail.com

Насонов Денис Александрович — кандидат технических наук, старший научный сотрудник, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, [sc 56086498600](https://orcid.org/0000-0002-2740-0173), <https://orcid.org/0000-0002-2740-0173>, denis.nasonov@gmail.com

Фирулик Михаил Юрьевич — директор департамента, ООО «Оператор Газпром ИД», Санкт-Петербург, 191028, Российская Федерация, [sc 56086498600](https://orcid.org/0000-0001-7114-1052), <https://orcid.org/0000-0001-7114-1052>, firulik@gmail.com

Authors

Maria A. Khodorchenko — Junior Researcher, ITMO University, Saint Petersburg, 197101, Russian Federation, [sc 57207458742](https://orcid.org/0000-0001-5446-5311), <https://orcid.org/0000-0001-5446-5311>, mariyaxod@yandex.ru

Nikolay A. Butakov — PhD, Senior Researcher, ITMO University, Saint Petersburg, 197101, Russian Federation, [sc 56218218400](https://orcid.org/0000-0002-2705-1313), <https://orcid.org/0000-0002-2705-1313>, alipoov.nb@gmail.com

Denis A. Nasonov — PhD, Senior Researcher, ITMO University, Saint Petersburg, 197101, Russian Federation, [sc 56086498600](https://orcid.org/0000-0002-2740-0173), <https://orcid.org/0000-0002-2740-0173>, denis.nasonov@gmail.com

Mikhail Yu. Firulik — Director of Department, LLC “Operator Gazprom ID”, Saint Petersburg, 191028, Russian Federation, [sc 56086498600](https://orcid.org/0000-0001-7114-1052), <https://orcid.org/0000-0001-7114-1052>, firulik@gmail.com

Статья поступила в редакцию 20.10.2022

Одобрена после рецензирования 08.12.2022

Принята к печати 24.01.2023

Received 20.10.2022

Approved after reviewing 08.12.2022

Accepted 24.01.2023



Работа доступна по лицензии
Creative Commons
«Attribution-NonCommercial»

doi: 10.17586/2226-1494-2023-23-1-121-135

УДК 51-77: [519.865+330.16]

Ценностно-ориентированное моделирование принятия экономических решений в условиях нестационарности внешней среды

Валентина Юрьевна Гулева¹✉, Антон Николаевич Кованцев²,
Анатолий Георгиевич Суриков³, Пётр Владимирович Чунаев⁴,
Галина Владимировна Горнова⁵, Александр Валерьевич Бухановский⁶

^{1,2,3,4,5,6} Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация

¹ guleva@itmo.ru ✉, <https://orcid.org/0000-0002-1555-9371>

² ankovantcev@itmo.ru, <https://orcid.org/0000-0003-1765-7001>

³ anatoliy.surikov@itmo.ru, <https://orcid.org/0000-0003-4411-271X>

⁴ chunaev@itmo.ru, <https://orcid.org/0000-0001-8169-8436>

⁵ gornova@itmo.ru, <https://orcid.org/0000-0002-6615-0701>

⁶ avbukhanovskii@itmo.ru, <https://orcid.org/0000-0003-1588-8164>

Аннотация

Предмет исследования. Изменяющаяся среда формирует условия для изменения поведения людей, что в совокупности может приводить к кризисным ситуациям. В случае принятия экономических решений возникающая нестационарность динамики разных компонентов системы может представлять собой экономический кризис. Исследована возможность учета общечеловеческих ценностей при моделировании принятия решений в условиях изменения среды. Внутренние процессы агента, предшествующие принятию решения, отражены в концепции «убеждения-желания-намерения-действие-реакция» (Beliefs-Desires-Intentions-Actions-Reactions, BDIAR). **Метод.** Выполнен обзор закономерностей и существующих подходов к моделированию экономических решений и предложен новый, авторский подход. Выявлены механизмы влияния стресса на принятие решений, факторы ограничения рациональности и оценки рисков в контексте поведенческой экономики. Рассмотрены известные теории влияния ценностей на изменение структуры потребления в кризисной ситуации. Показаны способы учета эмоций в архитектуре агента. В предлагаемой модели ценности рассмотрены как социальный фактор принятия решений. В силу их субъективности они представлены математически как базис оценки объектов среды. Субъективность оценок объектов выбора отражена в функциях привлекательности объектов, функциях динамики состояния агента и субъективности влияния решения на удовлетворенность агента.

Основные результаты. Показана возможная модификация компонент модели агента с учетом влияния ценностей на динамику потребления. Предложен метод учета ценностей в архитектуре BDIAR при моделировании принятия решения агентом, где уровни архитектуры соответствуют ценностям, функциям предпочтения и динамике состояния агента. Проанализированы псевдонимизированные транзакционные данные по дебетовым картам клиентов банка-партнера «Санкт-Петербург» отдельно за 2017–2019 гг. и 2020 г. Продемонстрирована субъективность влияния среды в кризисной ситуации на динамику изменения ценностей и потребностей для разных групп потребителей с учетом кластеризации типов их поведения. Показаны различия динамики ценностей на индивидуальном уровне и уровне групп; подтверждено повышение приоритета ценностей выживания и продемонстрирована разная скорость возвращения к докризисному состоянию для различных поведенческих групп. **Практическая значимость.** Полученные результаты могут быть полезны для развития методов моделирования экономического поведения в условиях нестационарной внешней среды, в частности, в случае кризисов.

Ключевые слова

модель принятия решений, кризис, ценности, потребности, нестационарность, поведенческая экономика, BDIAR

Благодарности

Статья подготовлена по результатам работы, выполненной при поддержке Российского научного фонда (соглашение № 17–71–30029) и банка «Санкт-Петербург».

Ссылка для цитирования: Гулева В.Ю., Кованцев А.Н., Суриков А.Г., Чунаев П.В., Горнова Г.В., Бухановский А.В. Ценностно-ориентированное моделирование принятия экономических решений в условиях нестационарности внешней среды // Научно-технический вестник информационных технологий, механики и оптики. 2023. Т. 23, № 1. С. 121–135. doi: 10.17586/2226-1494-2023-23-1-121-135

Value-based modeling of economic decision making in conditions of unsteady environment

Valentina Yu. Guleva^{1✉}, Anton N. Kovatsev², Anatoly G. Surikov³, Petr V. Chunaev⁴, Galina V. Gornova⁵, Alexander V. Boukhanovskiy⁶

^{1,2,3,4,5,6} ITMO University, Saint Petersburg, 197101, Russian Federation

¹ guleva@itmo.ru✉, <https://orcid.org/0000-0002-1555-9371>

² ankovantcev@itmo.ru, <https://orcid.org/0000-0003-1765-7001>

³ anatoliy.surikov@itmo.ru, <https://orcid.org/0000-0003-4411-271X>

⁴ chunaev@itmo.ru, <https://orcid.org/0000-0001-8169-8436>

⁵ gornova@itmo.ru, <https://orcid.org/0000-0002-6615-0701>

⁶ avbukhanovskii@itmo.ru, <https://orcid.org/0000-0003-1588-8164>

Abstract

The changing environment creates the conditions for changing people's behavior which together can lead to crisis situations. In the case of making economic decisions, the emerging non-stationarity of the dynamics of different components of the system can represent an economic crisis. The possibility of universal human values consideration in modeling decision-making under conditions of changing environment has been studied. The internal processes of an agent prior to making a decision are reflected in the concept of Beliefs-Desires-Intentions-Actions-Reactions (BDIAR). The article reviews the regularities and existing approaches to modeling economic decisions, and proposes a new, author's approach. The mechanisms of influence of stress on decision-making, factors of rationality limitation, and risk assessment in the context of behavioral economics are revealed. The well-known theories of the influence of values on the change in the structure of consumption in a crisis situation are considered. Ways of taking into account emotions in the architecture of the agent are shown. In the proposed model, values are considered as a social factor in decision making. Due to their subjectivity, they are presented mathematically as the basis for assessing environmental objects. The subjectivity of the objects of choice assessments is reflected in the functions of attractiveness of the objects, the functions of the agent's state dynamics, and in the subjectivity of the decision influence on the satisfaction of the agent. A possible modification of the components of the agent model is shown, taking into account the influence of values on the dynamics of consumption. A method is proposed for taking into account values in the BDIAR architecture when modeling an agent's decision-making where the levels of the architecture correspond to values, preference functions, and functions of the agent's state dynamics. The pseudonymized transactional data on debit cards of the partner bank customers were analyzed separately for 2017–2019 and 2020. The subjectivity of the environment influence in a crisis situation on the dynamics of changes in values and needs for different consumer groups is demonstrated, taking into account the clustering of their behavior types. Differences in the dynamics of values at the individual level and the level of groups are shown; an increase in the priority of survival values and a different rate of return to the pre-crisis state for different behavioral groups are also shown. The results obtained are useful for developing methods for modeling economic behavior in a non-stationary external environment, in particular, in the case of crises.

Keywords

decision making model, crisis, values, needs, non-stationarity, behavioral economics, BDIAR

Acknowledgements

This research is financially supported by the Russian Science Foundation, Agreement No. 17–71–30029 with co-financing of the Bank “Saint Petersburg”.

For citation: Guleva V.Yu., Kovatsev A.N., Surikov A.G., Chunaev P.V., Gornova G.V., Boukhanovskiy A.V. Value-based modeling of economic decision making in conditions of unsteady environment. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 2023, vol. 23, no. 1, pp. 121–135 (in Russian). doi: 10.17586/2226-1494-2023-23-1-121-135

Введение

Изменения среды формируют условия для изменения поведения людей, что в совокупности может приводить к кризисным ситуациям. В случае экономических решений возникающая нестационарность динамики разных компонент системы может наблюдаться как экономический кризис. Экономические кризисы оказывают огромное влияние на многие процессы в обществе, его благосостояние и развитие, что заставляет не только экономистов, политиков и научных работни-

ков, но и людей многих других профессий исследовать возможности упреждения, прекращения и управления кризисными ситуациями [1, 2], в которые оказываются втянутыми отдельные люди, домохозяйства, фирмы, банки и государства. Одна из таких проблем — преобразование систем потребления и производства для обеспечения устойчивого развития [3]. Решения, принимаемые любым из участников рынка, могут быть эмоциональными, не отражающими реальной ситуации, или рациональными. При этом каждое решение влияет на экономическую ситуацию на других масшта-

бах: шаблоны потребления для отдельных магазинов и микроуровневые поведенческие модели влияют на рынок и макроэкономические переменные.

Прогнозирование на основе статистических моделей может обеспечивать надежные результаты при незначительном числе параметров — они не требуют понимания моделируемых процессов на разных уровнях и способны воспроизводить ожидаемую динамику при наличии репрезентативной выборки наблюдений. Тем не менее, развитие критических ситуаций характеризуется сменой стационарности наблюдений, что не позволяет использовать для прогноза модели, ранее полученные по данным. В силу открытости экономических систем и своеобразности отдельных кризисов, возможности для множественного анализа и разработки методов их предупреждения ограничены [3–5]. Таким образом, именно понимание основополагающих процессов экономического поведения является ключевым для оценки причин итоговых наблюдений, прогнозирования возможных последствий, а также избежания кризисов.

Отсутствие стационарности в данных свойственно не только кризисам, но и развивающимся ситуациям, часто встречающимся в экономике при возникновении новых финансовых инструментов, когда наблюдается динамика их принятия индивидами и численность агентов с одним поведением сменяется агентами с новым типом поведения. Например, динамика принятия карт безналичного расчета или карт лояльности сетей розничной торговли связана с социально-демографическими характеристиками потребителей [6], а постепенность и неодновременность смены их поведения отражается на неполноте транзакционных данных и возникновении степенных законов, часто встречаемых в экономических системах [7]. Поведенческая стационарность обычно объясняется существованием фиксированных ограничений, существующих достаточное время для адаптации к ним индивидов и возникновения некоторой нормы поведения. Новые технологии определяют новые варианты поведения, расширяя границы нормы, что приводит к комбинированному распределению вариантов поведения. Таким образом, наблюдения, связанные с возникновением новых технологий, также нестационарны, что влечет неполноту соответствующих данных и усложняет задачу построения прогностических моделей, что приводит к необходимости реконструкции данных на основании небольших наблюдаемых компонент [8] и необходимости разработки моделей, объясняющих механизмы изменения динамики в условиях нестационарной среды.

Обработка экземпляров данных как индивидуальных решений позволяет объяснить причины принимаемого решения на уровне отдельных индивидов и извлечь параметры поведения для рассматриваемого человека, что позволяет более гибко использовать доступные данные, однако требует объяснения субъективности решений и различий их ценности для различных людей. Зависимость решения от субъекта, его текущего состояния в контексте изменяющейся среды и наблюдаемой информации о ней обеспечивает переход от функции полезности [8, 9], определяющей объектив-

ную полезность решения, к субъективной привлекательности [10] каждого решения как функции объекта, субъекта (актора) и окружения, что означает влияние внутренних свойств индивида, таких как эмоции и социальные ценности, которые находят отражение в идеях поведенческой экономики.

В рамках работы рассмотрены основные теории, затрагивающие особенности принятия решений в кризисных ситуациях, на основе которых строится архитектура модели агента принятия экономических решений с учетом ценностей и потребностей агента. Проведенный анализ транзакционных данных демонстрирует смену приоритетов платежного поведения в результате развития кризисной ситуации, а классификация платежных категорий по группам ценностей выживания, социализации и саморазвития демонстрирует возрастающую роль ценностей выживания.

Стресс и кризис как частный случай нестационарности среды — ограничения при принятии решений

Изменения среды, демонстрирующие нестационарность статистических свойств, приводят к активизации адаптационных процессов со стороны агентов, взаимодействующих со средой. При этом в контексте принятия решений формируются ограничения возможного выбора, что создает стрессовую/кризисную ситуацию. Поиск моделей принятия экономических решений в кризис раскрывает большое число результатов наблюдений за изменением поведения в стрессовых ситуациях. Среди них наблюдения за динамикой во время и после эпидемий [11, 12]. При планировании расходов наблюдается упрощение жизни, снижение трат в категориях роскоши и повышение расходов в категориях еды, образования и здравоохранения [11]. При этом общая динамика продаж в сетях розничной торговли претерпевает спад с последующим подъемом при сокращении паники. Среди факторов панического поведения отмечается значимость социального контекста [13] и новостей [12]. Индивидуальная адаптивность и эффективность в условиях стресса связывается соответственно с временем и восприимчивостью [14]. Йилмаз и Кафадар (2022) используют нейрофизиологические методы для исследования влияния стресса на процесс принятия решений, а также предпринимают попытку связать это с типологией личности [15]. В ходе эксперимента создана стрессовая ситуация, соматический ответ организма на которую измеряется на поверхности кожи с помощью специальных тестов, при этом оценивается влияние стресса на принятие решений. Эксперимент в симуляторе управления полетами показал, что навык принятия решений в условиях стресса коррелировал со страхом принятия ошибочных решений [16]. При этом тренировки способствовали выработке навыка принятия решения в условиях стресса и снижению страха перед ошибками.

Исследования основаны на анализе данных наблюдений и опросов и ограничены информационными моделями.

Поведенческая экономика

Экономическое поведение — набор решений, связанных с персональными финансовыми потоками и распределением экономических ресурсов с целью извлечения выгоды из финансового оборота. Наиболее доступные данные для наблюдения за финансовым поведением — данные сетей магазинов розничной торговли, статистика по картам лояльности и транзакционные данные, отражающие время, суммы и место совершения транзакции. Они отражают изменение поведения и позволяют строить гипотезы о роли ценностей.

Принятие решений в экономике, как и в других областях, использующих методы агентного моделирования, естественным образом принимает за основу рациональные предположения, отражающие явные преимущества и недостатки принимаемого решения [17]. Таким образом, каждый вариант решения характеризуется функцией полезности [8, 9], где ожидаемая польза или экономические потери измеряются конкретными денежными величинами, а итоговый выбор решения им пропорционален.

Поведенческая экономика [18] строит теории, исходя из отказа от рационального агента, в стремлении отразить персональные предпочтения в отказе от незначительной выгоды сейчас, для получения большего в будущем. Аналогичные соображения лежат в основе социальных ценностей, также учитываемых в рамках поведенческой экономики. Эксперимент [19], когда студенты осуществляли выбор — распределить деньги поровну в соответствии с принципами социальной справедливости или взять себе больше в соответствии с функцией полезности, не поддается интерпретации рациональными экономическими теориями и демонстрирует значимость влияния социальных ценностей на принятие экономических решений. Отнесение отдельными людьми ресторанов к категориям еды или развлечений отражает как ценности, так и благосостояние, и характеризует ценность отдельных объектов для удовлетворения различных классов потребностей. Таким образом, функция полезности для оценки текущих индивидуальных предпочтений также имеет субъективный характер.

Субъективность персональных ценностей подтверждается рядом исследований когнитивных психологов и поведенческих экономистов. Теория перспектив [20] исследует субъективность восприятия информации в условиях риска, что влияет на итоговое решение. При этом субъективность состояла в склонности к недооценке и переоцениванию различных вероятностей альтернатив, а предметом исследования выступали когнитивные механизмы оценки потенциального выигрыша [21]. В работе [22] исследована теория распределения внимания ввиду механизмов распределения ресурсов между отделами мозга. Влияние ограничений когнитивных процессов на рациональное принятие решений изучено в рамках динамической модели агента на макроэкономическом примере динамики цен [23]. Предположено, что ограничения внимания способствуют возникновению дефицита информации, препятствующему принятию действительно рациональ-

ных решений [24], что приводит к моделям динамики потребления в условиях ограничений на обработку информации [25].

Таким образом, поведенческая экономика фокусируется на влиянии когнитивных, социальных и эмоциональных факторов на принятие экономических решений, которые, в свою очередь, влияют на глобальные переменные, такие как цены и распределение ресурсов.

Ценности как социальный фактор принятия решений

Ценности, культура, этика — части социальной составляющей поведения, нацеленные на выработку стратегий сотрудничества в интересах популяции. Социальные ценности и оптимальные стратегии социального поведения — результат многовековой эволюции [26, 27], что приводит к многочисленным воздействиям на индивидуальное поведение, начиная с копирования отдельных решений, эффектов передачи знаний среди населения и до культурных норм, основанных на народной мудрости, на эффективных стратегиях поведения.

Ценностные модели направлены на объяснение и классификацию системы приоритетов индивидов или групп людей. Пирамида потребностей [28] отражает приоритет потребностей выживания и безопасности перед эстетическими потребностями, как бы демонстрируя их большую ценность и невозможность развития потребностей высшего уровня без приоритетных потребностей выживания.

В работе [29], отражающей стили жизни, выделены классы выживших (survivors), поддерживающих (sustainers), владельцев (belongers), подражающих (emulators), успешных (achievers), «Я — это я» (I-am-me), экспериментальных (experiential), социально сознательных (societally conscious) и комбинированных (integrated). Частично с пирамидой потребностей [28] связан перечень ценностей [30]: самоуважение, веселье, азарт, безопасность, взаимоотношения, чувство долга, самореализация, чувство принадлежности, уважаемость. Шкала ценностей [31], отражающая уровень консерватизма и интроверсии, классифицирует ценности универсализма, благожелательности, конформности, власти, безопасности, достижений, гедонизма и самостоятельности между группами «сохранения — открытости изменениям» и «самоутверждения — заботы о людях и природе».

В специфичных случаях может быть целесообразным рассмотрение предметно-ориентированных ценностей, которые, впрочем, неизбежно связываются с общечеловеческими ценностями. Например, в [32] выделены профессиональные ценности: компетенции, справедливость, уважение, честность, самостоятельность, ответственность.

Легко видеть, что все группы ценностей вводят некоторый субъективный базис, позволяющий оценить значимость поступков и решений для отдельного человека [33]. Кроме того, первоочередная цель данных исследований — развитие поведенческих теорий. Соотнесение фактов, событий и поступков с ценност-

ным базисом индивида позволяет делать предположения о механизмах удовлетворенности и психологического баланса в зависимости от рассматриваемого контекста [34]. Методы измерения ценностей по шкалам [28, 29] основаны на социально-психологических опросниках, а также могут быть автоматизированы на уровне анализа текстов с применением методов тематического моделирования [35].

Исследованию воздействия ценностей на социальную и культурную жизнь посвящен международный проект, объединяющий социологов по всему миру, — «Всемирный обзор ценностей» [36]. В его рамках проведено 7 раундов исследований 97 стран мира, включая 90 % совершеннолетнего населения, начиная с 1981 года, утверждена методология разработки национальных скриптов интервью и анкет, и стандарт подготовки данных, согласно которому минимальная группа составляет тысячу человек от каждой из стран. Культуры народов Африки, Латинской Америки, Азии, мусульмане и православные размещены в двумерном пространстве вдоль осей «традиционных — секулярно-рациональных» ценностей и ценностей «выживания — самовыражения». Динамика изменения культурных ценностей связывается с ростом экзистенциальной безопасности и ощущением уверенности в будущем [37]. Таким образом, при моделировании на уровне одного агента целесообразно предположение о влиянии окружающей ситуации: безопасности и знаний о будущем и текущих ресурсах — на механику смены приоритетов в ценностном базисе.

Потребности, приоритеты, удовлетворенность и механизмы исчисления ценностей как рациональная основа принятия решений

Объективный мотиватор агентов при принятии решений — потребности. В отличие от ценностей, они динамичны и отражают текущее состояние агента. Ценности являются результатом социализации, следовательно, более стационарны и характеризуют правильность поведения. Пирамида потребностей характеризует их первостепенность: отражает физические потребности и потребности к безопасности, иллюстрирует необходимость их удовлетворения для возникновения высших потребностей — социализации и саморазвития. В случае с финансовым поведением потребности соответствуют наблюдаемым в транзакциях категориям потребления: продуктовые магазины, рестораны, театры и концертные залы, развлечения, путешествия, аптеки. В то же время траты в рамках каждой категории могут быть направлены на удовлетворение потребностей каждой из групп приоритета.

Динамика возникновения потребности в рамках отдельных категорий расходов соответствует динамике дефицита. Тогда платежное поведение в рамках каждой из категорий способствует реализации потребности. Например, существует возможность закупить еды на какое-то время вперед, и тогда можно дольше не ходить в магазин, но чем ближе к моменту истечения запасов, тем выше потребность повторного пополнения. Такая динамика может быть смоделирована уравне-

нием для конденсатора или осциллятора, поскольку учитывает внешнее воздействие и изменение энергии. Предположение об аналогичной динамике возникает для совокупности потребностей как их упорядоченной иерархии. Другими словами, удовлетворенность агента в целом может определяться уровнем иерархии, на котором он находится, в то время как эта общая удовлетворенность складывается из уровня удовлетворенности отдельных потребностей в рамках каждого из уровней иерархии потребностей. Сложность моделирования возникает ввиду необходимости учета взаимосвязи потребностей одного уровня для перехода к потребностям более высокого уровня.

Ценности позволяют определять влияние одного и того же события на удовлетворенность агента. Если зафиксировать ценности как базис оценки пользователем отдельных событий в векторном пространстве, то можно говорить о ценностной позитивности и негативности выбора на основе коллинеарности векторов и знака их скалярного произведения, на основе чего может быть оценена удовлетворенность отдельными решениями. Тогда от отдельного решения к траектории в ценностном пространстве можно перейти к суммированию последовательности событий, что при проекции на отдельные оси ценностей даст агрегированную картину удовлетворенности и позволит оценить вклад отдельных событий. Траектории в ценностном пространстве индивида позволяют оценить соответствие поведения ценностным установкам, степень стабильности поведения и уровень удовлетворенности.

Аналогично теории ограниченной рациональности [38], где функция принятия решения скорректирована с учетом субъективности восприятия, ценностный базис (заданный для отдельного индивида) корректирует модель принятия решения с учетом субъективности оценки решений и модель изменения состояния индивида с учетом субъективной удовлетворенности.

Структура модели агента Belief-Desire-Intention-Action-Reaction (BDIAR)

Принятие решений — динамический процесс [39], обусловленный физиологическими, психологическими и социальными факторами [40]. В исследовании динамики принятия решений в процессе корректировки режима питания эти факторы, среди которых фигурируют и личные ценности, рассматриваются в роли мотиваторов [40]. Динамика переживания агентом кризисных ситуаций, как правило, отделяет момент обновления информации и состояний агента и среды от совершения действий в ответ на предполагаемые или полученные изменения [41]. Ввиду этого в рамках работы определена потенциальная роль ценностей и потребностей при принятии решений.

BDI-модели (убеждения (Belief), желания (Desire), намерения (Intention)) представляют собой класс агентных моделей, ориентированных на отражение внутреннего мира агента в ходе принятия решений [42–44] — процесс планирования отделяется от исполнения. В предположении моделирования агента в контексте открытой системы, где каждому действию агента/приня-

тому решению может быть сопоставлен ответ среды и изменения состояний агентов, аббревиатура BDI может быть расширена до BDIAR. При этом действия агентов (Action) и их последствия (Reaction) — необходимая составляющая любой модели агента и соответствуют процессу исполнения запланированного. Аббревиатура соответствует состояниям агента и переходам между ними: убеждения — информации агента о мире, не обязательно истинной; желания — мотивационному состоянию агента, целям, выбранным к исполнению; намерения — планам, принятым к исполнению; действия — фактам их совершения; реакции — событиям, последствиям действий или изменениям среды, влияющим на состояние агента и его убеждения. Таким образом, отражается степень готовности агента к действию и связь с его «свойствами»/состоянием.

Учет эмоций и взаимодействия агента со средой при принятии решений встречается в научных работах как расширение BDI-модели предварительной эмоциональной компонентой, осуществляющей обработку сигналов от среды [45]. При этом предполагается обновление всех компонент модели с учетом эмоционального восприятия. В других архитектурах эмоциональных агентных моделей в разной степени уделяется внимание роли темперамента, личности, физиологических процессов и влиянию на динамику на разных уровнях BDI-модели агента [46–50]. Похожий подход в рамках поведенческой экономики рассматривается в работе [13], где вклад осведомленности и намерений в итоговое действие рассматривается наравне с относительной ценностью денег и предпринимается попытка их учета при объяснении парадокса Алле [51], демонстрирующего, что агент максимизирует не прибыль, а надежность.

Место теорий поведенческой экономики и ценностных моделей в архитектуре экономического агента

Произвольная модель динамики системы содержит в качестве основных компонент элементы системы, зависящие от состояний элементов макросостояния системы и набор механизмов изменения состояний элементов системы (рис. 1). В частности, изменение состояний элементов системы может происходить при их взаимодействии между собой или со средой. Таким образом, для некоторого начального состояния системы могут существовать устойчивые состояния, достигаемые спустя некоторое время, которые могут меняться за некоторое время при смене механизмов изменения состояний агентов или изменении среды.

В контексте агентной модели выбор отдельных агентов определяется функцией привлекательности выбираемых элементов среды (действий) $\Phi(e, u)$, а результат взаимодействия — выражением для динамики состояния агента (ds/dt) . Функция привлекательности реализует вероятность выбора каждого из вариантов и в случае рациональной теории принятия решений будет соответствовать функции полезности. В примере выбора контрагента в системе межбанковского кредитования рациональное решение соответствует привлекательности значений капитала C_u и процентной ставки R_u . При этом для учета субъективной склонности к риску добавляется параметр ценностных предпочтений $\varepsilon_v \in [0; 1]$: $\Phi(v, u) \sim \varepsilon_v C_u + (1 - \varepsilon_v) R_u$. Стратегии выбора могут определяться динамикой обработки информации, повторяемостью используемых стратегий, числом объектов выбора, когнитивными затратами на принятие решения при наличии ограничений на каче-

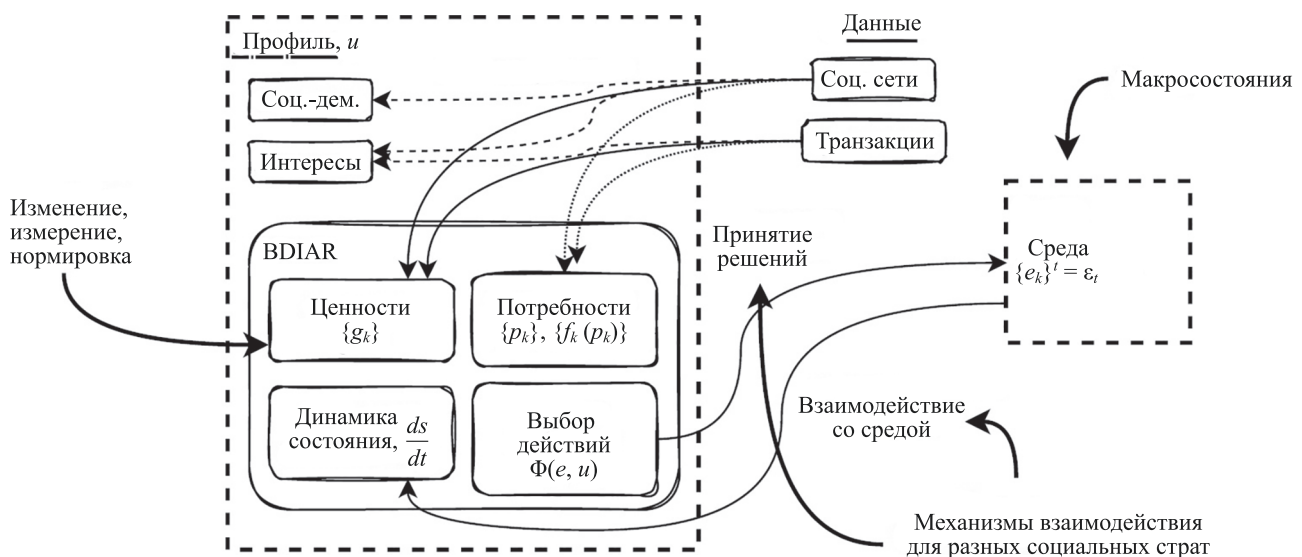


Рис. 1. Модель взаимодействия агента u и среды ε в момент t . Отражены блоки профиля интересов и социально-демографических характеристик, модели выбора действий $\Phi(e, u)$ и динамики состояний при выборе элементов, блоки ценностей и потребностей агента для субъективизации оценок. Параметры механизмов взаимодействий могут отличаться для различных страт (социальных групп)

Fig. 1. The model of interaction between agent u and environment ε at t moment. The blocks of user interests, socio-demographic features, choice model $\Phi(e, u)$ of environment elements, state dynamics model, agent values and needs for subjectivity reflection are displayed. Parameters of interaction functions are different for social strata

ство принимаемого решения и временные ресурсы [52]. Моделирование агента на основе ценностно-ориентированных принципов и теорий поведенческой экономики является определяющим для наблюдаемой динамики на макроуровне.

Помимо функций выбора и динамики состояния модель агента должна учитывать потребности и ценности агента в привычном процессе принятия решений. Теории поведенческой экономики свидетельствуют в пользу нерациональности или частичной рациональности решения и влияния когнитивных, эмоциональных и социальных процессов. В качестве социальной компоненты выступают ценности, а когнитивные процессы связаны с ограничениями восприятия.

Сопоставим уровни BDI-модели агента необходимым компонентам динамики системы из рис. 1 и ценностным ориентациям. Ориентация на ценности и потребности при принятии решений и выборе действий отражена в модели BDIAR состояний агентов. Модель демонстрирует не только состояние агента, функцию привлекательности и модель смены состояния агента в результате действий, но и включает ограничения на убеждения, что соответствует ценностям (табл. 1).

Для каждого агента моделируется динамика потребления, наблюдаемая, например, посредством транзакционной активности в различных категориях. Выбор категории расходов моделируется с учетом потребностей агентов на основании функции привлекательности. Таким образом, на макросостояние системы влияет состояние пользователя в рамках модели потребностей, которое определяется возможностями среды обеспе-

чения потребностей, что также связано с механизмом обеспечения потребностей (уравнение динамики состояния агента). При этом на макроуровне имеет значение перераспределение спроса в условиях ограничения предложения по ряду потребительских категорий.

Дополним табл. 1 учетом субъективности восприятия. Рассмотрим набор из N акторов u с состояниями $\{s^k\}_{k=1}^N$, где s^k — состояние k -го актора, описываемое вектором в линейном пространстве. Акторы взаимодействуют со средой $\epsilon = \{e^k\}_{k=1}^M$, где M — число объектов выбора. Макросостояние системы S описывается некоторой функцией от состояний акторов системы.

Каждый объект выбора e^i обладает некоторой ценностью для актора с состоянием s^k , так что выбор объекта пользователем осуществляется на основании функции привлекательности $\Phi(e, u)$, которая в случае с ценностно-ориентированными моделями связана с ценностями и потребностями агента. При этом ценность объекта выбора также влияет на динамику изменения состояния агента ds/dt . Таким образом, в контексте ценностно-ориентированных моделей агента компонентам динамической системы агентов, отвечающим за выбор объектов среды и влияние объектов среды на агентов, сопоставляются функции, учитывающие специфику предметной области (табл. 2).

Так, ценности представляют собой базис для оценки события, а динамика состояний агента учитывает проекции события на базис ценностей $e_k \cos(e_k, v)$. Таким образом, функции динамики агента учитывают изменение состояния с учетом ценностей и выбор объекта с учетом потребностей.

Таблица 1. Сопоставление уровней архитектуры BDIAR уровням модели агента с учетом ценностей и компонентам математической модели

Table 1. The correspondence between BDIAR agent architecture and agent model with values to mathematical components of an arbitrary agent model

Уровни BDIAR	Компоненты, которые необходимо учесть	Обозначение в модели	
Убеждения	Ценности	Ограничения (базис)	$\{g_k\}$
Желания	Потребности	Состояние и параметры агента	$s, \{p_i\}$
Намерения	План	Функция привлекательности объекта выбора	$\Phi(e_k)$
Действия	Действие	Вероятность выбора объекта	$P(e_k)$
Реакция	Результат	Модель смены состояния агента	ds/dt

Таблица 2. Адаптация модели рационального выбора агента в динамической системе к случаю субъективного выбора с учетом ценностей

Table 2. Adaptation of the rational choice model of an agent in a dynamic system to the case of subjective choice, taking into account values

Рациональный выбор		Субъективный выбор	
Агент, u	Среда, ε	Агент	Среда
Состояние s	Объекты выбора $\{e_k\} = \varepsilon$	Состояние s	Объекты выбора $\{e_k\} = \varepsilon$
Параметры		Параметры + ценности v	
Функция привлекательности $\Phi(e_k)$		$\Phi(e_k, u)$	
Влияние выбора для всех одинаково: $\frac{ds}{dt} \sim f(u, e)$		Проекция на оси ценностей для оценки значимости объекта выбора: $\frac{ds}{dt} \sim f(u, e \cos(e, v))$	

Анализ ценностной траектории на основе транзакционных данных

Рассмотрим данные о транзакциях по дебетовым картам клиентов банка «Санкт-Петербург» отдельно за 2017–2019 гг. и 2020 г. Платежные категории, стандартно отмеченные в транзакционных данных кодами MCC (Merchant Category Code — «код категории продавца»), относящими место совершения платежа к одной из категорий, приведем в соответствие с тремя базовыми ценностями: «выживание», «социализация», «саморазвитие» (пример сопоставления по ряду категорий в табл. 3).

Для этого трехуровневая модель ценностей сопоставлена с группами потребностей пирамиды Маслоу (рис. 2), а платежные категории (MCC) разделены по группам потребительских интересов. При измерении реализации потребностей через суммы транзакций, совершенных в соответствующей категории, динамика интереса к отдельным ценностям может быть оценена через динамику трат в соответствующих категориях как для отдельных клиентов (рис. 3), так и для их набора (рис. 4). Для оценки категорий трат на каждый день вычислим нормированные суммарные расходы по всем картам с разделением по группам:

Таблица 3. Сопоставление «действия»–«ценности» для трех базовых потребительских ценностей «выживание», «социализация», «саморазвитие» и действий, соответствующих категориям трат по дебетовым картам

Table 3. “Actions”–“values” map for the three basic consumer values “survival”, “socialization”, “self-development” and actions corresponding to the spending categories on debit cards

МСС-код	Группа потребительских интересов	Ценность
5411: 'Бакалейные магазины, супермаркеты'	Пища	«Выживание»
5812: 'Места общественного питания, рестораны'		
5462: 'Булочные', 5441: 'Кондитерские'		
5691: 'Магазины мужской и женской одежды'	Одежда	
5651: 'Одежда для всей семьи', 5661: 'Обувные магазины'		
5949: 'Магазины ткани, ниток, рукоделия, шитья'		
5211: 'Лесо- и строительный материал'	Жилище	
4900: 'Жилищно-коммунальные услуги'		
5714: 'Ткани, обивочный материал, гардины и портьеры, жалюзи'		
5912: 'Аптеки', 8043: 'Оптика, оптические товары и очки'	Здоровье	
8062: 'Больницы', 8021: 'Стоматологи, ортодонты'		
8071: 'Стоматологические и медицинские лаборатории'		
4111: 'Пассажирские перевозки — пригородные и местные пригородные рейсы, включая паромы'	Транспорт	«Социализация»
4112: 'Пассажирские железнодорожные перевозки'		
5533: 'Автозапчасти и аксессуары'		
4814: 'Телекоммуникационные услуги'	Информация и связь	
5732: 'Продажа электронного оборудования'		
9402: 'Почтовые услуги — только государственные'		
6012: 'Финансовые учреждения — торговля и услуги'	Финансы	«Саморазвитие»
9311: 'Налоговые платежи'		
5943: 'Магазины офисных, школьных принадлежностей, канцтоваров'	Дети и образование	
8299: 'Образовательные услуги, нигде более не классифицированные'		
8220: 'Колледжи, университеты, профессиональные училища и техникумы'		
7941: 'Атлетические поля, коммерческие виды спорта, профессиональные спортивные клубы, промоутеры спорта'	Развлечения и от- дых	
5942: 5192: 'Книги, периодические издания и газеты'		
5992: 'Флористика' 5735: 'Магазины звукозаписи'		
7922: 'Театральные продюсеры (кроме кинофильмов), билетные агентства'		
0742: 'Ветеринарные услуги'		
7991: 'Туристические достопримечательности и выставки'		

$$v_{i,k}^N = \frac{v_{i,k}}{\max_{l=1,\dots,n} \left(\sum_{j=1}^m v_{l,j} \right)}; i = 1, \dots, n; k = 1, \dots, m, \quad (1)$$

где $v_{i,j}$ — сумма транзакции j в момент i ; m — число ценностей; n — число наблюдений в ряду. Тогда пропорциональное соотношение суммарных расходов по группам трат имеет вид

$$v_{i,k}^R = \frac{v_{i,k}}{\sum_{j=1}^m v_{i,j}}; i = 1, \dots, n; k = 1, \dots, m. \quad (2)$$

Расчеты по формулам (1)–(2) лежат в основе визуализации динамики трат по ценностям для отдельных людей и их групп (рис. 3, 4).

В модели BDIAR, предназначенной для описания процессов принятия решений и выполнения действий во внешней среде финансовым актором, убеждениям (Belief) соответствуют ценности финансового актора, а потребности/желания (Desire) отражены отдельными тратами по группам категорий МСС. В рамках разработанного метода ценности могут меняться при возникновении кризисных состояний внешней среды.



Рис. 2. Отображение трехуровневой модели ценностей (выживание, социализация, саморазвитие) на пирамиду потребностей Маслоу

Fig. 2. Mapping a three-level value model (survival, socialization, self-development) onto Maslow's pyramid of needs

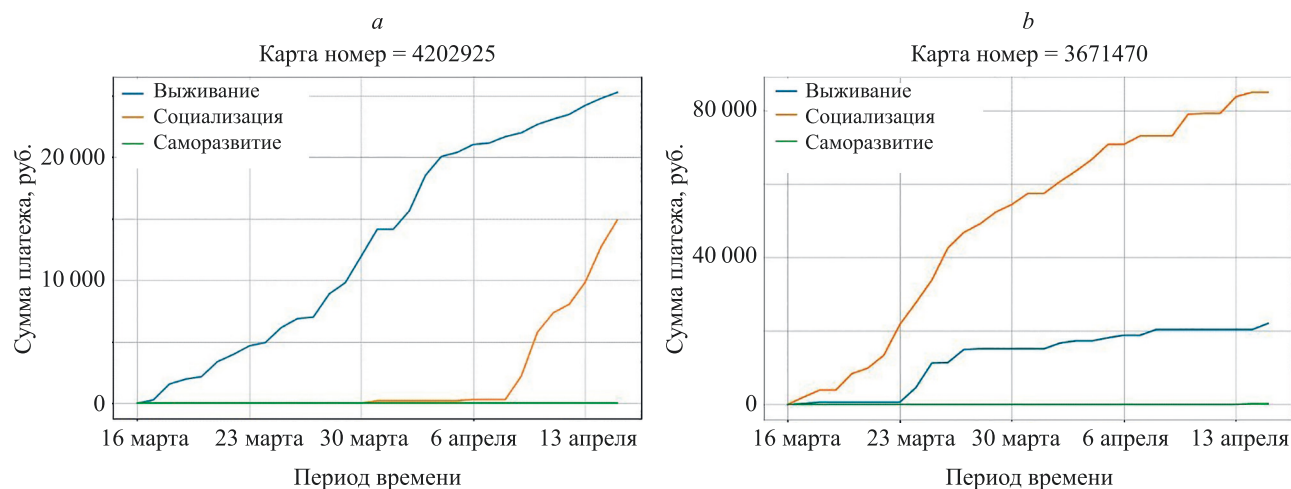


Рис. 3. Кумулятивный график трат для двух различных клиентов банка (в рамках одного месяца с 16 марта по 16 апреля 2020 года) для категорий, соответствующих различным базовым ценностям

Fig. 3. Cumulative spending schedule for two different bank customers (within one month from March 16 to April 16, 2020) for categories corresponding to different core values

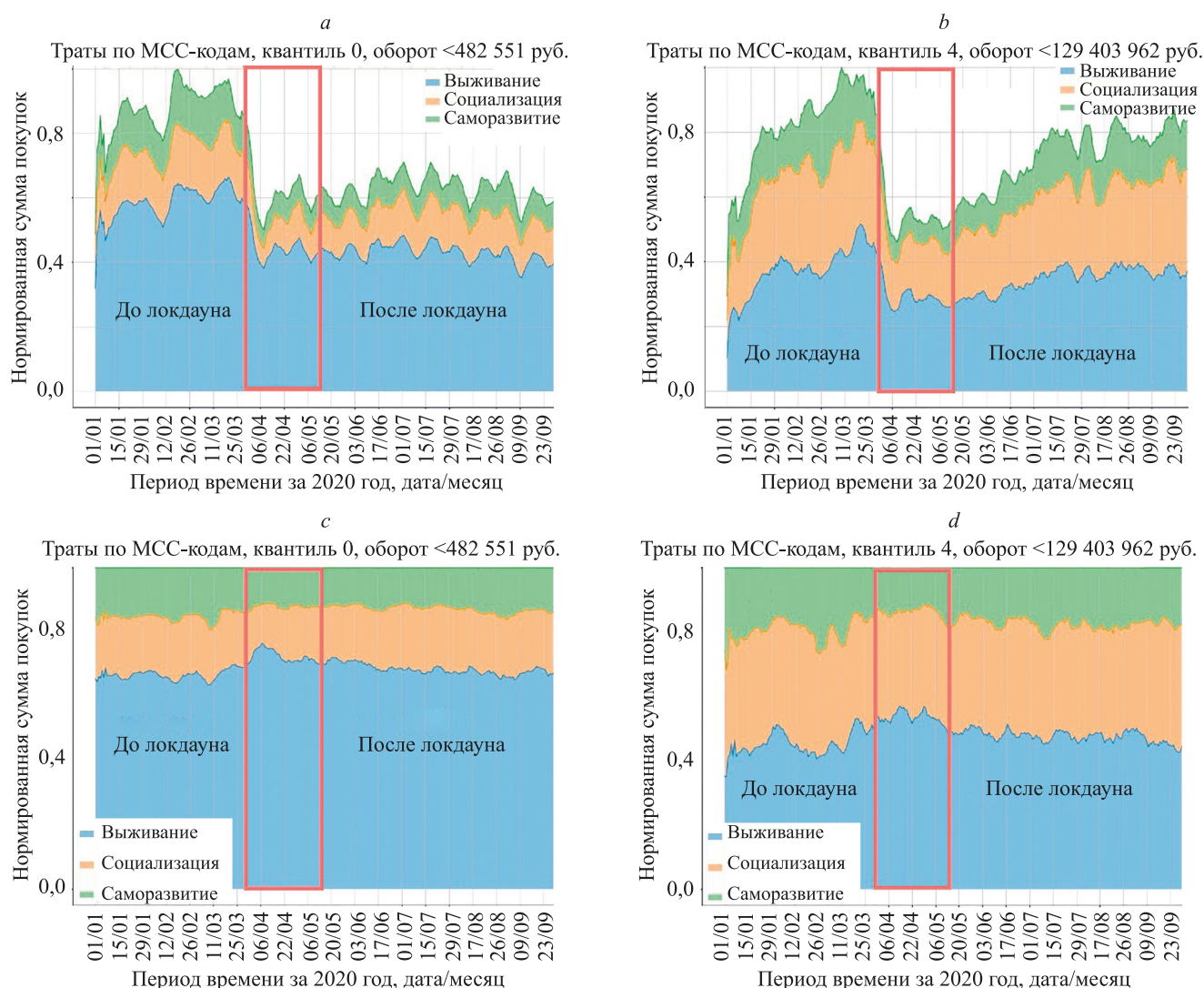


Рис. 4. Сравнение приоритета различных ценностей до, во время и после переходного процесса: у страты C1 (минимальный оборот) (a, c); у страты C5 (максимальный оборот) (b, d).

На рис. 4, a, b — абсолютные нормированные суммарные расходы по всем картам с разделением по выбранным ценностям (применено сглаживание методом скользящего среднего с окном, равным одной неделе); на рис. 4, c, d — пропорциональное соотношение суммарных расходов по выбранным ценностям

Fig. 4. Comparison of the priority of different values before, during and after the transition process: at stratum C1 (minimum turnover) (a, c); near stratum C5 (maximum turnover) (b, d).

Figures (a, b) show the absolute normalized total costs for all cards divided by selected values (smoothing was applied using the moving average method with a window equal to one week); (c, d) — the proportional ratio of the total costs for the selected values

Таким образом, модель позволяет описывать переживание кризиса (в том числе адаптацию) на уровне страт агентов. Кризис моделируется как изменение внешней среды (контекста), в частности, как изменение уровня доступности объектов для удовлетворения потребностей агентов, что может повлечь за собой изменение важности ценностей для страт агентов.

Идентификация BDIAR-модели для отдельных агентов затруднена вследствие разреженности данных и субъективности причин принимаемых решений, не позволяющих обучить отдельную модель агента для каждого прототипа. Исходя из этого, обучение BDIAR-моделей проведем на уровне групп (страт) финансовых акторов. Метод анализа включает: стратификацию финансовых акторов; расчет временных рядов изменения

приоритета потребностей/ценностей по отдельным стратам агентов на основании разметки (табл. 3); обучение параметрической модели изменения потребностей/ценностей для различных страт. Для идентификации модели переходного процесса на основе BDIAR время моделирования разделено на предкризисный, кризисный и посткризисный интервалы [T0, T1], [T1, T2], [T2, T3], где T0 и T3 — начало и конец интервала моделирования; T1 и T2 — начало и конец периода изменений внешней среды (контекста). Так, для моделирования действий карантинных ограничений на потребителей определено T1 = {30 марта 2020 г.}, T2 = {10 мая 2020 г.} на основе данных производственного календаря Российской Федерации (выделено красной рамкой на рис. 4). На рис. 4 видно: резкое снижение суммарно-

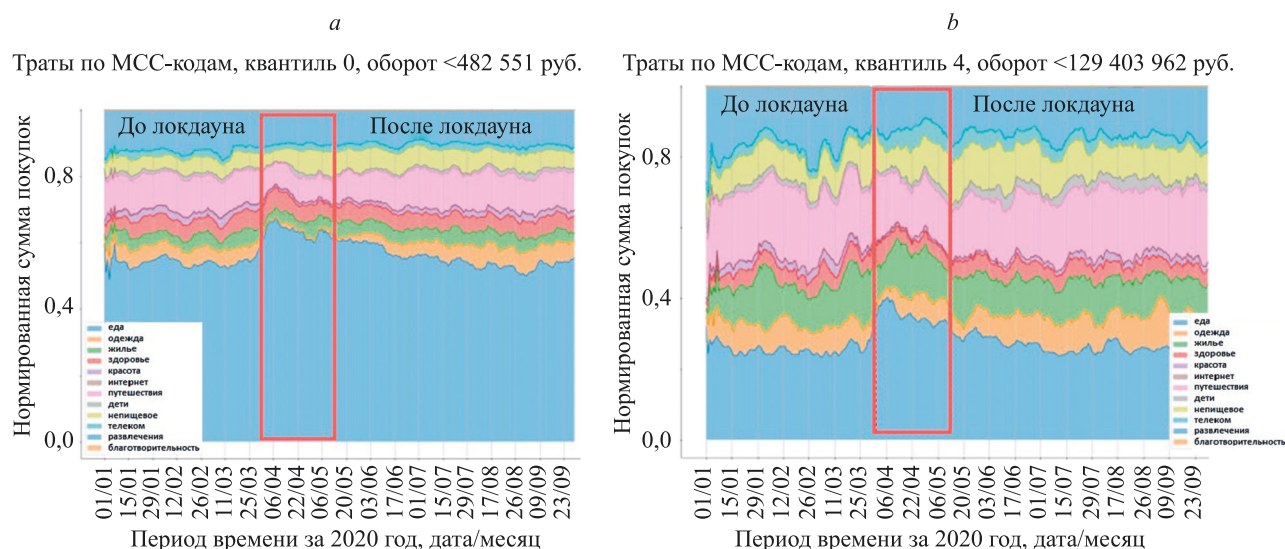


Рис. 5. Пропорциональное соотношение суммарных расходов, отражающих потребности клиентов банка в разные моменты времени: для страты 1 (минимальный оборот, C1) (a); для страты 5 (максимальный оборот, C5) (b).

Красной линией выделен интервал карантинных ограничений с 30 марта по 10 мая 2020 г.

Fig. 5. Proportional ratio of total expenses reflecting the needs of bank customers at different points in time: for stratum 1 (minimum turnover, C1) (a); for stratum 5 (maximum turnover, C5) (b)

The red line marks the interval of quarantine restrictions from March 30 to May 10, 2020

го количества трат; резкое увеличение важности ценности «выживание», с постепенным переходом в новое стационарное состояние для ситуации противоэпидемических мероприятий весны 2020 г., что соответствует ожидаемым закономерностям платежного поведения в кризис [11, 12]. С точки зрения BDIAR для страт финансовых акторов изменение внешней среды (снижение доходов, доступности, мобильности и др.) приводит к резкому изменению модели потребностей (что видно по изменению абсолютных сумм и долей трат в различных категориях на рис. 5), далее происходит постепенный переход к новому стабильному состоянию. Анализ данных показал субъективность влияния внешней среды на динамику ценностей, а следовательно, различие влияния принимаемого решения на динамику состояния агента. Выделение страт осуществлено методом кластеризации с учетом похожести рядов платежного поведения, причем один из признаков — среднемесячный оборот, а выбор страт C1 (минимальный оборот) и C5 (максимальный оборот) для визуализации обусловлен их существенными особенностями, обеспечивающими наглядность визуализации по сравнению с остальными стратами (рис. 4, 5).

Наблюдается качественно различные сценарии адаптации к кризису: обе страты сокращают абсолютный уровень трат в интервал ограничений («локдауна») приблизительно в два раза, однако клиенты страты с максимальным оборотом восстанавливают привычный уровень потребления уже через пять недель после кризиса, в то время как оборот по страте C1 не восстанавливается в течение всего периода наблюдений. В то же время и у C1, и у C5 наблюдается повышение приоритета ценности «выживание» в кризисное время, и ее стабилизация на более высоком уровне, чем до ограничений.

В результате наблюдается разная важность типов потребностей у страт для стабильного периода. Например, для периода «До локдауна» траты на продукты питания у C1 $> 0,5$, а у C5 $< 0,3$, у C5 в значительно большей степени удовлетворяются менее базовые потребности, такие как потребность в путешествиях, развлечениях, обустройстве дома. Отметим, что различные страты по-разному меняют потребности в кризисной ситуации: у C1 за счет возрастания расходов на продовольственные товары сокращаются траты на все категории потребностей, кроме категорий «здоровье» и «непродовольственные товары». Особенно снижены расходы на одежду, путешествия и уход за собой. Вследствие ограниченности ресурса происходит в первую очередь удовлетворение наиболее базовых потребностей, соответствующих ценности «выживание». У C5 отмечено снижение трат на развлечения и уход за собой, рост трат на категорию «телеком» и «продукты» и относительная стабильность остальных категорий. Таким образом, в случае избытка ресурса большее значение для страты приобретает факт недоступности ранее привлекательных объектов внешней среды (рестораны, кафе, бары, салоны красоты).

Заключение

Рассмотрена проблема моделирования экономического поведения агента в кризисной ситуации. Проведен обзор факторов влияния на поведение в условиях стресса, в том числе на платежное поведение. Рассмотрены теории поведенческой экономики, направленные на учет эмоциональных, когнитивных и социальных факторов, имеющих место в кризисной ситуации, где показано их влияние на поведение. Проведен обзор ценностных шкал как социальных ори-

ентиров при принятии решений. Проанализированы возможности BDI-архитектуры агента по учету эмоций. Выявлено место ценностно-ориентированных особенностей модели принятия решений в контексте BDI. Сигнатуры функций в модели динамики агента приведены в соответствие с ценностными моделями. В ходе анализа данных показана динамика смены ценностей агентов, динамика изменения потребностей, а также различие динамики в разных странах, что в совокупности предоставляет возможности исследования экономического поведения групп с разными ценностными установками. Смещение приоритетов людей в сторону ценностей выживания во время кризиса позволяет переопределить вектор ценностей и определить функции привлекательности объектов при принятии решений в соответствии с актуальным состоянием вектора ценностей. Таким образом, принятие решений может моделироваться в соответствии с фиксированными функциями привлекательности, и, предположительно,

моделирование динамики поведения в нестационарных условиях может быть сведено к моделированию динамики векторов ценностей как отдельных людей, так и страт. Разработка метода моделирования динамики ценностей, как и методов их идентификации, является предметом дальнейшего исследования, а проведенный обзор научных работ, анализ данных и их сопоставление в рамках единой архитектуры агента предоставляет основу для предполагаемого вида модели принятия решений и ее необходимых компонент.

Для дальнейшего моделирования экономического поведения агента в кризисной ситуации выявлена необходимость моделирования снижения приоритета эстетических потребностей и повышение приоритета потребностей выживания и безопасности. Выявлена необходимость моделирования динамики роста потребности и субъективности восприятия актуальной потребности с учетом ограничения информации, внимания и когнитивных особенностей агента.

Литература

1. Поспелов И. Моделирование российской экономики в условиях кризиса // Вопросы экономики. 2009. № 11. С. 50–75. <https://doi.org/10.32609/0042-8736-2009-11-50-75>
2. Hawkins D. Some conditions of macroeconomic stability // *Econometrica*. 1948. V. 16. N 3. P. 309–322. <https://doi.org/10.2307/1909272>
3. Bengtsson M., Alfredsson E., Cohen M., Lorek S., Schroeder P. Transforming systems of consumption and production for achieving the sustainable development goals: moving beyond efficiency // *Sustainability Science*. 2018. V. 13. N 6. P. 1533–1547. <https://doi.org/10.1007/s11625-018-0582-1>
4. Ashraf S., Félix E.G.S., Serrasqueiro Z. Do traditional financial distress prediction models predict the early warning signs of financial distress? // *Journal of Risk and Financial Management*. 2019. V. 12. N 2. P. 55. <https://doi.org/10.3390/jrfm12020055>
5. Edison H.J. Do indicators of financial crises work? An evaluation of an early warning system // *International Journal of Finance & Economics*. 2003. V. 8. N 1. P. 11–53. <https://doi.org/10.1002/ijfe.197>
6. Demoulin N.T.M., Zidda P. Drivers of customers' adoption and adoption timing of a new loyalty card in the grocery retail market // *Journal of Retailing*. 2009. V. 85. N 3. P. 391–405. <https://doi.org/10.1016/j.jretai.2009.05.007>
7. Gabaix X. Power Laws in Economics and Finance // *Annual Review of Economics*. 2009. V. 1. N 1. P. 255–294. <https://doi.org/10.1146/annurev.economics.050708.142940>
8. Squartini T., Caldarelli G., Cimini G., Gabrielli A., Garlaschelli D. Reconstruction methods for networks: the case of economic and financial systems // *Physics Reports*. 2018. V. 757. P. 1–47. <https://doi.org/10.1016/j.physrep.2018.06.008>
9. Houthakker H.S. Revealed preference and the utility function // *Economica*. 1950. V. 17. N 66. P. 159–174. <https://doi.org/10.2307/2549382>
10. Kalman P.J. Theory of consumer behavior when prices enter the utility function // *Econometrica*. 1968. V. 36. N 3-4. P. 497–510. <https://doi.org/10.2307/1909519>
11. Mehta S., Saxena T., Purohit N. The new consumer behaviour paradigm amid COVID-19: permanent or transient? // *Journal of Health Management*. 2020. V. 22. N 2. P. 291–301. <https://doi.org/10.1177/0972063420940834>
12. Loxton M., Truskett R., Scarf B., Sindone L., Baldry G., Zhao Y. Consumer behaviour during crises: Preliminary research on how coronavirus has manifested consumer panic buying, herd mentality, changing discretionary spending and the role of the media in influencing behaviour // *Journal of Risk and Financial Management*. 2020. V. 13. N 8. P. 166. <https://doi.org/10.3390/jrfm13080166>

References

1. Pospelov I. Modeling russian economy in the crisis period. *Voprosy Ekonomiki*, 2009, no. 11, pp. 50–75. (in Russian). <https://doi.org/10.32609/0042-8736-2009-11-50-75>
2. Hawkins D. Some conditions of macroeconomic stability. *Econometrica*, 1948, vol. 16, no. 3, pp. 309–322. <https://doi.org/10.2307/1909272>
3. Bengtsson M., Alfredsson E., Cohen M., Lorek S., Schroeder P. Transforming systems of consumption and production for achieving the sustainable development goals: moving beyond efficiency. *Sustainability Science*, 2018, vol. 13, no. 6, pp. 1533–1547. <https://doi.org/10.1007/s11625-018-0582-1>
4. Ashraf S., Félix E.G.S., Serrasqueiro Z. Do traditional financial distress prediction models predict the early warning signs of financial distress? *Journal of Risk and Financial Management*, 2019, vol. 12, no. 2, pp. 55. <https://doi.org/10.3390/jrfm12020055>
5. Edison H.J. Do indicators of financial crises work? An evaluation of an early warning system. *International Journal of Finance & Economics*, 2003, vol. 8, no. 1, pp. 11–53. <https://doi.org/10.1002/ijfe.197>
6. Demoulin N.T.M., Zidda P. Drivers of customers' adoption and adoption timing of a new loyalty card in the grocery retail market. *Journal of Retailing*, 2009, vol. 85, no. 3, pp. 391–405. <https://doi.org/10.1016/j.jretai.2009.05.007>
7. Gabaix X. Power Laws in Economics and Finance. *Annual Review of Economics*, 2009, vol. 1, no. 1, pp. 255–294. <https://doi.org/10.1146/annurev.economics.050708.142940>
8. Squartini T., Caldarelli G., Cimini G., Gabrielli A., Garlaschelli D. Reconstruction methods for networks: the case of economic and financial systems. *Physics Reports*, 2018, vol. 757, pp. 1–47. <https://doi.org/10.1016/j.physrep.2018.06.008>
9. Houthakker H.S. Revealed preference and the utility function. *Economica*, 1950, vol. 17, no. 66, pp. 159–174. <https://doi.org/10.2307/2549382>
10. Kalman P.J. Theory of consumer behavior when prices enter the utility function. *Econometrica*, 1968, vol. 36, no. 3-4, pp. 497–510. <https://doi.org/10.2307/1909519>
11. Mehta S., Saxena T., Purohit N. The new consumer behaviour paradigm amid COVID-19: permanent or transient? *Journal of Health Management*, 2020, vol. 22, no. 2, pp. 291–301. <https://doi.org/10.1177/0972063420940834>
12. Loxton M., Truskett R., Scarf B., Sindone L., Baldry G., Zhao Y. Consumer behaviour during crises: Preliminary research on how coronavirus has manifested consumer panic buying, herd mentality, changing discretionary spending and the role of the media in influencing behavior. *Journal of Risk and Financial Management*, 2020, vol. 13, no. 8, pp. 166. <https://doi.org/10.3390/jrfm13080166>

13. Jeffrey H.J., Putman A.O. The irrationality illusion: A new paradigm for economics and behavioral economics // *Journal of Behavioral Finance*. 2013. V. 14. N 3. P. 161–194. <https://doi.org/10.1080/15427560.2013.790388>
14. Poole K.B. *Adaptability and Decision Making under Stress in the Workplace*: a thesis submitted in partial fulfillment of the requirements for the degree of Master of Arts in Psychology. Middle Tennessee State University, 2018. 66 p.
15. Yilmaz S., Kafadar H. Decision-making under stress: Executive functions, analytical intelligence, somatic markers, and personality traits in young adults // *Applied Neuropsychology: Adult*. 2022. <https://doi.org/10.1080/23279095.2022.2122829>
16. Pope J., Hein M., Russell M.A., Burkholder Ch. A Continuation of research: Student decision making under stress in a flight control center simulation // *Proc. of the 20th International Symposium on Aviation Psychology (ISAP)*. 2019. P. 91–96.
17. Coleman J.S., Farraro Th.J. *Rational Choice Theory: Advocacy and Critique*. SAGE Publications, 1992. 232 p.
18. Поведенческая экономическая теория // Большая российская энциклопедия [Электронный ресурс]. URL: <https://old.bigenc.ru/economics/text/3147419> (дата обращения: 19.10.2022).
19. Kahneman D., Knetsch J.L., Thaler R.H. Fairness and the assumptions of economics // *The Journal of Business*. 1986. V. 59. N 4. P. S285. <https://doi.org/10.1086/296367>
20. Kahneman D., Tversky A. Prospect theory: An analysis of decision under risk // *Handbook of the fundamentals of financial decision making: Part I. World Scientific*. 2013. P. 99–127. https://doi.org/10.1142/9789814417358_0006
21. Tversky A., Kahneman D. Judgment under uncertainty: Heuristics and biases // *Science*. 1974. V. 185. N 4157. P. 1124–1131. <https://doi.org/10.1126/science.185.4157.1124>
22. Egeth H., Kahneman D. Attention and effort // *The American Journal of Psychology*. 1975. V. 88. N 2. P. 339–340. <https://doi.org/10.2307/1421603>
23. Maćkowiak B., Matějka F., Wiederholt M. Dynamic rational inattention: Analytical results // *Journal of Economic Theory*. 2018. V. 176. P. 650–692. <https://doi.org/10.1016/j.jet.2018.05.001>
24. Sims C.A. Implications of rational inattention // *Journal of Monetary Economics*. 2003. V. 50. N 3. P. 665–690. [https://doi.org/10.1016/s0304-3932\(03\)00029-1](https://doi.org/10.1016/s0304-3932(03)00029-1)
25. Luo Y. Consumption dynamics under information processing constraints // *Review of Economic Dynamics*. 2008. V. 11. N 2. P. 366–385. <https://doi.org/10.1016/j.red.2007.07.003>
26. Bergstrom T.C. Evolution of social behavior: Individual and group selection // *Journal of Economic Perspectives*. 2002. V. 16. N 2. P. 67–88. <https://doi.org/10.1257/0895330027265>
27. Taborsky B. The Evolution of Social Behaviour // *Ethology*. 2021. V. 127. N 10. P. 751–757. <https://doi.org/10.1111/eth.13212>
28. Maslow A.H. Preface to motivation theory // *Psychosomatic Medicine*. 1943. V. 5. N 1. P. 85–92. <https://doi.org/10.1097/00006842-194301000-00012>
29. Mitchell A. *The Nine American Lifestyles: Who We are and where We're Going*. Scribner Book Company, 1983. 302 p.
30. Kahle L.R. *Social Values and Social Change: Adaptation to Life in America*. Praeger Publishers, 1983. 324 p.
31. Schwartz S.H. An overview of the Schwartz theory of basic values // *Online Readings in Psychology and Culture*. 2012. V. 2. N 1. <https://doi.org/10.9707/2307-0919.1116>
32. Boroel B., Aramburo V., Gonzalez M. Development of a scale to measure attitudes toward professional values: An analysis of dimensionality using Rasch measurement // *Procedia — Social and Behavioral Sciences*. 2017. V. 237. P. 292–298. <https://doi.org/10.1016/j.sbspro.2017.02.079>
33. Gawel J.E. Herzberg's theory of motivation and Maslow's hierarchy of needs // *Practical Assessment, Research, and Evaluation*. 1996. V. 5. P. 11. <https://doi.org/10.7275/31qy-ea53>
34. Lester D., Hvezda J., Sullivan Sh., Plourde R. Maslow's hierarchy of needs and psychological health // *The Journal of General Psychology*. 1983. V. 109. N 1. P. 83–85. <https://doi.org/10.1080/00221309.1983.9711513>
35. Wilson S.R., Mihalcea R., Boyd R.L., Pennebaker J.W. Cultural influences on the measurement of personal values through words // *AAAI Spring Symposia*. 2016.
36. Inglehart R. et al. *World Values Survey*. 2005.
13. Jeffrey H.J., Putman A.O. The irrationality illusion: A new paradigm for economics and behavioral economics. *Journal of Behavioral Finance*, 2013, vol. 14, no. 3, pp. 161–194. <https://doi.org/10.1080/15427560.2013.790388>
14. Poole K.B. *Adaptability and Decision Making under Stress in the Workplace*. A thesis submitted in partial fulfillment of the requirements for the degree of Master of Arts in Psychology. Middle Tennessee State University, 2018. 66 p.
15. Yilmaz S., Kafadar H. Decision-making under stress: Executive functions, analytical intelligence, somatic markers, and personality traits in young adults. *Applied Neuropsychology: Adult*, 2022. <https://doi.org/10.1080/23279095.2022.2122829>
16. Pope J., Hein M., Russell M.A., Burkholder Ch. A Continuation of research: Student decision making under stress in a flight control center simulation. *Proc. of the 20th International Symposium on Aviation Psychology (ISAP)*, 2019, pp. 91–96.
17. Coleman J.S., Farraro Th.J. *Rational Choice Theory: Advocacy and Critique*. SAGE Publications, 1992, 232 p.
18. Behavioral Economics. *Great Russian Encyclopedia*. Available at: <https://old.bigenc.ru/economics/text/3147419> (accessed: 19.10.2022). (in Russian)
19. Kahneman D., Knetsch J.L., Thaler R.H. Fairness and the assumptions of economics. *The Journal of Business*, 1986, vol. 59, no. 4, pp. S285. <https://doi.org/10.1086/296367>
20. Kahneman D., Tversky A. Prospect theory: An analysis of decision under risk. *Handbook of the fundamentals of financial decision making: Part I. World Scientific*, 2013, pp. 99–127. https://doi.org/10.1142/9789814417358_0006
21. Tversky A., Kahneman D. Judgment under uncertainty: Heuristics and biases. *Science*, 1974, vol. 185, no. 4157, pp. 1124–1131. <https://doi.org/10.1126/science.185.4157.1124>
22. Egeth H., Kahneman D. Attention and effort. *The American Journal of Psychology*, 1975, vol. 88, no. 2, pp. 339–340. <https://doi.org/10.2307/1421603>
23. Maćkowiak B., Matějka F., Wiederholt M. Dynamic rational inattention: Analytical results. *Journal of Economic Theory*, 2018, vol. 176, pp. 650–692. <https://doi.org/10.1016/j.jet.2018.05.001>
24. Sims C.A. Implications of rational inattention. *Journal of Monetary Economics*, 2003, vol. 50, no. 3, pp. 665–690. [https://doi.org/10.1016/s0304-3932\(03\)00029-1](https://doi.org/10.1016/s0304-3932(03)00029-1)
25. Luo Y. Consumption dynamics under information processing constraints. *Review of Economic Dynamics*, 2008, vol. 11, no. 2, pp. 366–385. <https://doi.org/10.1016/j.red.2007.07.003>
26. Bergstrom T.C. Evolution of social behavior: Individual and group selection. *Journal of Economic Perspectives*, 2002, vol. 16, no. 2, pp. 67–88. <https://doi.org/10.1257/0895330027265>
27. Taborsky B. The Evolution of Social Behaviour. *Ethology*, 2021, vol. 127, no. 10, pp. 751–757. <https://doi.org/10.1111/eth.13212>
28. Maslow A.H. Preface to motivation theory. *Psychosomatic Medicine*, 1943, vol. 5, no. 1, pp. 85–92. <https://doi.org/10.1097/00006842-194301000-00012>
29. Mitchell A. *The Nine American Lifestyles: Who We are and where We're Going*. Scribner Book Company, 1983, 302 p.
30. Kahle L.R. *Social Values and Social Change: Adaptation to Life in America*. Praeger Publishers, 1983, 324 p.
31. Schwartz S.H. An overview of the Schwartz theory of basic values. *Online Readings in Psychology and Culture*, 2012, vol. 2, no. 1. <https://doi.org/10.9707/2307-0919.1116>
32. Boroel B., Aramburo V., Gonzalez M. Development of a scale to measure attitudes toward professional values: An analysis of dimensionality using Rasch measurement. *Procedia — Social and Behavioral Sciences*, 2017, vol. 237, pp. 292–298. <https://doi.org/10.1016/j.sbspro.2017.02.079>
33. Gawel J.E. Herzberg's theory of motivation and Maslow's hierarchy of needs. *Practical Assessment, Research, and Evaluation*, 1996, vol. 5, pp. 11. <https://doi.org/10.7275/31qy-ea53>
34. Lester D., Hvezda J., Sullivan Sh., Plourde R. Maslow's hierarchy of needs and psychological health. *The Journal of General Psychology*, 1983, vol. 109, no. 1, pp. 83–85. <https://doi.org/10.1080/00221309.1983.9711513>
35. Wilson S.R., Mihalcea R., Boyd R.L., Pennebaker J.W. Cultural influences on the measurement of personal values through words. *AAAI Spring Symposia*, 2016.
36. Inglehart R. et al. *World Values Survey*. 2005.

37. Инглхарт Р. Постмодерн: меняющиеся ценности и изменяющиеся общества // Полис. Политические исследования. 1997. № 4. С. 6–32.
38. Maćkowiak B., Matějka F., Wiederholt M. Rational Inattention: A Review. ECB Working Paper, 2021.
39. Usher M., Tsetsos K., Yu E.C., Lagnado D.A. Dynamics of decision-making: From evidence accumulation to preference and belief // *Frontiers in Psychology*. 2013. V. 4. <https://doi.org/10.3389/fpsyg.2013.00758>
40. Poltawski L., van Beurden S.B., Morgan-Trimmer S., Greaves C. The dynamics of decision-making in weight loss and maintenance: A qualitative enquiry // *BMC Public Health*. 2020. V. 20. N 1. P. 573. <https://doi.org/10.1186/s12889-020-08664-y>
41. Negre E., Arru M., Rosenthal-Sabroux C. Toward a modeling of population behaviors in crisis situations // *How Information Systems Can Help in Alarm/Alert Detection*. Elsevier, 2018. P. 199–218. <https://doi.org/10.1016/b978-1-78548-302-8.50007-1>
42. De Silva L., Meneguzzi F.R., Logan B. BDI agent architectures: A survey // *Proc. of the 29th International Joint Conference on Artificial Intelligence (IJCAI)*. 2020. P. 4914–4921. <https://doi.org/10.24963/ijcai.2020/684>
43. Rao A.S., Georgeff M.P. BDI agents: from theory to practice // *Proc. of the First International Conference on Multiagent Systems*. 1995. P. 312–319.
44. Lin C.-E., Kavi K.M., Sheldon F.T., Daley K.M., Abercrombie R.K. A methodology to evaluate agent oriented software engineering techniques // *Proc. of the 40th Annual Hawaii International Conference on System Sciences (HICSS'07)*. 2007. P. 60. <https://doi.org/10.1109/hicss.2007.20>
45. Puica M.-A., Florea A.-M. Emotional belief-desire-intention agent model: Previous work and proposed architecture // *International Journal of Advanced Research in Artificial Intelligence*. 2013. V. 2. N 2. <https://doi.org/10.14569/ijarai.2013.020201>
46. Pereira D., Oliveira E., Moreira N. Modelling Emotional BDI Agents // *Proc. of the Formal Approaches to Multiagent Systems (FAMAS) Workshop*. 2006.
47. Van Dyke Parunak H., Bisson R., Brueckner S., Matthews R., Sauter J. A model of emotions for situated agents // *Proc. of the fifth International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS'06)*. 2006. P. 993–995. <https://doi.org/10.1145/1160633.1160810>
48. Jiang H., Vidal J.M., Huhns M.N. EBDI: An architecture for emotional agents // *Proc. of the 6th International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS'07)*. 2007. P. 38–40. <https://doi.org/10.1145/1329125.1329139>
49. Jones H., Saunier J., Lourdeaux D. Personality, emotions and physiology in a BDI agent architecture: The PEP→BDI model // *Proc. of the IEEE/WIC/ACM International Conference on Intelligent Agent Technology (IAT 2009)*. V. 2. 2009. P. 263–266. <https://doi.org/10.1109/wi-iat.2009.160>
50. Hernández D.J., Deniz Ó., Lorenzo J., Hernández M. BDIE: A BDI like architecture with emotional capabilities // *AAAI Spring Symposium — Technical Report*. 2004. V. 2. P. 60–67.
51. Allais M. The so-called allais paradox and rational decisions under uncertainty // *Expected Utility Hypotheses and the Allais Paradox*. 1979. P. 437–681. https://doi.org/10.1007/978-94-015-7629-1_17
52. Beresford B., Sloper P. Understanding the dynamics of decision-making and choice: A scoping study of key psychological theories to inform the design and analysis of the panel study. Social Policy Research Unit, University of York York, 2008.
37. Inglehart R. Postmodernity: Changing values and changing societies. *Polis. Politicheskie issledovaniya*, 1997, no. 4, pp. 6–32. (in Russian)
38. Maćkowiak B., Matějka F., Wiederholt M. *Rational Inattention: A Review*. ECB Working Paper, 2021.
39. Usher M., Tsetsos K., Yu E.C., Lagnado D.A. Dynamics of decision-making: From evidence accumulation to preference and belief. *Frontiers in Psychology*, 2013, vol. 4. <https://doi.org/10.3389/fpsyg.2013.00758>
40. Poltawski L., van Beurden S.B., Morgan-Trimmer S., Greaves C. The dynamics of decision-making in weight loss and maintenance: A qualitative enquiry. *BMC Public Health*, 2020, vol. 20, no. 1, pp. 573. <https://doi.org/10.1186/s12889-020-08664-y>
41. Negre E., Arru M., Rosenthal-Sabroux C. Toward a modeling of population behaviors in crisis situations. *How Information Systems Can Help in Alarm/Alert Detection*, Elsevier, 2018, pp. 199–218. <https://doi.org/10.1016/b978-1-78548-302-8.50007-1>
42. De Silva L., Meneguzzi F.R., Logan B. BDI agent architectures: A survey. *Proc. of the 29th International Joint Conference on Artificial Intelligence (IJCAI)*, 2020, pp. 4914–4921. <https://doi.org/10.24963/ijcai.2020/684>
43. Rao A.S., Georgeff M.P. BDI agents: from theory to practice. *Proc. of the First International Conference on Multiagent Systems*, 1995, pp. 312–319.
44. Lin C.-E., Kavi K.M., Sheldon F.T., Daley K.M., Abercrombie R.K. A methodology to evaluate agent oriented software engineering techniques. *Proc. of the 40th Annual Hawaii International Conference on System Sciences (HICSS'07)*, 2007, pp. 60. <https://doi.org/10.1109/hicss.2007.20>
45. Puica M.-A., Florea A.-M. Emotional belief-desire-intention agent model: Previous work and proposed architecture. *International Journal of Advanced Research in Artificial Intelligence*, 2013, vol. 2, no. 2. <https://doi.org/10.14569/ijarai.2013.020201>
46. Pereira D., Oliveira E., Moreira N. Modelling Emotional BDI Agents. *Proc. of the Formal Approaches to Multiagent Systems (FAMAS) Workshop*, 2006.
47. Van Dyke Parunak H., Bisson R., Brueckner S., Matthews R., Sauter J. A model of emotions for situated agents. *Proc. of the fifth International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS '06)*, 2006, pp. 993–995. <https://doi.org/10.1145/1160633.1160810>
48. Jiang H., Vidal J.M., Huhns M.N. EBDI: An architecture for emotional agents. *Proc. of the 6th International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS'07)*, 2007, pp. 38–40. <https://doi.org/10.1145/1329125.1329139>
49. Jones H., Saunier J., Lourdeaux D. Personality, emotions and physiology in a BDI agent architecture: The PEP→BDI model. *Proc. of the IEEE/WIC/ACM International Conference on Intelligent Agent Technology (IAT 2009)*. V. 2, 2009, pp. 263–266. <https://doi.org/10.1109/wi-iat.2009.160>
50. Hernández D.J., Deniz Ó., Lorenzo J., Hernández M. BDIE: A BDI like architecture with emotional capabilities. *AAAI Spring Symposium — Technical Report*, 2004, vol. 2, pp. 60–67.
51. Allais M. The so-called allais paradox and rational decisions under uncertainty. *Expected Utility Hypotheses and the Allais Paradox*, 1979, pp. 437–681. https://doi.org/10.1007/978-94-015-7629-1_17
52. Beresford B., Sloper P. *Understanding the dynamics of decision-making and choice: A scoping study of key psychological theories to inform the design and analysis of the panel study*. Social Policy Research Unit, University of York York, 2008.

Авторы

Гулева Валентина Юрьевна — кандидат технических наук, старший научный сотрудник, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, <https://orcid.org/0000-0002-1555-9371>, guleva@itmo.ru

Кованцев Антон Николаевич — инженер, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, <https://orcid.org/0000-0003-1765-7001>, ankovantcev@itmo.ru


Суриков Анатолий Георгиевич — кандидат технических наук, старший научный сотрудник, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, <https://orcid.org/0000-0003-4411-271X>, anatoliy.surikov@itmo.ru


Authors


Valentina Yu. Guleva — PhD, Senior Researcher, ITMO University, Saint Petersburg, 197101, Russian Federation, <https://orcid.org/0000-0002-1555-9371>, guleva@itmo.ru


Anton N. Kovatsev — Engineer, ITMO University, Saint Petersburg, 197101, Russian Federation, <https://orcid.org/0000-0003-1765-7001>, ankovantcev@itmo.ru


Anatoly G. Surikov — PhD, Senior Researcher, ITMO University, Saint Petersburg, 197101, Russian Federation, <https://orcid.org/0000-0003-4411-271X>, anatoliy.surikov@itmo.ru


Чунаев Пётр Владимирович — кандидат физико-математических наук, старший научный сотрудник, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация,  36522457300, <https://orcid.org/0000-0001-8169-8436>, chunaev@itmo.ru

Горнова Галина Владимировна — доктор философских наук, доцент, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация,  57204463579, <https://orcid.org/0000-0002-6615-0701>, gornova@itmo.ru

Бухановский Александр Валерьевич — доктор технических наук, профессор, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация,  6603474810, <https://orcid.org/0000-0003-1588-8164>, avbukhanovskii@itmo.ru

Petr V. Chunaev — PhD (Physics & Mathematics), Senior Researcher, ITMO University, Saint Petersburg, 197101, Russian Federation,  36522457300, <https://orcid.org/0000-0001-8169-8436>, chunaev@itmo.ru

Galina V. Gornova — D.Sc. (Philosophy), Associate Professor, ITMO University, Saint Petersburg, 197101, Russian Federation,  57204463579, <https://orcid.org/0000-0002-6615-0701>, gornova@itmo.ru

Alexander V. Boukhanovskiy — D.Sc., Professor, ITMO University, Saint Petersburg, 197101, Russian Federation,  6603474810, <https://orcid.org/0000-0003-1588-8164>, avbukhanovskii@itmo.ru

Статья поступила в редакцию 11.10.2022
Одобрена после рецензирования 12.12.2022
Принята к печати 31.01.2023

Received 11.10.2022
Approved after reviewing 12.12.2022
Accepted 31.01.2023



Работа доступна по лицензии
Creative Commons
«Attribution-NonCommercial»

doi: 10.17586/2226-1494-2023-23-1-136-141

УДК 159.9.072

Методика организации и проведения исследования по оценке потребительской способности

Сергей Андреевич Алейников¹, Ольга Олеговна Гофман², Олег Олегович Басов³✉

^{1,2,3} Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация

¹ aleynikov.sergey.a@gmail.com, <https://orcid.org/0000-0002-6884-2322>

² ms.gofman.oo@gmail.com, <https://orcid.org/0000-0002-4750-5415>

³ oobasov@mail.ru✉, <https://orcid.org/0000-0001-5788-4845>

Аннотация

Предмет исследования. Предложена методика определения особенностей поведения пользователей (их потребности, покупательской активности и структуры затрат) в зависимости от изменения и катастрофизации ситуации. **Метод.** Для выполнения полунатурных исследований (опросов в игровой форме) в области поведенческой экономики создан стенд на основе информационной технологии цифровых ассистентов-аватаров. Применение стенда позволяет осуществлять симуляцию реальных жизненных ситуаций. В ходе исследования участникам каждую внутриигровую неделю необходимо проводить закупки на следующую неделю с учетом собственных представлений о ситуации. Ситуация моделируется путем предъявления участникам сообщений в формате новостной ленты мобильного приложения. После предоставления игровой информации (стимульного материала) участники должны ответить на вопросы о субъективном отношении к стабильности текущей игровой ситуации; самочувствию; интересе к каждой конкретной новости; склонности к импульсивным покупкам в стрессовой ситуации; потребительской активности; возможных внутриигровых планах на будущее.

Основные результаты. Предложено использование мобильного приложения на основе информационной технологии цифровых ассистентов-аватаров с целью сбора и анализа данных пользователей. Разработана методика организации и проведения исследования по оценке потребительской способности. Методика включает в себя следующие этапы: подготовка и сбор предварительной (доигровой) информации о целевой аудитории; проведение исследования с предоставлением стимульного материала (информационного контента), в котором описывается ситуация в окружающей среде (мир, страна, город и т. д.); предоставление вариантов выбора дальнейших действий пользователя как потребителя. **Практическая значимость.** Разработанная методика позволяет выявить особенности потребительского поведения и структуру потребления на общей выборке пользователей в условиях событийной динамики. Появляется возможность определения стратегии поведения потребителей и ошибок, которые совершают участники в ходе игры.

Ключевые слова

потребительская способность, цифровые ассистенты-аватары, игровое исследование, катастрофизация, оценка рисков

Благодарности

Исследование выполнено за счет гранта Российского научного фонда (проект № 17-71-30029).

Ссылка для цитирования: Алейников С.А., Гофман О.О., Басов О.О. Методика организации и проведения исследования по оценке потребительской способности // Научно-технический вестник информационных технологий, механики и оптики. 2023. Т. 23, № 1. С. 136–141. doi: 10.17586/2226-1494-2023-23-1-136-141

Methodology for organizing and conducting a study to assess consumer ability

Sergey A. Aleynikov¹, Olga O. Gofman², Oleg O. Basov³✉

^{1,2,3} ITMO University, Saint Petersburg, 197101, Russian Federation

¹ aleynikov.sergey.a@gmail.com, <https://orcid.org/0000-0002-6884-2322>

² ms.gofman.oo@gmail.com, <https://orcid.org/0000-0002-4750-5415>

³ oobasov@mail.ru✉, <https://orcid.org/0000-0001-5788-4845>

© Алейников С.А., Гофман О.О., Басов О.О., 2023

Abstract

A technique is proposed for determining the characteristics of user behavior (their needs, purchasing activity and cost structure) depending on the change and catastrophization of the situation. To perform semi-natural studies (surveys in a game form) in the field of behavioral economics, a stand was created based on the information technology of digital avatar assistants. The use of the stand makes it possible to simulate real life situations during which participants need to make purchases for the next week every in-game week, taking into account their own ideas about the situation which is modeled by presenting messages to participants in the format of a mobile application news feed. After providing game information (stimulus material), participants must answer questions about their subjective attitude to the stability of the current game situation: about well-being, about interest in each specific news, about the tendency to impulsive purchases in a stressful situation, about consumer activity, about possible in-game plans for future. The use of a mobile application based on the information technology of digital assistants-avatars for the purpose of collecting and analyzing user data is proposed. A methodology for organizing and conducting a study to assess consumer ability has been developed. The methodology includes the following steps: preparation and collection of preliminary (pre-game) information about the target audience; conducting research with the provision of stimulus material (information content) which describes the situation in the environment (world, country, city, etc.); providing options for choosing further actions for the user as a consumer. The developed methodology makes it possible to identify the features of consumer behavior and the structure of consumption in the general sample of users in the context of event dynamics. It becomes possible to determine the strategy of consumer behavior and the mistakes that participants make during the game.

Keywords

consumer ability, digital assistants-avatars, catastrophization, risk assessment

Acknowledgements

The research was carried out at the expense of a grant from the Russian Science Foundation (project No. 17-71-30029).

For citation: Aleynikov S.A., Gofman O.O., Basov O.O. Methodology for organizing and conducting a study to assess consumer ability. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 2023, vol. 23, no. 1, pp. 136–141 (in Russian). doi: 10.17586/2226-1494-2023-23-1-136-141

Введение

Анализ поведения потребителей является традиционным в сфере маркетинга. Так, например, исследования А. Тверски и Д. Канемана [1] посвящены изучению рациональности покупателя. Согласно их модели, предпочтения потребителей часто зависят не только от фактических различий между товарами, но и от того, как эта разница представлена потребителю. Для проверки своей модели исследователи провели серию экспериментов. В них покупателям предлагалось сделать выбор (одни варианты были рациональнее других), при этом задачи отличались формулировками: в первом варианте человек гарантированно получал 25 % выигрыша, во втором — 100 %, но с вероятностью 75 %. Было выявлено, что формулировки задач оказывали больший эффект на покупателя, чем исходная рациональность вариантов. Так, в случае если информация представлялась как потеря, — люди хотели ее избежать (например, потерять 100 % выигрыша с вероятностью 25 %).

Между тем пандемия 2019 года серьезным образом изменила поведение потребителей. Так, фокус внимания постепенно стал смещаться с рациональных выгод выбора товара или услуги на социальные и психологические предикторы поведения покупателя. Согласно работе [2], в США существенная часть покупок перешла в онлайн, они стали носить более локальный характер и осуществляться ближе к дому; увеличилась популярность онлайн-кинотеатров и стриминговых сервисов. Помимо этого, во время пандемии потребители стали меньше тратить деньги на «необязательные» (например, новый телефон, автомобиль) покупки в пользу жизненно важных (например, товаров первой необходимости). В России, согласно [3], покупки также начали переходить в онлайн-формат, а доля трат на хобби, требующих выход из дома, также снизилась. Автор

исследования связывает это не только с опасностью пандемии, но и с мерами, на которые пошло государство для предотвращения распространения вируса. Таким образом, актуальной задачей оказывается оценка рисков, связанных с прогнозированием поведения потребителей с учетом политических, социально-экономических и прочих факторов.

Материалы и методы

Оценка рисков — довольно обширная область исследований, которая затрагивает множество различных отраслей жизнедеятельности. Цель настоящей работы — определение особенностей поведения пользователей (их потребности, покупательскую активность и структуру затрат) в зависимости от изменения и катастрофизации ситуации. Подобные исследования достаточно распространены. Например, в работе [4] использованы сгенерированные новости. Участникам предъявлялись три заголовка новостей, связанных с тематикой — акулы, из различных австралийских средств массовой информации на Facebook. Стимулы в данном случае также варьировались по преднамеренности/непреднамеренности. После демонстрации новостей задавались вопросы (по шкале Лайкерта) о том, как необходимо регулировать жизнедеятельность акул. В исследовании [5] применены не просто новостные заголовки, а сверстаны макетные газетные статьи (один сюжет касался бочек с химикатами, другой — планов по строительству завода по сжиганию опасных отходов). Использовались разные версии макетов: в одной за реализацию проекта отвечало государство, в другой — информация раскрывалась не полностью, а представитель компании указывал, что читатели могут не беспокоиться. После прочтения газеты предлагалось оценить по шкале Лайкерта серьезность риска

и вопрос, что мог бы чувствовать человек, находясь в таком обществе (злость, беспомощность, безопасность, тревогу и т. д.). В работе [6] для оценки рисков также задействованы новостные сюжеты. Таким образом, использование симуляции чрезвычайной ситуации (посредством демонстрации подготовленных новостей или представления ситуации) является методом, который может применяться в исследованиях по оценке рисков. В настоящей работе для создания эффекта «реального времени» были предоставлены участникам исследования не только новостные «газетные» заголовки, но и новости в формате новостной ленты из известных социальных сетей, а также в стиле приложений для общения, просмотра видеоконтента и др.

Проектирование контента направлено на описание общей ситуации в стране и мире, и далее возможного поведения в постепенно «усложняющихся» условиях с коррекцией планов и субъективной оценки потребительской активности с учетом представленной ситуации. В начале исследования ситуация в мире и стране описывается как обычная: международные съезды, обсуждение вопросов социальной политики в стране, проходят концерты и городские мероприятия. Со временем появляются новости о возникновении очагов заболевания новым вирусом, и с каждым днем контент начинает доминировать в ленте участников исследования, оказывая влияние на уровень их частной жизни. Стимульный материал — предоставляемый контент — основывается на реальных новостях, которые появлялись в период пандемии COVID-19. В оригинальных новостных заголовках менялось место действия (например Ухань как очаг заражения был заменен на Санкт-Петербург) и масштабы бедствия для создания постепенного усложнения ситуации. Участникам предлагалось добавить или изменить запланированные встречи с учетом динамики событий. Данные активности были нацелены на усиления «погружения» в созданную ситуацию.

При планировании исследования также учитывалась целевая аудитория воздействия. Известно, что существуют социально-демографические и возрастные особенности восприятия риска. Так, в различных исследованиях установлены различия между женщинами и мужчинами, людьми с разным уровнем дохода и образования, людьми разного возраста [7]. Например, обнаружена возрастная специфика восприятия риска: молодые люди в возрасте 18–29 лет в большей степени склонны игнорировать риск, чем люди более старшего возраста [8]. Отличительной особенностью молодежи в России является динамичность, стремление узнать что-то новое, попробовать новые продукты, воспользоваться новыми услугами, приобрести новые товары, а уровень платежеспособности у молодых людей в возрасте 18–29 лет оценивается как средний и выше среднего [9]. Молодые люди — самые активные потребители на рынке товаров и услуг, которые более других влияют на его развитие, именно поэтому целевой аудиторией в данном исследовании стали студенты старших курсов. Участники были разделены на две группы: не имеющая заработка и получающая деньги от родителей и тех, кто имеет самостоятельный заработок.

В целях исследования поставлены следующие задачи: выявить особенности потребительского поведения и структуру потребления на общей выборке молодежи и в зависимости от самостоятельности дохода; выявить ситуационные маркеры, которые меняют потребительское поведение, в частности катастрофизация ситуации; определить стратегии поведения потребителей в условиях катастрофизации ситуации и динамику изменения потребностей в условиях чрезвычайных ситуаций.

Перед началом исследования были собраны данные об участниках: особенностях образа жизни, ценностях, индивидуально-психологических особенностях, потребительских предпочтениях и пр. Кроме анкетных данных, применены психологические апробированные методики: TIPI-RU (Ten-Item Personality Inventory) в адаптации А.С. Сергеевой, Б.А. Кириллова, А.Ф. Джумагуловой¹ для определения: открытости (экстраверсия — интроверсия), критичности, тревожности, разносторонности (открытость новому), дружелюбия; Шварца, а именно «Профиль личности», который позволяет определять ценности на уровне поведения, т. е. индивидуальные приоритеты, наиболее часто проявляющиеся в социальном поведении личности; склонность к риску А.М. Шуберта; шкала воспринимаемого стресса для диагностики существующего уровня перенапряжения, способности противодействовать стрессу и шкала воспринимаемого стресса.

Проведен анализ временного ряда, предполагающий, что замеры в отношении одних и тех же переменных повторяются с течением времени. После получения игровой информации (стимульного материала) участники должны ответить на следующие вопросы:

- субъективное отношение к стабильности текущей игровой ситуации (по шкале Лайкерта): 1 — нестабильна и непредсказуема, 7 — максимально стабильна;
- самочувствие (по шкале Лайкерта): 1 — плохое настроение, 7 — настроение позитивное, много планов на будущее;
- интерес к каждой конкретной новости — дихотомическая шкала с вариантами ответа «да» и «нет». Например, «Вам интересна данная новость?»;
- склонность к импульсивным покупкам в стрессовой ситуации — дихотомическая шкала с вариантами ответа «да» и «нет» и, далее, открытый вопрос, связанный с тем, на какой товар или услугу человек бы хотел в данный момент потратить деньги. Например, «Хотели бы вы что-нибудь купить, услышав эту новость? Если «да», то что именно?»;
- о потребительской активности (определение потребности осуществить различного рода покупки) — дихотомические вопросы и открытые вопросы. Например, «Хотите сходить на мероприятие? Если «да», то на какое именно?»;
- возможные внутриигровые планы на будущее — открытые вопросы. Например: «Что планируете на лето? Отметьте пункты, которые в большей степе-

¹ Краткий личностный опросник TIPI-RU [Электронный ресурс]. Режим доступа: <https://psytests.org/big5/tipi.html> (дата обращения: 20.11.2022).

Таблица. План проведения исследования
Table. Research plan

Реальный день	Внутриигровые дни	Характеристика ситуации во внутриигровые дни	Действия участников в конце реального дня
1	1, 2, 7, 10	Происходит мирная жизнь без вируса и первые очаги заражения	Производят внутриигровые покупки на следующую неделю
2	14, 15, 17, 21	Уже известны тысячи случаев заболеваний и начинают закрываться страны	
3	26, 27, 28, 33	Темпы распространения вируса начинают приобретать большой размах. Количество заразившихся с тысяч переходит на миллионы	
4	34, 35, 36, 37	Начало тестирования вакцины и увеличение количества зараженных	
5	38, 39, 50	На момент окончания реального дня количество смертей от вируса превышает 6 млн, а российская вакцина не показывает эффективных результатов	

ни соответствуют вашим планам: путешествие за границу, путешествие по России и т. д.» с возможностью ввода собственного варианта времяпрепровождения.

По данной структуре в течение всего игрового исследования был оценен каждый стимульный материал — информационный контент. Использование анализа временного ряда позволило увидеть динамику изменения потребительской активности с повышением катастрофизации ситуации.

Для выполнения полунатурных исследований (опросов в игровой форме) в области поведенческой экономики создан стенд на основе информационной технологии цифровых ассистентов-аватаров¹. Стенд имеет микросервисную архитектуру и включает мобильное приложение для взаимодействия с пользователями, а также серверную часть платформы, реализующую функции обработки и хранения данных. Стенд развернут на вычислительных серверах Национального

¹ Платформа экосистемы цифровой личности [Электронный ресурс]. Режим доступа: <https://actcognitive.org/platformy/platforma-ekosistemy-tsifrovoy-lichnosti> (дата обращения: 10.11.2022).

Результаты исследования

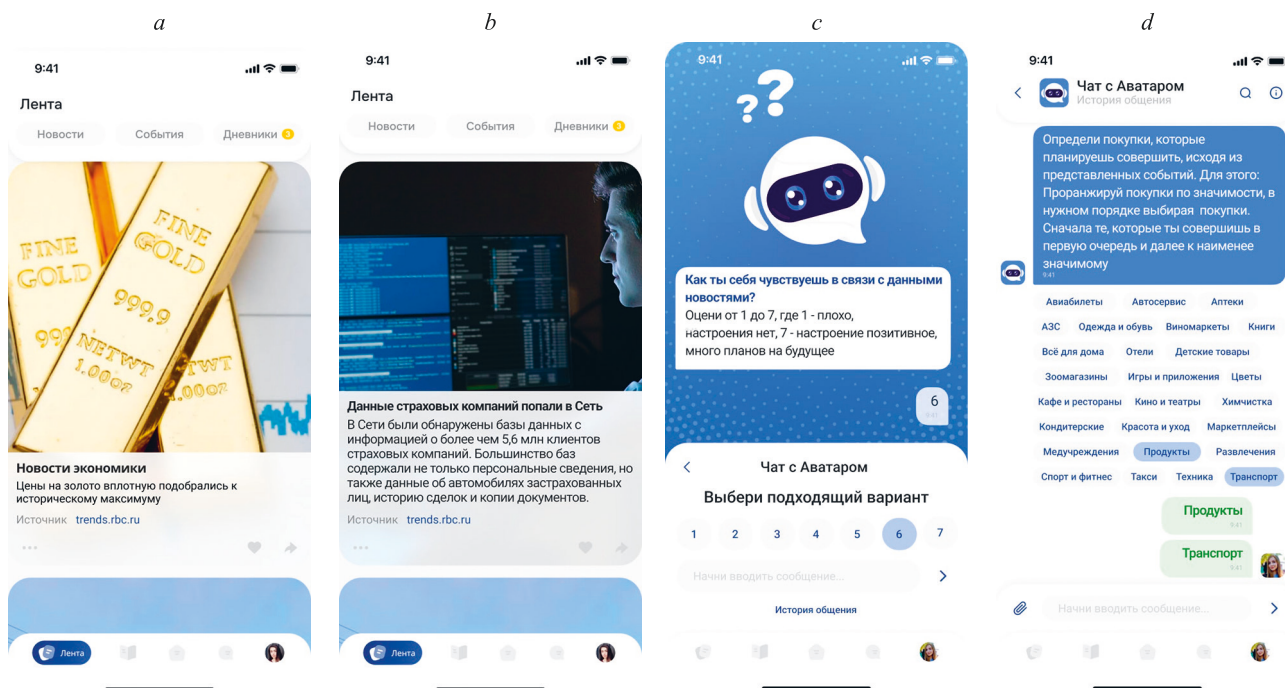


Рисунок. Пролистывание ленты событий (a); просмотр описания события (b); оценка предъявленного стимульного материала (c); осуществление закупки на следующую неделю (d)

Figure. Scrolling through the event feed (a); view event description (b); assessment of the presented stimulus material (c); making a purchase for the next week (d)

центра когнитивных разработок Университета ИТМО и функционирует в режиме 24×7 с сентября 2021 г. В настоящее время объем исследуемой группы составляет более 200 человек (студентов и преподавателей ИТМО).

Игровое исследование выполнено в реальных условиях, когда каждый «игрок» играет себя. Для минимизации времени «игра» проведена в условиях ускоренного времени (например, 3–5 дней игры за один реальный день), что позволило за две недели эмулировать развитие глубокого кризиса, связанного с чрезвычайным положением (таблица).

Участники каждую внутриигровую неделю осуществляли закупки на следующую неделю с учетом собственных представлений о ситуации и ее катастрофизации. В данном случае игровое исследование приближено к симуляции реальной жизни. Участникам был представлен бюджет, который соответствовал их среднему бюджету, указанному при первичной диагностике. Кроме того, участники имели две возможности управления своим бюджетом: потратить на продукты и/или отложить. Для выполнения покупки необходимо было проранжировать список из возможных категорий покупок с учетом субъективной значимости участника, после чего распределить сумму трат между этими категориями.

Так, например, на 21-й внутриигровой день участникам исследования объявлена следующая ситуация в мире: «Европейские лидеры проголосовали за то, чтобы закрыть как минимум 26 стран для почти всех

посетителей из остального мира как минимум на 60 дней. В России также говорят о возможном закрытии границ с ближайшими государствами. Ранее подобное было только с COVID-19». С использованием исследовательского стенда участники реализовали следующую последовательность действий (рисунок).

Заключение

Актуальность проведения игрового исследования связана с необходимостью выявления стратегий потребительского поведения в условиях неопределенности. Для реализации данной цели предложено использовать информационную технологию цифровых ассистентов-аватаров и разработанную методику организации и проведения исследования по оценке потребительской способности.

Полученные результаты разработки методики игрового исследования позволили: выявить особенности потребительского поведения и структуру потребления на общей выборке молодежи и в зависимости от самостоятельности дохода; выявить ситуационные маркеры, которые меняют потребительское поведение в условиях событийной динамики; определить стратегии поведения потребителей и ошибки, которые совершают участники в ходе игры.

Проведение исследования по оценке потребительской способности планируется после стабилизации внешнеполитической ситуации.

Литература

1. Tversky A., Kahneman D. Rational choice and the framing of decisions // *Multiple Criteria Decision Making and Risk Analysis Using Microcomputers*. 1989. P. 81–126. https://doi.org/10.1007/978-3-642-74919-3_4
2. Sheth J. Impact of Covid-19 on consumer behavior: Will the old habits return or die? // *Journal of Business Research*. 2020. V. 117. P. 280–283. <https://doi.org/10.1016/j.jbusres.2020.05.059>
3. Плотноков В.А. Пандемия COVID-19, потребительский рынок и цифровизация // *Экономическое возрождение России*. 2021. № 3(69). С. 92–104. <https://doi.org/10.37930/1990-9780-2021-3-69-92-104>
4. Le Busque B., Dorrian J., Litchfield C. The impact of news media portrayals of sharks on public perception of risk and support for shark conservation // *Marine Policy*. 2021. V. 124. P. 104341. <https://doi.org/10.1016/j.marpol.2020.104341>
5. Sandman P.M., Miller P.M., Johnson B.B., Weinstein N.D. Agency communication, community outrage, and perception of risk: Three simulation experiments // *Risk Analysis*. 1993. V. 13. N 6. P. 585–598. <https://doi.org/10.1111/j.1539-6924.1993.tb01321.x>
6. Durfee J.L. “Social change” and “status quo” framing effects on risk perception: an exploratory experiment // *Science Communication*. 2006. V. 27. N 4. P. 459–495. <https://doi.org/10.1177/1075547005285334>
7. Nan X., Kim J., Verrill L., Daily K. Influence of risk severity and intentionality on perceptions of food contamination risks among pregnant women: an experimental investigation // *Journal of Health Communication*. 2020. V. 25. N 6. P. 514–521. <https://doi.org/10.1080/10810730.2020.1818149>
8. Otani H., Leonard S.D., Ashford V.L., Bushroe M., Reeder G. Age differences in perception of risk // *Perceptual and Motor Skills*. 1992. V. 74. N 2. P. 587–594. <https://doi.org/10.2466/pms.1992.74.2.587>
9. Мельникова Т.Ф., Мельникова А.В., Евченко М.Н., Нияскина А.В. Особенности молодежного сегмента на рынке потребительских товаров // *Молодой ученый*. 2014. № 17. С. 301–304.

References

1. Tversky A., Kahneman D. Rational choice and the framing of decisions. *Multiple Criteria Decision Making and Risk Analysis Using Microcomputers*, 1989, pp. 81–126. https://doi.org/10.1007/978-3-642-74919-3_4
2. Sheth J. Impact of Covid-19 on consumer behavior: Will the old habits return or die? *Journal of Business Research*, 2020, vol. 117, pp. 280–283. <https://doi.org/10.1016/j.jbusres.2020.05.059>
3. Plotnikov V.A. The Covid-19 pandemic, consumer market and digitalization. *Economic Revival of Russia*, 2021, no. 3(69), pp. 92–104. (in Russian). <https://doi.org/10.37930/1990-9780-2021-3-69-92-104>
4. Le Busque B., Dorrian J., Litchfield C. The impact of news media portrayals of sharks on public perception of risk and support for shark conservation. *Marine Policy*, 2021, vol. 124, pp. 104341. <https://doi.org/10.1016/j.marpol.2020.104341>
5. Sandman P.M., Miller P.M., Johnson B.B., Weinstein N.D. Agency communication, community outrage, and perception of risk: Three simulation experiments. *Risk Analysis*, 1993, vol. 13, no. 6, pp. 585–598. <https://doi.org/10.1111/j.1539-6924.1993.tb01321.x>
6. Durfee J.L. “Social change” and “status quo” framing effects on risk perception: an exploratory experiment. *Science Communication*, 2006, vol. 27, no. 4, pp. 459–495. <https://doi.org/10.1177/1075547005285334>
7. Nan X., Kim J., Verrill L., Daily K. Influence of risk severity and intentionality on perceptions of food contamination risks among pregnant women: an experimental investigation. *Journal of Health Communication*, 2020, vol. 25, no. 6, pp. 514–521. <https://doi.org/10.1080/10810730.2020.1818149>
8. Otani H., Leonard S.D., Ashford V.L., Bushroe M., Reeder G. Age differences in perception of risk. *Perceptual and Motor Skills*, 1992, vol. 74, no. 2, pp. 587–594. <https://doi.org/10.2466/pms.1992.74.2.587>
9. Melnikova T.F., Melnikova A.V., Evchenko M.N., Niiaskina A.V. Special aspects of the youth segment in the consumer goods market. *Young Scientist*, 2014, no. 17, pp. 301–304. (in Russian)

Авторы

Алейников Сергей Андреевич — инженер, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, <https://orcid.org/0000-0002-6884-2322>, aleynikov.sergey.a@gmail.com

Гофман Ольга Олеговна — кандидат психологических наук, старший научный сотрудник, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, [sc 57219905431](https://orcid.org/0000-0002-4750-5415), <https://orcid.org/0000-0002-4750-5415>, ms.gofman.oo@gmail.com

Басов Олег Олегович — доктор технических наук, доцент, профессор, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, [sc 16400670700](https://orcid.org/0000-0001-5788-4845), <https://orcid.org/0000-0001-5788-4845>, oobasov@mail.ru

Authors

Sergey A. Aleynikov — Engineer, ITMO University, Saint Petersburg, 197101, Russian Federation, <https://orcid.org/0000-0002-6884-2322>, aleynikov.sergey.a@gmail.com

Olga O. Gofman — PhD (Psychology), Senior Researcher, ITMO University, Saint Petersburg, 197101, Russian Federation, [sc 57219905431](https://orcid.org/0000-0002-4750-5415), <https://orcid.org/0000-0002-4750-5415>, ms.gofman.oo@gmail.com

Oleg O. Basov — D.Sc., Associate Professor, Professor, ITMO University, Saint Petersburg, 197101, Russian Federation, [sc 16400670700](https://orcid.org/0000-0001-5788-4845), <https://orcid.org/0000-0001-5788-4845>, oobasov@mail.ru

Статья поступила в редакцию 08.10.2022

Одобрена после рецензирования 12.12.2022

Принята к печати 26.01.2023

Received 08.10.2022

Approved after reviewing 12.12.2022

Accepted 26.01.2023



Работа доступна по лицензии
Creative Commons
«Attribution-NonCommercial»

doi: 10.17586/2226-1494-2023-23-1-142-149

УДК 004.946

Автоматизированный кластерный анализ коммуникативных стратегий образовательных telegram-каналов

Борис Абдуллохонвич Низомутдинов¹✉, Анна Борисовна Углова²,
Ирина Марковна Богдановская³

¹ Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация

^{2,3} Российский государственный педагогический университет им. А. И. Герцена, Санкт-Петербург, 191186, Российская Федерация

¹ boris@itmo.ru✉, <https://orcid.org/0000-0002-4090-9564>

² anna.uglova@list.ru, <https://orcid.org/0000-0002-8072-0539>

³ ibogdanovs@ Herzen.spb.ru, <https://orcid.org/0000-0001-7303-615X>

Аннотация

Предмет исследования. Вопросы мониторинга образовательной коммуникации, анализа коммуникативных стратегий и тактик в презентации учебных материалов мало изучены. Изучение основных тематик содержания публикаций на каналах с разным рейтингом популярности у пользователей может рассматриваться как один из этапов разработки инструментов для анализа образовательной коммуникации в Telegram. В данной работе исследован дидактический дизайн виртуального образовательного канала на примере Telegram. Изучена его коммуникативная направленность, стратегии и тактики взаимодействия, которые используются преподавателями для достижения высоких результатов своих студентов и повышения вовлеченности аудитории. **Метод.** С применением методов машинного обучения по имеющемуся набору публикаций образовательных telegram-каналов выполнено разбиение текстового массива на кластеры для дальнейшего экспертного анализа и определения приблизительной тематики. С этой целью использована программная платформа для анализа данных PolyAnalyst, разработанная компанией Megaputer Intelligence. Платформа обеспечивает кластеризацию документов по методу k -средних (k -means), поддерживает этапы процесса анализа данных от загрузки и обработки данных до расширенного анализа текста и данных, а также создание пользовательских отчетов. **Основные результаты.** Представлена тематическая структура контента образовательных telegram-каналов с высоким и низким рейтингами и статистическая информация о дидактическом наполнении образовательных ресурсов. Показано, что высокорейтинговые образовательные telegram-каналы осуществляют смысловую стратегию интеграции образовательного и карьерного маршрутов. Образовательные telegram-каналы с низким рейтингом выполняют коммуникативную стратегию, направленную на предоставление узкоспециализированной, логически разобщенной справочной, коммерческой или развлекательной информации. Одним из признаков коммуникативных тактик на низкорейтинговых каналах выступают приемы манипулятивного воздействия, которые позволяют влиять на мнение аудитории. К ним относятся тактики: косвенного убеждения, актуализации мотива финансовой выгоды и заполнения информационных «пробелов». **Практическая значимость.** Полученные результаты могут найти применение при разработке инструментов для анализа и мониторинга образовательной коммуникации в сети Интернет. Методика автоматизированного кластерного анализа коммуникативных стратегий образовательных telegram-каналов может быть востребована широким кругом специалистов в сфере менеджмента образования, разработчиками контента образовательных интернет-каналов, маркетологам, педагогам, работающим в виртуальной среде.

Ключевые слова

автоматизированный кластерный анализ, коммуникативные стратегии, коммуникативные тактики, образовательные ресурсы, Telegram

Благодарности

Исследование выполнено за счет гранта Российского научного фонда № 22-78-10047 «Конструктивные и деструктивные коммуникативные практики специалистов помогающих профессий в цифровых медиа», <https://rscf.ru/project/22-78-10047/>

© Низомутдинов Б.А., Углова А.Б., Богдановская И.М., 2023

Ссылка для цитирования: Низомутдинов Б.А., Углова А.Б., Богдановская И.М. Автоматизированный кластерный анализ коммуникативных стратегий образовательных telegram-каналов // Научно-технический вестник информационных технологий, механики и оптики. 2023. Т. 23, № 1. С. 142–149. doi: 10.17586/2226-1494-2023-23-1-142-149

Automated cluster analysis of communication strategies of educational telegram channels

Boris A. Nizomutdinov¹✉, Anna B. Uglova², Irina M. Bogdanovskaya³

¹ ITMO University, Saint Petersburg, 197101, Russian Federation

^{2,3} Herzen State Pedagogical University of Russia, Saint Petersburg, 191186, Russian Federation

¹ boris@itmo.ru✉, <https://orcid.org/0000-0002-4090-9564>

² anna.uglova@list.ru, <https://orcid.org/0000-0002-8072-0539>

³ ibogdanovs@herzen.spb.ru, <https://orcid.org/0000-0001-7303-615X>

Abstract

The issues of educational communication monitoring, analysis of communication strategies and tactics in the presentation of educational materials have been little studied. The study of the main topics of the content of publications on channels with different popularity ratings among users can be considered as one of the stages of developing tools for analyzing educational communication in Telegram. In this paper, the didactic design of a virtual educational channel is studied on the example of Telegram. Its communicative orientation, strategies and tactics of interaction, which are used by teachers to achieve high results of their students and increase audience engagement, are studied. Using machine learning methods based on the existing set of publications of educational Telegram channels, the text array was divided into clusters for further expert analysis and determination of approximate topics. For this purpose, the PolyAnalyst data analysis software platform developed by Megaputer Intelligence was used. The platform provides clustering of documents using the *k*-means method and supports the stages of the data analysis process from data loading and processing to advanced text and data analysis as well as supports the creation of custom reports. The thematic structure of the content of educational Telegram channels with high and low ratings and statistical information on the didactic content of educational resources is presented. It is shown that highly rated educational Telegram channels implement a semantic strategy for integrating educational and career routes. Educational Telegram channels with a low rating implement a communicative strategy aimed at providing highly specialized, logically disconnected reference, commercial or entertainment information. One of the signs of communicative tactics on low-rating channels are manipulative techniques that allow you to influence the opinion of the audience. These include the tactic of indirect persuasion, the tactic of actualizing the motive of financial gain, the tactic of filling information “gaps”. The results obtained can be used in the development of tools for the analysis and monitoring of educational communication on the Internet. The methodology of automated cluster analysis of communicative strategies of educational Telegram channels can be in demand by a wide range of specialists in the field of education management, content developers of educational Internet channels, marketers, teachers working in a virtual environment.

Keywords

automated cluster analysis, communication strategies, communication tactics, educational resources, Telegram

Acknowledgements

The research was carried out at the expense of the grant of the Russian Science Foundation No. 22-78-10047, <https://rscf.ru/project/22-78-10047/> “Constructive and destructive communicative practices of specialists of helping professions in digital media”.

For citation: Nizomutdinov B.A., Uglova A.B., Bogdanovskaya I.M. Automated cluster analysis of communication strategies of educational telegram channels. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 2023, vol. 23, no. 1, pp. 142–149 (in Russian). doi: 10.17586/2226-1494-2023-23-1-142-149

Введение

Постоянный рост потребности общества в новых знаниях и квалифицированных специалистах, технологический прогресс и практики социального дистанцирования приводят к развитию образовательных технологий и активному внедрению электронного обучения [1]. Электронное обучение позволяет упростить процесс передачи образовательных материалов и сделать процесс взаимодействия более интерактивным, используя как синхронные, так и асинхронные формы обучения [2]. Общение учителя и ученика в реальном времени вне зависимости от геолокации позволяет синхронизировать обучение. С другой стороны, асинхронная форма обучения дает возможность постоянного до-

ступа к учебным материалам. Учащиеся могут учиться в любое время и самостоятельно регулировать процесс коммуникации [3].

Платформы социальных сетей и мессенджеры хорошо интегрировались в систему образования. Однако вопрос оценки их роли и влияния на процесс обучения остается не до конца изученным.

Основная из популярных коммуникационных площадок для обучения — кроссплатформенная система «Telegram». Многие исследователи указывают, что технический функционал данной платформы (коммерциализация, функциональность, архитектура, безопасность) делает ее оптимальной для поддержки образовательного процесса [4]. Данная платформа позволяет преодолеть социально-экономические и эпидемиоло-

гические проблемы, предоставив доступ к обучению людям с ограниченными возможностями здоровья, социально незащищенным слоям населения и лицам с ограниченными возможностями мобильности [5].

Предмет исследования

В последнее время Telegram привлекает внимание исследователей как источник информации для изучения различных общественных процессов.

В работе [6] предложена обработка данных из мессенджера WhatsApp с помощью методов машинного обучения для поиска фейков. Представлен метод FakeWhatsApp.BR, использованный в работе, с помощью которого собран набор данных сообщений WhatsApp на бразильском варианте португальского языка, из бразильских публичных групп. Разработан классификатор дезинформации, сочетающий методы извлечения признаков на основе обработки естественного языка.

В работе [7] представлена система «Telegram Monitor», которая отслеживает политические дебаты в данной среде и позволяет анализировать наиболее распространенный контент на нескольких каналах и в общественных группах. По мнению авторов, система призвана помогать журналистам, исследователям и агентствам по проверке фактов выявления актуальных теорий заговора, кампаний по дезинформации или просто отслеживания политических дебатов во время выборов в Бразилии 2022 года.

Профессор Санкт-Петербургского государственного университета С.С. Бодрунова [8] исследовала Telegram сообщества, выступающие против вакцинации. При этом была проведена оценка более 282 тыс. комментариев в крупнейшем сообществе antivaxxer в российском Telegram (anti_covid21: с января по июль 2021).

В отечественных и зарубежных научных работах представлено большое количество исследований, посвященных изучению самого процесса электронного обучения и отношения студентов и преподавателей к его особенностям. В работе [9] отмечено, что возможности системы Telegram позволяют регулировать процесс взаимодействия преподавателя и обучающихся, в реальном времени отслеживая посещаемость и вовлеченность студентов.

В [10] доказано, что использование групповых чатов, поддержки различных типов онлайн-взаимодействия в Telegram эффективнее способствует повышению вовлеченности учащихся, чем использование других мессенджеров.

Одно из неоспоримых преимуществ применения данной платформы — возможность стимулирования активного диалога между обучающимися, который позволяет активизировать профессиональные и учебные знания [11].

Отметим, что платформа Telegram в первую очередь полезна для закрепления уже полученных знаний и обмена учебными материалами для обучающихся, которые уже получили базовые знания [12].

В работе [13] рассмотрен субъективный опыт студентов в использовании платформы Telegram для об-

учения языку и отмечено, что она оказалась удобна для развития имеющихся навыков (чтение, письмо, аудирование), но не показала своей эффективности в освоении новых разговорных навыков.

Возникает вопрос об эффективности обучения с использованием электронных платформ для широкого круга обещающих, с разным уровнем подготовки.

В систематическом обзоре исследований [14], посвященных использованию Telegram, показано, что в большинстве исследований доказана успешность использования данной платформы. Это связано в первую очередь с профессионализмом преподавателя, который сможет выстроить успешную онлайн-стратегию обучения и, которая будет отличаться от классических офлайн-методов работы.

Таким образом, можно говорить о том, что актуальным становится вопрос изучения дидактического дизайна виртуального образовательного канала, его коммуникативной направленности, стратегий и тактик взаимодействия, которые используют преподаватели для достижения высоких результатов своих студентов и повышения вовлеченности аудитории. На данный момент реализовано большое количество исследований, посвященных экспертному анализу политической коммуникации в telegram-каналах [15], медицинской коммуникации врачей и пациентов [16], производственной коммуникации [17], доверия индивидуальным блогам [18]. Однако существует дефицит научных работ, посвященных мониторингу образовательной коммуникации, анализу коммуникативных стратегий и тактик в презентации учебных материалов.

Предположим, что одним из этапов разработки инструментов для анализа образовательной коммуникации в Telegram может стать изучение основных тематик содержания публикаций на каналах с разным рейтингом популярности у пользователей.

Сбор информации

В России аудитория Telegram превышает 35 млн человек. По данным TGStat, в русскоязычном Telegram сейчас больше 300 тыс. каналов. Первым шагом в данном исследовании стал отбор каналов и сбор статистики посещаемости. Для этого задействован сервис TGStat, который имеет готовые подборки по различным тематикам. Также TGStat предлагает полный спектр опций для быстрого и удобного отслеживания публикаций в telegram-каналах.

Для исследования выбран рейтинг каналов по образованию, включающий 100 позиций образовательных каналов и построенный по количеству подписчиков. Все ссылки на каналы сохранены в порядке убывания популярности, а также имеется расширенная статистика по каждому каналу. Сервис предоставляет различную информацию, в том числе индекс цитирования, охваты публикаций, внешний трафик, топ публикаций и др.

На следующем шаге загружены все посты, которые опубликованы в собранной подборке. Telegram имеет развитый интерфейс прикладного программирования API (Application Programming Interface) для создания

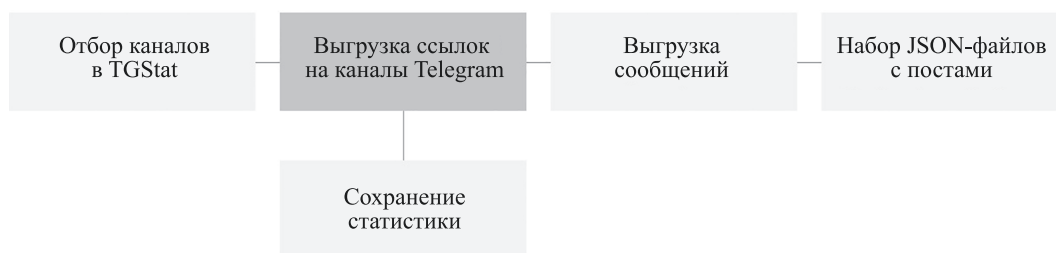


Рис. 1. Общая схема сбора информации
Fig. 1. General information collection scheme

ботов. Бот API позволяет легко создавать программы, которые используют сообщения Telegram для интерфейса. API Telegram и TDLlib создают собственные настраиваемые клиенты Telegram. При этом чаты можно выгружать без API с использованием стандартного функционала, доступного в мессенджере. Для сбора информации применены встроенные функции мессенджера Telegram, в интерфейсе программы для персонального компьютера (ПК) возможно выгрузить сообщения в формате JSON (JavaScript Object Notation — нотация объектов в Java Script) из канала. В результате сбора скачено 100 JSON-файлов. На рис. 1 приведена общая схема сбора информации.

Выполним сравнение каналов из подборки: 10 первых (самых популярных) и 10 последних (менее популярных).

Обработка информации

Цель, которая была поставлена в части обработки текста, заключалась в следующем — по имеющемуся набору публикаций разных тематик в сфере образования, собранных из Telegram, автоматически (без ручной разметки, с помощью методов машинного обучения) определить, к какой тематике в сфере образования относится пост в виде текста и какой тон (позитивный или отрицательный) имеет каждый пост. Для достижения данной цели поставлены следующие задачи.

- Для автоматического определения тематики текста необходимо решить задачу кластеризации текста, т. е. разбиения текстового массива на произвольное количество групп (кластеров) для дальнейшего экспертного анализа по определению приблизительной тематики определенной группы новостей.
- Для автоматического определения тональности новости (позитивный или негативный оттенок) необходимо выбрать готовый и публичный русскоязычный массив текста с заданной разметкой на положительные и отрицательные группы, и на основе данного массива подготовить модель по автоматическому определению тональности текста новостей.

На данном этапе исследования, в качестве эксперимента, использован готовый сервис. В [19] представлены результаты исследования круга проблем и вопросов финансового положения студентов в период пандемии по цифровым следам в социальных сети ВКонтакте студенческой аудитории. Для интеллектуального анализа данных использовано специализированное

программное обеспечение Polyanalyst — система извлечения полезной информации из структурированных и неструктурированных данных с применением методов машинного обучения. В настоящей работе также использовано решение от компании Megaruter Intelligence — PolyAnalyst. Данное программное обеспечение поддерживает все этапы процесса анализа данных: от загрузки и обработки данных до расширенного анализа текста и данных, а также создания пользовательских отчетов.

PolyAnalyst обеспечивает кластеризацию документов, с использованием метода k -means. Работа алгоритма заключается в минимизации общего квадратичного отклонения точек кластера от центров самих этих кластеров. По каждому из образованных кластеров созданы списки с топ-10 ключевыми словами, характерными для каждого кластера.

Дополнительно проведен анализ публикационной активности telegram-каналов по годам, статистика по количеству символов в постах и другие параметры, а также выделены ключевые слова. На рис. 2 приведена общая схема работы по обработке текста.

Анализ полученных данных

В результате анализа статистической информации о дидактическом наполнении образовательных telegram-каналов можно сделать вывод, что образовательные telegram-каналы с высоким рейтингом представляют собой долгосрочные проекты с большим объемом публикаций за последние 6 лет. В это же время образовательные telegram-каналы с низким рейтингом большую часть информации публиковали в 2022 году, что указывает на краткосрочность данных проектов.

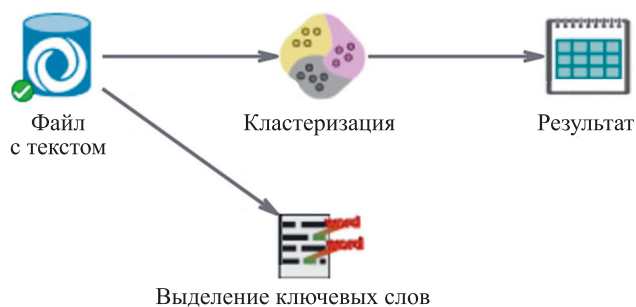


Рис. 2. Общая схема анализа данных
Fig. 2. General scheme of data analysis

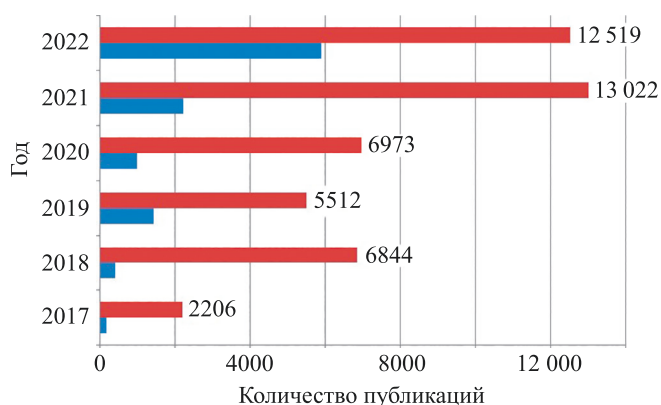


Рис. 3. Публикационная активность по годам. Красным цветом показаны 10 каналов из подборки (наиболее популярные), синим — 10 каналов (наименее популярных)

Fig. 3. Publication activity by year. 10 channels from the selection (the most popular) are shown in red, 10 channels (the least popular) are shown in blue

Анализ объема символов в отдельных постах показал, что образовательные telegram-каналы с высоким рейтингом в основном публикуют достаточно короткие

сообщения (18 % контента — посты до 500 знаков), в то время как контент telegram-каналов с низким рейтингом достаточно разноформатный, чаще публикуются объемные сообщения. Возможно из-за того, что образовательные telegram-каналы с высоким рейтингом публикуют более лаконичную, сжатую информацию, ее проще воспринимать и она привлекает большую аудиторию.

Выполним кластерный анализ содержательных аспектов контента, выкладываемого из 10 наиболее популярных каналов и 10 наименее популярных. В результате по каждому из образованных кластеров созданы списки с топ-10 ключевыми словами, характерными для каждого кластера. Результаты визуализации распределения постов по кластерам представлены на рис. 4 и 5.

Анализ контента образовательных telegram-каналов с высоким рейтингом (рис. 4) позволил выделить три большие группы коммуникативных стратегий, объединяющих 10 кластеров (табл. 1). В кластере №11 собраны посты, для которых не удалось выделить категории.

Таким образом, анализ тематик высокорейтинговых образовательных telegram-каналов позволил реконструировать коммуникативные тактики, которые ориентированы на взрослых пользователей и сконцентрированы в первую очередь на процессе обучения и его связи с

Таблица 1. Коммуникативные стратегии и кластеры образовательных telegram-каналов с высоким рейтингом

Table 1. Communicative strategies and clusters of the educational Telegram channels with a high rating

Номер стратегии	Название коммуникативной стратегии	Кластер		
		номер	название	ключевые слова
1	Организация тайм-менеджмента учебной деятельности	1	Организация тайм-менеджмента учебной деятельности	объединяет темы, призванные побуждать к активной деятельности, описывающие пространственно-временную организацию процесса обучения (цель, день, жизнь, минута, сила, человек, делать, привычка, сделать)
2	Информирование о построении образовательного маршрута и начале карьеры	2	Организация образования	объединил темы, связанные с организацией высшего университетского образования (стипендия, обучение, университет, дедлайн, студент, программа, академический, подробный)
		3	Участие в конкурсах	дедлайн, участие, конкурс, организатор, заявка, молодой, подробный, проживание, расход, месяц
		4	Стажировка	стажировка, стажер, английский, опыт, подробный, требования, компания, область, месяц, владеть
		5	Поступление в университет	вуз, поступление, образование, поступить, университет, страна, учеба, рубеж, получить, учиться
		6	Начало карьеры	курс, бесплатный, навык, карьера, компания, сертификат, вакансия, онлайн курс, пройти, стажировка
3	Предложения по теме образования	7	Обучение физике	физика, механика, закон, движение, электрический, квантовый, волна, ток, поле, теория
		8	Обучение программированию	программирование, код, книга, разработка, приложение, издание, язык, пример, программист, проектирование
		9	Обучение математике	задача, математика, алгебра, теория, математический, решение, уравнение, дифференциальный, геометрия, вероятность
		10	История математики	математика, наука, математический, книга, история, задача, физика, занимательный, читатель

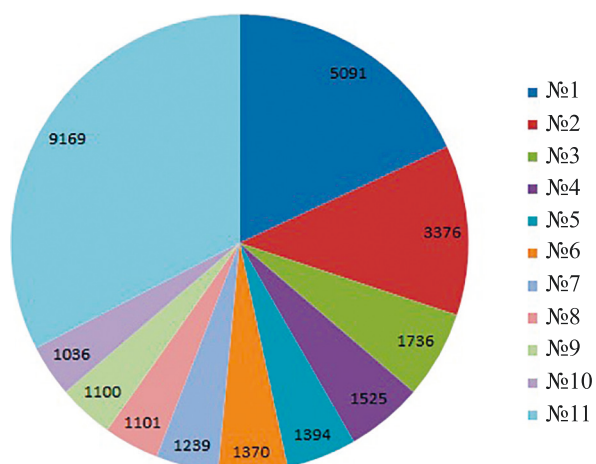


Рис. 4. Категориальная структура коммуникативных тактик образовательных telegram-каналов с высоким рейтингом.

Общее количество сообщений по каждому кластеру

Fig. 4. Categorical structure of communicative tactics of educational telegram channels with a high rating. Total number of messages for each cluster

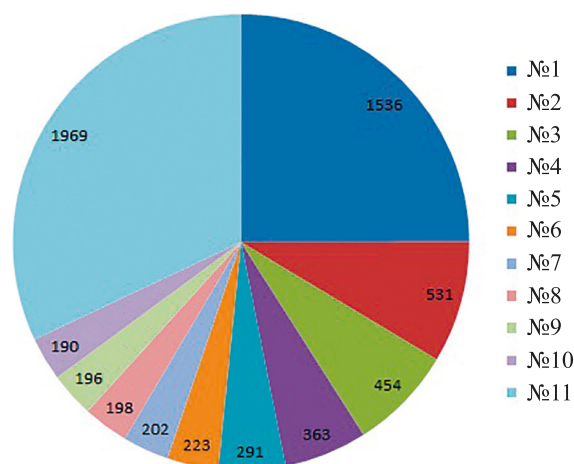


Рис. 5. Категориальная структура коммуникативных тактик образовательных telegram-каналов с низким рейтингом.

Общее количество сообщений по каждому кластеру

Fig. 5. Categorical structure of communication tactics of educational telegram channels with a low rating. Total number of messages for each cluster

практикой. Это говорит о том, что высокорейтинговые образовательные telegram-каналы реализуют коммуникативную стратегию объединения образовательного и карьерного маршрутов, что логически связывает всю информацию в единый контент.

Анализ контента образовательных telegram-каналов с низким рейтингом позволил выделить 10 узкотематических кластеров (табл. 2), которыми можно описать как диалоговые коммуникативные тактики, направленные на описание конкретной области знания и выбор

Таблица 2. Коммуникативные стратегии и кластеры образовательных telegram-каналов с низким рейтингом

Table 2. Communicative strategies and clusters of the educational Telegram channels with a low rating

Номер стратегии	Название коммуникативной стратегии	Кластер		
		номер	название	ключевые слова
1	Предоставление справочной информации	1	Школьная подготовка к тестам и экзаменам	класс, регион, друг, звать, ответ, язык, математика, выложить, вариант
		2	Научные государственные проекты	наука, научный, ученый, российский, проект, организация, финансирование, развитие, федеральный
		3	Информация об образовательных программах	вуз, выпускник, образование, программа, цифровой, обучение, экзамен, образовательный, абитуриент, показатель
		4	Информация о вакцинации	студент, дистанционный, вакцинация, обучение, формат, вуз, прививка, режим, диктант, корона-вирус
2	Предложение коммерческих образовательных проектов	5	Бизнес-курсы	бизнес, команда, компания, проект, бесплатный, продукт, узнать, эксперт, курс, маркетинг
		6	Медицинская информация	вебинар, лечение, медицинский, специалист, заболевание, уважаемый, модуль, пациент, клетка, болезнь
		7	Коммерческие курсы	курс, профессия, освоить рынок, теория, методист, скидка, программирование, специальность, нуль
3	Предложение образовательно-развлекательного контента	8	Правовая информация	дело, уголовный, диссертация, ученый, наука, суд, степень, взятка, следствие, мошенничество
		9	Познавательная видеoinформация	канал, узнать, онлайн, рассказать, видео, минута, хотеть, курс, инструмент, время
		10	Информация о кино	театр, фильм, роль, актер, кино, актриса, родиться, сняться, режиссер, артист

узкой темы (рис. 5). В кластере №11 собраны посты, для которых не удалось выделить категории.

В результате анализа образовательных тематик telegram-каналов с низким рейтингом можно сказать, что они реализуют коммуникативную стратегию, направленную на предоставление достаточно узкоспециализированной, но не всегда логически связанной справочной, коммерческой или развлекательной информации. Анализ ключевых слов показал, что одним из важных маркеров коммуникативных тактик на низкорейтинговых каналах выступают приемы манипулятивного воздействия, позволяющие реализовывать влияние на мнение аудитории. К ним относятся: тактика косвенного убеждения, связанная с желанием усилить впечатление о профессионализме того или иного субъекта при помощи таких лексем как «уважаемый», «специалист», «профессионал», «эксперт» и др.; тактика актуализации мотива финансовой выгоды, при помощи таких лексем как «скидка», «продажи», «бесплатно», тактика заполнения информационных «пробелов», связанная с не всегда уместным использованием развлекательного контента.

Заключение

В работе рассмотрена возможность использования результатов автоматизированного анализа контента образовательных telegram-каналов для выявления коммуникативных образовательных стратегий. С этой целью

с помощью методов машинного обучения определено к какой тематике в сфере образования относится пост в виде текста. Описана тематическая структура контента образовательных telegram-каналов с высоким и низким рейтингами, а также статистическая информация о дидактическом наполнении образовательных ресурсов. Выявлено, что высокорейтинговые образовательные telegram-каналы реализуют интерактивную коммуникацию, логически связывая всю информацию в единый контент, объединенный основной смысловой стратегией интеграции образовательного и карьерного маршрутов. Образовательные telegram-каналы с низким рейтингом реализуют коммуникативную стратегию, направленную на предоставление узкоспециализированной, логически разобщенной справочной, коммерческой или развлекательной информацией. Одним из маркеров коммуникативных тактик на низкорейтинговых каналах выступают приемы манипулятивного воздействия, позволяющие реализовывать влияние на мнение аудитории: тактика косвенного убеждения, тактика актуализации мотива финансовой выгоды, тактика заполнения информационных «пробелов». Предложенный метод может использоваться для анализа и мониторинга образовательной коммуникации в сети Интернет широким кругом специалистов в сфере менеджмента образования и педагогами, работающим в виртуальной среде.

Литература

1. Butnaru G.I., Niță V., Anichiti A., Brînză G. The effectiveness of online education during Covid 19 pandemic — a comparative analysis between the perceptions of academic students and high school students from Romania // *Sustainability*. 2021. V. 13. N 9. P. 5311. <https://doi.org/10.3390/su13095311>
2. Elfahal H.S., Saeid E., Elfatih Y. Assessment of the applicability of using Telegram as a learning management system // *International Journal of Information Technology and Language Studies*. 2021. V. 5. N 3. P. 1–7.
3. Hrastinski S. What do we mean by blended learning? // *TechTrends*. 2019. V. 63. N 5. P. 564–569. <https://doi.org/10.1007/s11528-019-00375-5>
4. Nosenko O., Nosenko Y., Shevchuk R. Telegram messenger for supporting educational process under the conditions of quarantine restrictions // *Communications in Computer and Information Science*. 2022. V. 1635. P. 308–319. https://doi.org/10.1007/978-3-031-14841-5_20
5. Swartz B.C., Valentine L.Z., Jaftha D.V. Participatory parity through teaching with Telegram // *Perspectives in Education*. 2022. V. 40. N 1. P. 96–111. <https://doi.org/10.18820/2519593X/pie.v40.i1.6>
6. Cabral L., Monteiro J.M., da Silva J.W.F., Mattos C.L., Mourão P.J.C. FakeWhatsApp.BR: NLP and machine learning techniques for misinformation detection in Brazilian Portuguese WhatsApp messages // *Proc. of the 23rd International Conference on Enterprise Information Systems (ICEIS 2021)*. P. 63–74. <https://doi.org/10.5220/0010446800630074>
7. Júnior M., Melo P., Kansaon D., Mafra V., Sa K., Benevenuto F. Telegram monitor: Monitoring Brazilian political groups and channels on Telegram // *HT'22: Proc. of the 33rd ACM Conference on Hypertext and Social Media*. 2022. P. 228–231. <https://doi.org/10.1145/3511095.3536375>
8. Bodrunova S., Nepiyushchik D. Dynamics of distrust, aggression, and conspiracy thinking in the anti-vaccination discourse on Russian Telegram // *Lecture Notes in Computer Science*. 2022. V. 13315. P. 468–484. https://doi.org/10.1007/978-3-031-05061-9_33

References

1. Butnaru G.I., Niță V., Anichiti A., Brînză G. The effectiveness of online education during Covid 19 pandemic — a comparative analysis between the perceptions of academic students and high school students from Romania. *Sustainability*, 2021, vol. 13, no. 9, pp. 5311. <https://doi.org/10.3390/su13095311>
2. Elfahal H.S., Saeid E., Elfatih Y. Assessment of the applicability of using Telegram as a learning management system. *International Journal of Information Technology and Language Studies*, 2021, vol. 5, no. 3, pp. 1–7.
3. Hrastinski S. What do we mean by blended learning? *TechTrends*, 2019, vol. 63, no. 5, pp. 564–569. <https://doi.org/10.1007/s11528-019-00375-5>
4. Nosenko O., Nosenko Y., Shevchuk R. Telegram messenger for supporting educational process under the conditions of quarantine restrictions. *Communications in Computer and Information Science*, 2022, vol. 1635, pp. 308–319. https://doi.org/10.1007/978-3-031-14841-5_20
5. Swartz B.C., Valentine L.Z., Jaftha D.V. Participatory parity through teaching with Telegram. *Perspectives in Education*, 2022, vol. 40, no. 1, pp. 96–111. <https://doi.org/10.18820/2519593X/pie.v40.i1.6>
6. Cabral L., Monteiro J.M., da Silva J.W.F., Mattos C.L., Mourão P.J.C. FakeWhatsApp.BR: NLP and machine learning techniques for misinformation detection in Brazilian Portuguese WhatsApp messages. *Proc. of the 23rd International Conference on Enterprise Information Systems (ICEIS 2021)*, pp. 63–74. <https://doi.org/10.5220/0010446800630074>
7. Júnior M., Melo P., Kansaon D., Mafra V., Sa K., Benevenuto F. Telegram monitor: Monitoring Brazilian political groups and channels on telegram. *HT'22: Proc. of the 33rd ACM Conference on Hypertext and Social Media*, 2022, pp. 228–231. <https://doi.org/10.1145/3511095.3536375>
8. Bodrunova S., Nepiyushchik D. Dynamics of distrust, aggression, and conspiracy thinking in the anti-vaccination discourse on Russian Telegram. *Lecture Notes in Computer Science*, 2022, vol. 13315, pp. 468–484. https://doi.org/10.1007/978-3-031-05061-9_33

9. Iksan Z.H., Saufian S.M. Mobile learning: Innovation in Teaching and learning using Telegram // *International Journal of Pedagogy and Teacher Education*. 2017. V. 1. N 1. P. 19–26. <https://doi.org/10.20961/ijpte.v1i1.5120>
10. Molina O.E. The effects of WhatsApp and Telegram on student engagement: An analysis from the mixed-methods approach // *Education Research International*. 2022. P. 2881404. <https://doi.org/10.1155/2022/2881404>
11. Patahuddin P., Yuliati Y., Syawal S. Evaluating Telegram application to empower the students' vocabulary mastery // *International Journal of English Linguistics*. 2022. V. 12. N 4. P. 106. <https://doi.org/10.5539/ijel.v12n4p106>
12. Soon M.K.S., Martinengo L., Lu J., Car L.T., Chia C.L.K. The use of Telegram in surgical education: Exploratory study // *JMIR Medical Education*. 2022. V. 8. N 3. P. e35983. <https://doi.org/10.2196/35983>
13. Alharbi A., Sultan N.Y. Jeddah University Students' Attitude toward Using Telegram to Support their EFL Learning. 2022 [Электронный ресурс]. URL: https://www.researchgate.net/publication/361567240_Jeddah_University_Students%27_Attitude_toward_Using_Telegram_to_Support_their_EFL_Learning (дата обращения: 01.11.2022)
14. Utomo M.R., Cahyaningrum N., Aliah N., Nafi Z.C. A systematic review: Telegram as learning media in teaching english // *Proc. of the Conference on English Language Teaching*. V. 1. 2021. P. 32–40. <https://doi.org/10.24090/celti.v1.4>
15. Ляховенко О.И. Телеграм-каналы в системе экспертной и политической коммуникации в современной России // *Galactica Media: Journal of Media Studies*. 2022. T. 4. № 1. С. 114–144. <https://doi.org/10.46539/gmd.v4i1.230>
16. Muharom S., Ramadhan A., Jue A.G.K. Design and build a Telegram — Based infusion droplet control and monitoring system // *Procedia of Engineering and Life Science*. 2022. V. 2. N 2. <https://doi.org/10.21070/pels.v2i2.1225>
17. Hakim D.K., Nugroho S.A. Implementasi Telegram bot untuk monitoring mikrotik router // *Sainteks*. 2020. V. 16. N 2. <https://doi.org/10.30595/st.v16i2.7132>
18. Богдановская И.М., Королева Н.Н., Углова А.Б. Психологические факторы доверия к популярным видеоблогерам у современной молодежи // *Психология. Журнал Высшей школы экономики*. 2021. T. 18. № 3. С. 451–467. <https://doi.org/10.17323/1813-8918-2021-3-451-467>
19. Романова Е.В., Калаврий Т.Ю. Анализ реакции студентов на изменение финансового положения в период пандемии по цифровым следам в социальной сети ВКонтакте // *Вестник Северо-Восточного федерального университета им. М.К. Аммосова. Серия: Экономика. Социология. Культурология*. 2021. № 4(24). С. 54–64. <https://doi.org/10.25587/SVFU.2021.24.4.006>
9. Iksan Z.H., Saufian S.M. Mobile learning: Innovation in Teaching and learning using Telegram. *International Journal of Pedagogy and Teacher Education*, 2017, vol. 1, no. 1, pp. 19–26. <https://doi.org/10.20961/ijpte.v1i1.5120>
10. Molina O.E. The effects of WhatsApp and Telegram on student engagement: An analysis from the mixed-methods approach. *Education Research International*, 2022, pp. 2881404. <https://doi.org/10.1155/2022/2881404>
11. Patahuddin P., Yuliati Y., Syawal S. Evaluating Telegram application to empower the students' vocabulary mastery. *International Journal of English Linguistics*, 2022, vol. 12, no. 4, pp. 106. <https://doi.org/10.5539/ijel.v12n4p106>
12. Soon M.K.S., Martinengo L., Lu J., Car L.T., Chia C.L.K. The use of Telegram in surgical education: Exploratory study. *JMIR Medical Education*, 2022, vol. 8, no. 3, pp. e35983. <https://doi.org/10.2196/35983>
13. Alharbi A., Sultan N.Y. *Jeddah University Students' Attitude toward Using Telegram to Support their EFL Learning*. 2022. Available at: https://www.researchgate.net/publication/361567240_Jeddah_University_Students%27_Attitude_toward_Using_Telegram_to_Support_their_EFL_Learning (accessed: 01.11.2022)
14. Utomo M.R., Cahyaningrum N., Aliah N., Nafi Z.C. A systematic review: Telegram as learning media in teaching English. *Proc. of the Conference on English Language Teaching*. V. 1, 2021, pp. 32–40. <https://doi.org/10.24090/celti.v1.4>
15. Lyakhovenko O.I. Telegram channels in the system of expert and political communication in modern Russia. *Galactica Media: Journal of Media Studies*, 2022, vol. 4, no. 1, pp. 114–144. (in Russian). <https://doi.org/10.46539/gmd.v4i1.230>
16. Muharom S., Ramadhan A., Jue A.G.K. Design and build a Telegram — Based infusion droplet control and monitoring system. *Procedia of Engineering and Life Science*, 2022, vol. 2, no. 2. <https://doi.org/10.21070/pels.v2i2.1225>
17. Hakim D.K., Nugroho S.A. Implementasi Telegram bot untuk monitoring mikrotik router. *Sainteks*, 2020, vol. 16, no. 2. <https://doi.org/10.30595/st.v16i2.7132>
18. Bogdanovskaya I.M., Uglova A.B., Koroleva N.N. Psychological factors of trust in popular video bloggers among modern youth. *Psychology. Journal of the Higher School of Economics*, 2021, vol. 18, no. 3, pp. 451–467. (in Russian). <https://doi.org/10.17323/1813-8918-2021-3-451-467>
19. Romanova E.V., Kalavriy T.Y. An analysis of students' response to changes in financial situation during the pandemic by digital traces in the VKontakte social network. *Vestnik of North-Eastern Federal University. «Economics. Sociology. Culturology» series*, 2021, no. 4(24), pp. 54–64. (in Russian). <https://doi.org/10.25587/SVFU.2021.24.4.006>

Авторы

Низомутдинов Борис Абдуллохонвич — ведущий аналитик, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, <https://orcid.org/0000-0002-4090-9564>, boris@itmo.ru

Углова Анна Борисовна — кандидат психологических наук, доцент, Российский государственный педагогический университет им. А. И. Герцена, Санкт-Петербург, 191186, Российская Федерация, <https://orcid.org/0000-0002-8072-0539>, anna.uglova@list.ru

Богдановская Ирина Марковна — кандидат психологических наук, доцент, доцент, Российский государственный педагогический университет им. А. И. Герцена, Санкт-Петербург, 191186, Российская Федерация, <https://orcid.org/0000-0001-7303-615X>, ibogdanovs@herzen.spb.ru

Authors

Boris A. Nizomutdinov — Leading Analyst, ITMO University, Saint Petersburg, 197101, Russian Federation, <https://orcid.org/0000-0002-4090-9564>, boris@itmo.ru

Anna B. Uglova — PhD (Psychology), Associate Professor, Herzen State Pedagogical University of Russia, Saint Petersburg, 191186, Russian Federation, <https://orcid.org/0000-0002-8072-0539>, anna.uglova@list.ru

Irina M. Bogdanovskaya — PhD (Psychology), Associate Professor, Associate Professor, Herzen State Pedagogical University of Russia, Saint Petersburg, 191186, Russian Federation, <https://orcid.org/0000-0001-7303-615X>, ibogdanovs@herzen.spb.ru

Статья поступила в редакцию 10.10.2022
Одобрена после рецензирования 14.12.2022
Принята к печати 29.01.2023

Received 10.10.2022
Approved after reviewing 14.12.2022
Accepted 29.01.2023



Работа доступна по лицензии
Creative Commons
«Attribution-NonCommercial»

МАТЕМАТИЧЕСКОЕ И КОМПЬЮТЕРНОЕ МОДЕЛИРОВАНИЕ MODELING AND SIMULATION

doi: 10.17586/2226-1494-2023-23-1-150-160

УДК 004.94

Компьютерное моделирование немарковских процессов на основе принципа баланса «комплексных вероятностей»

Ярослав Николаевич Гусеница¹✉, Олег Анатольевич Ширямов²

¹ Военный инновационный технополис «ЭРА», Анапа, 353456, Российская Федерация

² Военно-космическая академия А.Ф. Можайского, Санкт-Петербург, 197198, Российская Федерация

¹ yaromir226@gmail.com✉, <https://orcid.org/0000-0003-4757-5066>

² shiryamov_oa@mail.ru, <https://orcid.org/0000-0002-8985-8251>

Аннотация

Предмет исследования. Значительная часть исследований эффективности различных систем посвящена изучению их функционирования в стационарном режиме. Однако с точки зрения их практического применения представляют интерес исследования функционирования таких систем с изменяющейся интенсивностью рабочей нагрузки в переходных, нестационарных режимах работы. В отличие от моделей исследования нестационарных систем, существенно опирающихся на статические значения распределений, в данной работе предложена модель с использованием произвольных распределений вероятностей во времени. **Метод.** Математическая формализация модели основана не на применении классической дифференциальной модели во временной области, а на формальном представлении вероятностей состояний систем в преобразовании Лапласа, т. е. в комплексном виде. Определение значений вероятностей состояний систем основывается на принципе баланса «комплексных вероятностей», что позволяет разрабатывать модели нестационарных систем обслуживания с произвольными распределениями вероятностей времени поступления заявок и их обслуживания, учитывающие случайные или детерминированные временные задержки. Для оперативного расчета систем предлагается использовать разработанное приложение с графическим пользовательским интерфейсом. **Основные результаты.** Представлена архитектура данного приложения в виде диаграммы пакетов. Показан алгоритм работы приложения. Выполнено сравнение работы приложения с программами для решения задач технических вычислений MATLAB и MathCad при моделировании процесса функционирования эталона единицы величины и системы управления роботом. Приведены преимущества использования разработанного приложения. **Практическая значимость.** Представленные результаты могут быть применены специалистами, занимающимися исследованиями эффективности различных систем.

Ключевые слова

немарковский процесс, принцип баланса, компьютерное моделирование, изображение Лапласа, Python, граф состояния

Благодарности

Авторы работы выражают благодарность своему научному наставнику — действительному члену Международной академии информатизации, заслуженному деятелю науки Российской Федерации, доктору технических наук, профессору Владимиру Александровичу Смагину.

Ссылка для цитирования: Гусеница Я.Н., Ширямов О.А. Компьютерное моделирование немарковских процессов на основе принципа баланса «комплексных вероятностей» // Научно-технический вестник информационных технологий, механики и оптики. 2023. Т. 23, № 1. С. 150–160. doi: 10.17586/2226-1494-2023-23-1-150-160

Computer modeling of non-Markovian processes based on the principle of balance of “complex probabilities”

Yaroslav N. Gusenitsa¹✉, Oleg A. Shiryamov²

¹ Military Innovative Technopolis “ERA”, Anapa, 353456, Russian Federation

² Mozhaisky Military Aerospace Academy, Saint Petersburg, 197198, Russian Federation

¹ yaromir226@gmail.com✉, <https://orcid.org/0000-0003-4757-5066>

² shiryamov_oa@mail.ru, <https://orcid.org/0000-0002-8985-8251>

Abstract

A significant part of the research on the effectiveness of various systems is devoted to the study of their functioning in a stationary mode. However, from the point of view of their practical application, it is of interest to study the functioning of such systems with varying workload intensity in transient, non-stationary modes of operation. And unlike the models for studying non-stationary systems, which are essentially based on the static values of distributions, this paper proposes a model using arbitrary probability distributions over time. The mathematical formalization of the model is based not on the application of the classical differential model in the time domain, but on the formal representation of the probabilities of the system states in the Laplace transform, i.e., in a complex way. Determining the values of the probabilities of the systems states is based on the principle of balance of “complex probabilities” which allows developing models of non-stationary queuing systems with arbitrary probability distributions of the arrival time of requests and their service, taking into account random or deterministic time delays. For the operational calculation of systems, it is proposed to use the developed application with a graphical user interface. The architecture of this application is presented in the form of a package diagram. The algorithm of the application is shown. Comparison of the application operation with programs MATLAB and MathCad for solving the problems of technical calculations was made when modeling the process of functioning of the standard unit of quantity and the robot control system. The advantages of using the developed application are given. The presented results can be applied by specialists involved in research on the effectiveness of various systems.

Keywords

non-Markovian process, balance principle, computer simulation, Laplace image, Python, state graph

Acknowledgements

The authors of the work express their gratitude to their scientific mentor — member of the International Academy of Informatization, Honored Scientist of the Russian Federation, Doctor of Technical Sciences, Professor Vladimir Alexandrovich Smagin.

For citation: Gusenitsa Ya.N., Shiryamov O.A. Computer modeling of non-Markovian processes based on the principle of balance of “complex probabilities”. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 2023, vol. 23, no. 1, pp. 150–160 (in Russian). doi: 10.17586/2226-1494-2023-23-1-150-160

Введение

В настоящее время при разработке сложных систем моделирование, как правило, остается единственным способом оценки показателей качества до момента изготовления. Кроме того, применение моделирования на начальных стадиях жизненного цикла систем позволяет значительно сократить сроки проведения испытаний и уменьшить финансовые и материальные затраты.

Важным аспектом в данной проблематике является создание моделей, которые позволяют описывать с высокой степенью адекватности различные аспекты поведения систем на всех стадиях жизненного цикла, учитывать изменяющуюся интенсивность рабочей нагрузки в переходных и нестационарных режимах работы.

Модели функционирования систем с изменяющейся интенсивностью рабочей нагрузки и методы их расчета рассмотрены в [1, 2]. Однако в практическом применении они сложны и громоздки.

Подобные математические структуры получили название моделей нестационарных систем массового обслуживания. В [3] изучены нестационарные одноканальные и многоканальные модели с различными видами распределений между поступающими заявками и длительностями их обслуживания. В [4] предложено расширение свойств нестационарной модели за

счет учета в ней времени задержки на обслуживание, построен соответствующий процессу граф, записана и решена система дифференциальных уравнений. При этом предложенные модели являются марковскими, так как в них используются экспоненциальные законы распределения и постоянное значение времени задержки.

В работах [5, 6] предложен принцип баланса «комплексных вероятностей» для моделирования нестационарных систем массового обслуживания, представленных в виде ациклического графа переходов состояний. В [7, 8] данный принцип применен для моделирования нестационарных систем массового обслуживания, рассмотренных в виде замкнутой системы переходов состояний. В [9] проведена систематизация и обобщение принципа баланса «комплексных вероятностей» для разработки моделей стохастических процессов с произвольными распределениями вероятностей времени поступления заявок и их обслуживания, учитывающих случайные или детерминированные временные задержки.

Учитывая, что в рассмотренных работах применение принципа баланса «комплексных вероятностей» в инженерной практике не до конца раскрыт, в рамках настоящей работы представлены предложения по программной реализации компьютерного моделирования немарковских процессов.

Основные положения принципа баланса «комплексных вероятностей»

Термин «комплексные вероятности» впервые введен Дэвидом Коксом [10]. В работе [11] отмечено, что «...применив метод «комплексных вероятностей» Кокса, мы можем охватить случай, когда плотность длительности обслуживания при $x > 0$ задается любой функцией вида»:

$$p(x) = \sum_i C_i x^{r_i} e^{-\lambda_i x},$$

где C_i — i -я комплексно-сопряженная переменная; r_i — количество появлений λ_i в ряду $\lambda_1, \dots, \lambda_{i-1}$; λ_i — интенсивность обслуживания в i -м состоянии.

«...лишь бы она была неотрицательной и имела интеграл, равный 1». Кроме того, «...любые законы распределения длительности обслуживания допускают аппроксимацию суммой экспонент с полиномиальными множителями».

Данная идея была применена в [12] при разложении распределений вероятностей на сумму экспоненциальных плотностей с комплексными сопряженными коэффициентами и параметрами с целью решения исследования немарковских процессов. В работе [13] изучен вопрос о вероятностном анализе комплексной переменной с введением комплексной дельта-функции Дирака. В работе [14] приведено строгое обоснование комплексных функций Хевисайда и Дирака, а также показан численный пример приложения дельта-функции к исследованию альтернирующего случайного процесса с накоплением и потерей информации.

Результаты перечисленных исследований составляют основу принципа баланса «комплексных вероятностей».

В общем случае под балансом вероятностей понимается равенство сумм произведений значений интенсивностей переходов на вероятности состояний — отправители заявок (начальные вершины дуг графа состояний) и сумм произведений значений интенсивностей переходов на вероятности состояний — получатели заявок (конечные вершины дуг графа состояний). Таким образом, состояние процесса определяется как равновесное, когда среднестатистические характеристики входящих и выходящих случайных потоков взаимно уравновешиваются. Полученные уравнения для состояний процесса необходимо решить, чтобы найти стационарные значения вероятности состояний. Во временной области математическая формализация процесса основывается на действиях умножения независимых интенсивностей, вероятностей и сложении полученных произведений несовместных событий состояний процесса.

В случае баланса «комплексных вероятностей» стохастический процесс описывается не во временной, а в комплексной области. Для этого применяется преобразование Лапласа, позволяющее представить систему дифференциальных или интегральных уравнений стохастического процесса в виде системы алгебраических уравнений [9].

При использовании баланса «комплексных вероятностей» временные характеристики (вероятности

состояний) заменяются их изображениями в преобразовании Лапласа. Попаданию системы в некоторое состояние предшествует суммирование случайных временных интервалов траектории случайного процесса. В преобразовании Лапласа суммирование траекторий соответствует произведению изображений плотностей вероятностей. В результате для составления уравнений баланса состояний необходимо использовать произведение изображений с их последующим суммированием [9].

Пусть требуется найти вероятности состояний двухфазного немарковского процесса. Для исследования баланса «комплексных вероятностей» изобразим граф переходов (рис. 1).

Данный граф содержит переходы (стрелки, направленные вправо), связанные не с интенсивностями поступления в систему прибывающих заявок, а с изображениями по Лапласу плотностей вероятностей времени между заявками $a^*(s)$, где «*» — символ преобразования; s — комплексная переменная Лапласа.

Переходы (стрелки, направленные влево), связаны не с интенсивностями обслуживания прибывших заявок, а с изображениями по Лапласу плотностей вероятностей времен их обслуживания — $b^*(s)$. Кроме того, при ветвлении случайного процесса используются условные вероятности переходов α и β . Благодаря такому представлению происходит переход от марковских случайных процессов к немарковским. Отметим, что для перехода требуется не модель дифференциальных уравнений типа «гибели и размножения», а другая математическая форма — модель «комплексного представления вероятностей».

При данных условиях составим систему уравнений баланса:

$$\begin{cases} a^*(s)P_0^*(s) = \beta b^*(s)P_1^*(s); \\ (\alpha a^*(s) + \beta b^*(s))P_1^*(s) = a^*(s)P_0^*(s) + b^*(s)P_2^*(s); \\ b^*(s)P_2^*(s) = \alpha a^*(s)P_1^*(s). \end{cases} \quad (1)$$

Примем во внимание условие нормирования изображений вероятностей

$$\sum_{i=0}^2 P_i^*(s) = \frac{1}{s},$$

где $P_i^*(s)$ — комплексная вероятность нахождения эталона единицы величины в i -м состоянии, и решив систему уравнений (1), получим:

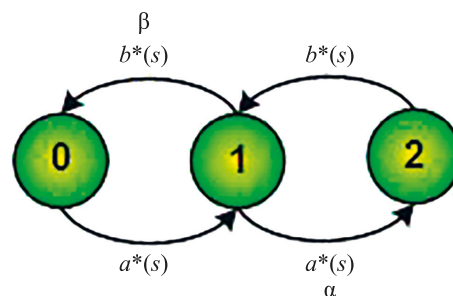


Рис. 1. Граф переходов нестационарной системы обслуживания

Fig. 1. Graph of transitions of a non-stationary queuing system

$$\begin{aligned}
P_0^*(s) &= \frac{\beta(b^*(s))^2}{s(\beta(b^*(s))^2 + a^*(s)b^*(s) + \alpha(a^*(s))^2)}, \\
P_1^*(s) &= \frac{a^*(s)b^*(s)}{s(\beta(b^*(s))^2 + a^*(s)b^*(s) + \alpha(a^*(s))^2)}, \\
P_2^*(s) &= \frac{\alpha(a^*(s))^2}{s(\beta(b^*(s))^2 + a^*(s)b^*(s) + \alpha(a^*(s))^2)}.
\end{aligned} \quad (2)$$

Пусть интервалы времени поступления заявок и их обслуживания подчиняются нормальному закону распределения.

Примем следующие исходные данные:

- распределение времени t между поступающими заявками нормальное с плотностью вероятности

$$a(t) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(t-m)^2}{2\sigma^2}},$$

математическим ожиданием $m = 20$ ч, среднеквадратичным отклонением $\sigma = 5$ ч;

- распределение времени t обслуживания требований нормальное с плотностью вероятности

$$b(t) = \frac{1}{\theta\sqrt{2\pi}} e^{-\frac{(t-n)^2}{2\theta^2}},$$

математическим ожиданием $n = 15$ ч, среднеквадратичным отклонением $\theta = 3$ ч.

Соответствующие функции распределения равны:

$$\begin{aligned}
A(t) &= \int_0^t a(z) dz; \\
B(t) &= \int_0^t b(z) dz.
\end{aligned}$$

Определим условные вероятности переходов при втвлении случайного процесса:

$$\begin{aligned}
\alpha &= \int_0^\infty (1 - B(z)) dA(z) = 0,196; \\
\beta &= \int_0^\infty (1 - A(z)) dB(z) = 0,804.
\end{aligned}$$

Найдем изображения Лапласа плотностей вероятностей, используя гипердельтную аппроксимацию нор-

мальной плотности по четырем начальным моментам, приведенную в [13], которую представим в виде:

$$f_\delta(t) = \frac{1}{2}(\delta(t - m - \sigma) + \delta(t - m + \sigma)), \quad (3)$$

где δ — дельта-функция Дирака; m и σ — параметры аппроксимируемого распределения.

Применив формулу (3), получим аппроксимационные формулы плотностей вероятностей:

$$a(t) \approx \frac{1}{2}(\delta(t - 15) + \delta(t - 25));$$

$$b(t) \approx \frac{1}{2}(\delta(t - 12) + \delta(t - 18)),$$

которые в изображениях Лапласа примут вид:

$$\begin{aligned}
a^*(s) &= \frac{1}{2} \left(e^{-15s} + (e^{-15s})^{\frac{5}{3}} \right); \\
b^*(s) &= \frac{1}{2} \left(e^{-12s} + (e^{-12s})^{\frac{3}{2}} \right).
\end{aligned} \quad (4)$$

Подставим формулы (4) в уравнения (2), получим их в развернутом виде для приведенных числовых данных примера.

Чтобы получить выражения для вероятностей состояний во временной области, необходимо применить обратное преобразование Лапласа. Однако достичь ожидаемых результатов достаточно затруднительно из-за сложности прямых изображений Лапласа искомых вероятностей и вычислительных трудностей работы с нормальными распределениями. В связи с этим переход от изображений $P_0^*(s)$, $P_1^*(s)$, $P_2^*(s)$ к оригиналам $P_0(t)$, $P_1(t)$, $P_2(t)$ произведем, применяя приближенный способ обращения преобразования Лапласа с использованием формулы Алфрея. Формула Алфрея является результатом формулы Уайдера на основе свойства фильтрации преобразования Лапласа с помощью дельта-функции. Она имеет следующий вид [15]:

$$f(t) \approx sf^*(s), \text{ при } s = \frac{1}{t}. \quad (5)$$

В результате применения формулы (5) получим выражения для вероятностей состояний процесса:

$$\begin{aligned}
P_0(t) &= \frac{0,804 \left(\frac{e^{-\frac{12}{t}}}{2} + \frac{\left(e^{-\frac{12}{t}} \right)^{\frac{3}{2}}}{2} \right)}{\left(\frac{e^{-\frac{12}{t}}}{2} + \frac{\left(e^{-\frac{12}{t}} \right)^{\frac{3}{2}}}{2} \right) \left(\frac{e^{-\frac{15}{t}}}{2} + \frac{\left(e^{-\frac{15}{t}} \right)^{\frac{5}{3}}}{2} \right) + 0,804 \left(\frac{e^{-\frac{12}{t}}}{2} + \frac{\left(e^{-\frac{12}{t}} \right)^{\frac{3}{2}}}{2} \right) + 0,196 \left(\frac{e^{-\frac{15}{t}}}{2} + \frac{\left(e^{-\frac{15}{t}} \right)^{\frac{5}{3}}}{2} \right)}, \\
P_1(t) &= \frac{\left(\frac{e^{-\frac{15}{t}}}{2} + \frac{\left(e^{-\frac{15}{t}} \right)^{\frac{5}{3}}}{2} \right) \left(\frac{e^{-\frac{12}{t}}}{2} + \frac{\left(e^{-\frac{12}{t}} \right)^{\frac{3}{2}}}{2} \right)}{\left(\frac{e^{-\frac{12}{t}}}{2} + \frac{\left(e^{-\frac{12}{t}} \right)^{\frac{3}{2}}}{2} \right) \left(\frac{e^{-\frac{15}{t}}}{2} + \frac{\left(e^{-\frac{15}{t}} \right)^{\frac{5}{3}}}{2} \right) + 0,804 \left(\frac{e^{-\frac{12}{t}}}{2} + \frac{\left(e^{-\frac{12}{t}} \right)^{\frac{3}{2}}}{2} \right) + 0,196 \left(\frac{e^{-\frac{15}{t}}}{2} + \frac{\left(e^{-\frac{15}{t}} \right)^{\frac{5}{3}}}{2} \right)},
\end{aligned}$$

$$P_2(t) = \frac{0,196 \left(\frac{e^{-\frac{15}{t}}}{2} + \frac{\left(e^{-\frac{15}{t}} \right)^{\frac{5}{3}}}{2} \right)}{\left(\frac{e^{-\frac{12}{t}}}{2} + \frac{\left(e^{-\frac{12}{t}} \right)^{\frac{3}{2}}}{2} \right) \left(\frac{e^{-\frac{15}{t}}}{2} + \frac{\left(e^{-\frac{15}{t}} \right)^{\frac{5}{3}}}{2} \right) + 0,804 \left(\frac{e^{-\frac{12}{t}}}{2} + \frac{\left(e^{-\frac{12}{t}} \right)^{\frac{3}{2}}}{2} \right)^2 + 0,196 \left(\frac{e^{-\frac{15}{t}}}{2} + \frac{\left(e^{-\frac{15}{t}} \right)^{\frac{5}{3}}}{2} \right)}.$$

Численные результаты представлены в виде графиков зависимостей вероятностей на рис. 2.

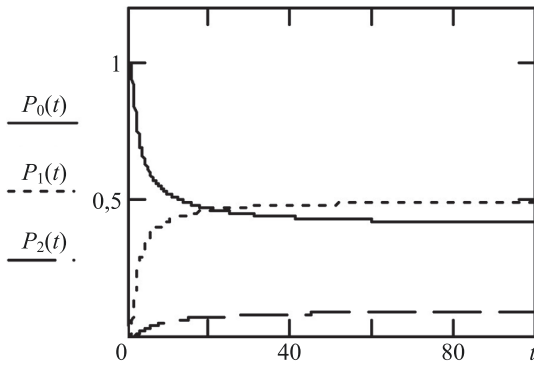


Рис. 2. Графики вероятностей состояний в течение времени
Fig. 2. Graphs of state probability functions over time

Немарковские модели процесса функционирования эталона единицы величины и системы управления роботом

В работах [16, 17] исследован процесс функционирования нижестоящего по точности эталона единицы величины средства измерений в условиях увеличения нагрузки.

В течение времени t с определенной интенсивностью поступает поток нижестоящих по точности N

эталонов на метрологическое обслуживание. Эталонам передается значение единицы величины с определенной интенсивностью. Таким образом, процесс функционирования эталона единицы величины может быть представлен с помощью графа переходов в изображениях Лапласа (рис. 3).

На рис. 3 приняты следующие обозначения: $a_i^*(s)$ и $b^*(s)$ — плотности вероятности поступления i -го эталона на метрологическое обслуживание и передачи единицы величины, представленных в изображении Лапласа; α_i и β_i — условные вероятности перехода, которые определяют выбор дальнейшего движения случайного процесса из состояний графа.

Для графа переходов (рис. 3) составим систему уравнений баланса:

$$\begin{cases} a_0^*(s)P_0^*(s) = \beta_1 b^*(s)P_1^*(s) \\ (\alpha_1 a_1^*(s) + \beta_1 b^*(s))P_1^*(s) = a_0^*(s)P_0^*(s) + \beta_2 b^*(s)P_2^*(s) \\ \dots \\ b^*(s)P_N^*(s) = \alpha_{N-1} a_{N-1}^*(s)P_{N-1}^*(s) \end{cases} \quad (6)$$

Принимая во внимание условие нормирования изображений вероятностей состояний и, решив систему уравнений (6), получим общую формулу для вычисления комплексной вероятности состояния, в которой всем поступившим эталонам будет передана единица величины:

$$P_0^*(s) = \frac{\left(\prod_{i=1}^{N-1} \beta_i \right) (b^*(s))^N}{s \left(\left(\prod_{i=1}^{N-1} \beta_i \right) (b^*(s))^N + \prod_{i=1}^{N-1} \alpha_i \prod_{i=0}^{N-1} a_i^*(s) + \sum_{i=0}^{N-2} \alpha_i a_i^*(s) \beta_{i+2} b^*(s) \right)}.$$

Учитывая, что вероятности переходов между состояниями подчиняются нормальному закону распределе-

ния, запишем аналитическое выражение для вероятности передачи единицы величины эталоном:

$$P_0(t) = \frac{\left(\prod_{i=1}^{N-1} \beta_i \right) \left(\frac{1}{2} \left(e^{-\frac{(m^b - \sigma^b)}{t}} + \left(e^{-\frac{-(m^b + \sigma^b)}{t}} \right)^{\frac{m^b + \sigma^b}{m^b - \sigma^b}} \right) \right)^N}{\left(\prod_{i=1}^{N-1} \beta_i \right) \left(\frac{1}{2} \left(e^{-\frac{(m^b - \sigma^b)}{t}} + \left(e^{-\frac{-(m^b + \sigma^b)}{t}} \right)^{\frac{m^b + \sigma^b}{m^b - \sigma^b}} \right) \right)^N + \prod_{i=1}^{N-1} \alpha_i \prod_{i=0}^{N-1} \left(\frac{1}{2} \left(e^{-\frac{(m_i^a - \sigma_i^a)}{t}} + \left(e^{-\frac{-(m_i^a + \sigma_i^a)}{t}} \right)^{\frac{m_i^a + \sigma_i^a}{m_i^a - \sigma_i^a}} \right) \right)} + \sum_{i=0}^{N-2} \alpha_i \left(\frac{1}{2} \left(e^{-\frac{(m_i^a - \sigma_i^a)}{t}} + \left(e^{-\frac{-(m_i^a + \sigma_i^a)}{t}} \right)^{\frac{m_i^a + \sigma_i^a}{m_i^a - \sigma_i^a}} \right) \right) \beta_{i+2} \left(\frac{1}{2} \left(e^{-\frac{(m^b - \sigma^b)}{t}} + \left(e^{-\frac{-(m^b + \sigma^b)}{t}} \right)^{\frac{m^b + \sigma^b}{m^b - \sigma^b}} \right) \right) \quad (7)$$

где α_i и β_i — условные вероятности переходов, которые определяют выбор дальнейшего движения случайного процесса из состояний графа; m^b и σ^b — математическое ожидание и среднее квадратическое отклонение вероятности перехода в предыдущее состояние; m_i^a

и σ_i^a — математическое ожидание и среднее квадратическое отклонение вероятности перехода в $i + 1$ состояние.

В [18, 19] исследован процесс функционирования системы управления роботом в условиях динамиче-

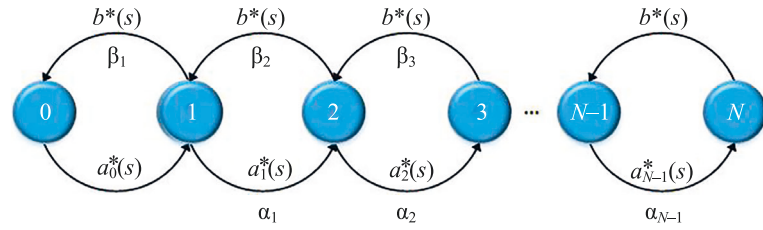


Рис. 3. Граф переходов процесса функционирования эталона единицы величины

Fig. 3. Graph of transitions for the standard unit of magnitude functioning process

ски изменяющихся значений параметров окружающей среды.

Процесс функционирования системы управления роботом рассмотрим в следующих состояниях систем управления: A_{00} — не идентифицирует изменения значений параметров окружающей среды и не формирует командную информацию; A_{10} — идентифицирует изменения значений параметров окружающей среды, но не формирует командную информацию; A_{01} — не идентифицирует изменения значений параметров окружающей среды, но формирует командную информацию; A_{11} — идентифицирует изменения значений параметров окружающей среды и формирует командную информацию.

Для объекта исследования получим граф переходов в изображениях Лапласа (рис. 4).

В соответствии с рис. 4 составим систему уравнений баланса:

$$\begin{cases} a_0^*(s)P_{00}^*(s) = \beta_2 b_2^*(s)P_{01}^*(s) \\ a_1^*(s)P_{10}^*(s) = a_0^*(s)P_{00}^*(s) + \alpha_3 a_3^*(s)P_{11}^*(s) \\ (\alpha_2 a_2^*(s) + \beta_2 b_2^*(s))P_{01}^*(s) = a_1^*(s)P_{10}^*(s) + \beta_3 b_3^*(s)P_{11}^*(s) \\ (\alpha_3 a_3^*(s) + \beta_3 b_3^*(s))P_{11}^*(s) = \alpha_2 a_2^*(s)P_{01}^*(s) \end{cases}$$

Решив систему уравнений (7), получим общую формулу для вычисления комплексной вероятности

$$P_{00}^*(s) = \frac{a_1^*(s)\beta_2 b_2^*(s)(\alpha_3 a_3^*(s) + \beta_3 b_3^*(s))}{s((\alpha_3 a_3^*(s) + \beta_3 b_3^*(s))((\beta_2 b_2^*(s))(a_1^*(s) + a_0^*(s)) + a_0^*(s)a_1(s)) + (a_0^*(s)\alpha_2 a_2^*(s))(\alpha_3 a_3^*(s) + a_1^*(s)))}$$

Учитывая, что вероятности переходов между состояниями подчиняются экспоненциальному закону распределения, составим аналитическое выражение для вероятности того, что на произвольные изменения

$$P_{00}(s) = \frac{a_1(t)\beta_2 b_2(t)(\alpha_3 a_3(t) + \beta_3 b_3(t))}{(\alpha_3 a_3(t) + \beta_3 b_3(t))((\beta_2 b_2(t))(a_1(t) + a_0(t)) + a_0(t)a_1(t)) + (a_0(t)\alpha_2 a_2(t))(\alpha_3 a_3(t) + a_1(t))}$$

где α_i и β_i — условные вероятности переходов, которые определяют выбор дальнейшего движения случайного процесса из состояний графа; $\alpha_i(t)$ и $\beta_i(t)$ — вероятности перехода из i -го в i -е состояние.

Программное решение

На основе рассмотренного в разделе «Основные положения принципа баланса «комплексных вероятностей» разработано приложение с графическим пользовательским интерфейсом «Система компьютерного моделирования немарковских процессов». Приложение

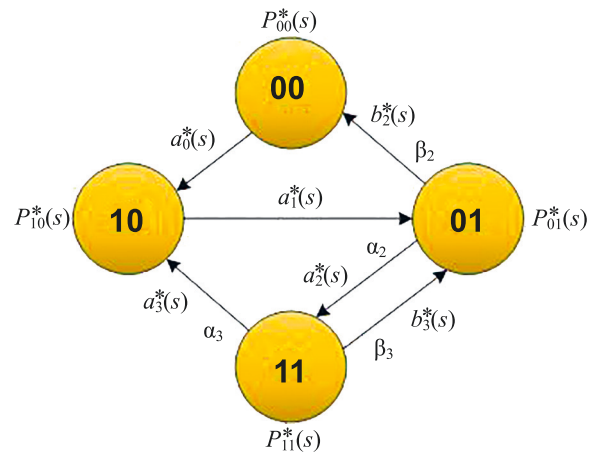


Рис. 4. Граф переходов процесса функционирования системы управления роботом

Fig. 4. Graph of transitions of the robot control system functioning process

состояния, в котором на произвольные изменения значений параметров окружающей среды сформирована командная информация:

значений параметров окружающей среды сформирована командная информация для остальных подсистем робота:

написано на языке Python (версии 3.10) с использованием библиотек SciPy (1.9.1), Kivy (2.0.0), SymPy (1.10.1), Matplotlib (3.6.0), Mpmath (1.2.1) и Pillow (8.4.0) [20–25]. Архитектура приложения в виде диаграммы пакетов представлена на рис. 5.

Программа (ср) состоит из двух блоков: пользовательского интерфейса (frontend), реализующего взаимодействие с оператором; расчетного блока (backend), осуществляющего вычисления.

Модуль main.py является входной точкой приложения и инициализирует его главное окно.

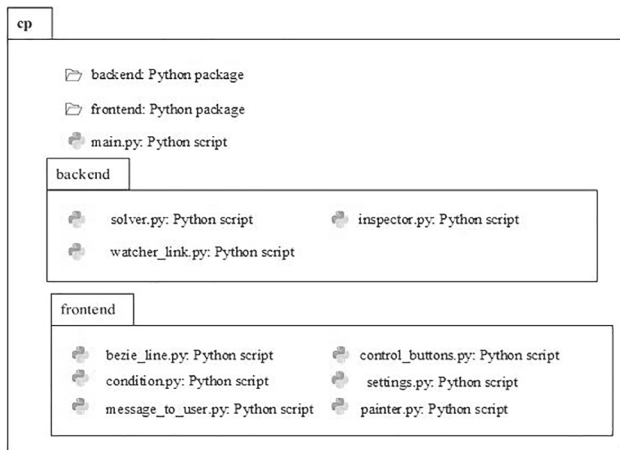


Рис. 5. Диаграмма пакетов приложения «Система компьютерного моделирования немарковских процессов»
Fig. 5. Package diagram of the application “Computer Modeling System for Non-Markovian Processes”



Рис. 6. Алгоритм работы приложения
Fig. 6. Application algorithm

Расчетный блок состоит из модулей solver.py, inspector.py и watcher_link.py, отвечающих за: определение условных вероятностей переходов; гипердельтную аппроксимацию произвольных плотностей распределения; формирование и решение системы уравнений баланса «комплексных вероятностей».

Пользовательский интерфейс включает модули bezier_line.py, condition.py, control_buttons.py, message_to_user.py, painter.py и settings.py, реализующие различные графические элементы приложения, логику их взаимодействия с оператором, обработку вводимых данных, а также обработку исключений.

Алгоритм работы приложения приведен на рис. 6.

Полученное приложение позволяет [26]: создавать в интерактивном режиме состояния и систему переходов между ними для объекта исследований; выбирать произвольные законы распределения и настраивать их числовые характеристики для описания системы переходов между состояниями; сохранять граф системы переходов в формате «*.png»; вычислять и строить графики вероятностей нахождения объекта исследований во всех состояниях.

Компьютерное моделирование процессов функционирования эталона единицы величины и системы управления роботом

Выполним расчетный эксперимент на моделях процессов функционирования эталона единицы величины и системы управления роботом с помощью приложения «Система компьютерного моделирования немарковских процессов», системы MATLAB и среды MathCad с целью получения сравнительной характеристики программных решений для компьютерного моделирования немарковских процессов.

Для модели функционирования эталона единицы величины численные расчеты проведем в MATLAB (версии R2018a). При расчете рассмотрим только первые три состояния. Вероятности между состояниями подчиняются нормальному закону распределения и имеют следующие значения для числовых характеристик: $m_0^1 = 5$ ч; $\sigma_0^1 = 0,2$ ч; $m_1^2 = 3$ ч; $\sigma_0^2 = 0,1$ ч; $m_1^1 = 6$ ч; $\sigma_1 = 0,4$ ч; $m_2 = 2$ ч; $\sigma_2 = 0,2$ ч.

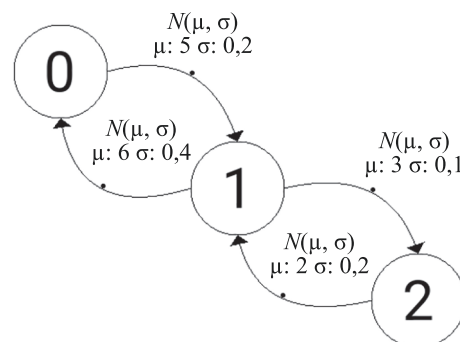


Рис. 7. Граф переходов процесса функционирования эталона единицы величины

Fig. 7. Graph of transitions of the standard unit of magnitude functioning process

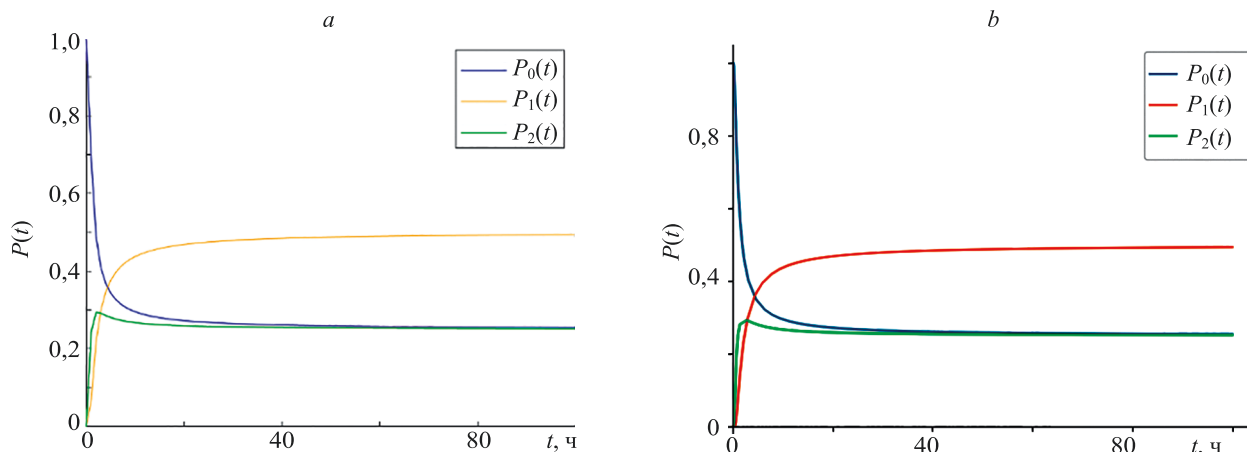


Рис. 8. Графики вероятностей нахождения эталона единицы величины в состояниях, полученные в MATLAB (a) и в разработанном приложении (b)

Fig. 8. Graphs of probabilities of finding a unit of magnitude reference in states obtained in MATLAB (a) and in the developed application (b)

Построим граф переходов в разработанном приложении «Система компьютерного моделирования немарковских процессов» (рис. 7).

На рис. 7 $N(\mu, \sigma)$ обозначает нормальный закон распределения с математическим ожиданием μ и среднеквадратичным отклонением σ .

По результатам расчета получим численные значения вероятности нахождения эталона в начальном состоянии, представленные для удобства в виде графика (рис. 8, a).

Выполним расчет значений вероятностей нахождения эталона во всех состояниях. Результаты представлены на рис. 8, b.

Как видно из рис. 8, результаты, полученные в приложении «Система компьютерного моделирования немарковских процессов», обладают высокой схожимостью с результатами, вычисленными в системе моделирования MATLAB. Однако у разработанного

приложения есть ряд преимуществ (рис. 9): временная сложность расчета значений вероятностей ниже, так как приложение работает с оперативной памятью на более низком уровне; условное время изменения параметров системы значительно ниже за счет пользовательского интерфейса, реализующего взаимодействие с оператором.

Для модели процесса функционирования системы управления роботом для различных состояний численные расчеты проведены в среде MathCad (версии 14.0). При расчете использованы следующие исходные данные для плотностей вероятностей переходов:

- из состояния 0 в состояние 1 подчиняется равномерному закону распределения с числовой характеристикой $m_0^1 = 2$ мс;
- из состояния 1 в состояние 2 подчиняется гамма-распределению с числовыми характеристиками $k_1^2 = 5$ мс и $\lambda_1^2 = 2$ мс;

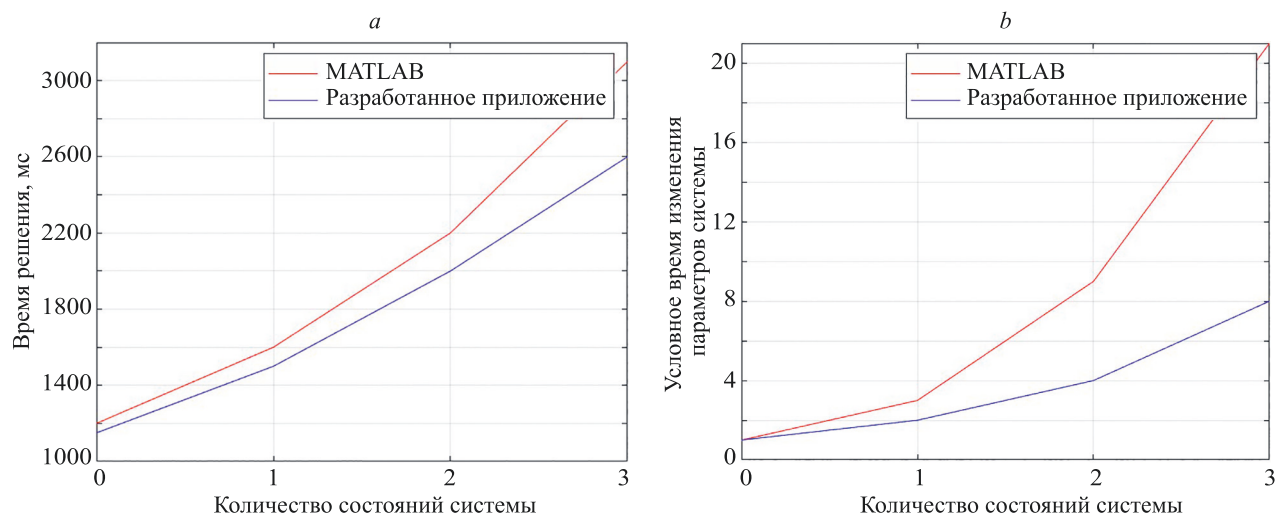


Рис. 9. Сравнительные характеристики разработанного приложения с MATLAB: временная сложность (a) и оперативность изменения начальных условий (b)

Fig. 9. Comparative characteristics of the developed application with MATLAB: time complexity (a) and speed of changing initial conditions (b)

- из состояния 2 в состояние 0 подчиняется экспоненциальному закону распределения с числовой характеристикой $\lambda_2^0 = 3$ мс;
- из состояния 2 в состояние 3 подчиняется нормальному закону распределения с числовыми характеристиками $m_2^3 = 5$ мс и $\sigma_2^3 = 2$ мс;
- из состояния 3 в состояние 1 подчиняется экспоненциальному закону распределения с числовой характеристикой $\lambda_3^1 = 6$ мс;
- из состояния 3 в состояние 2 подчиняется закону Рэлея с числовой характеристикой $\sigma_3^2 = 4$ мс.

Построим в приложении граф переходов в приложении «Система компьютерного моделирования немарковских процессов» (рис. 10).

На рис. 10: $U(a)$ — равномерный закон распределения с числовой характеристикой a ; $\text{Exp}(\lambda)$ — экспоненциальный закон распределения с числовой характеристикой λ ; $\Gamma(k, \theta)$ — экспоненциальный закон

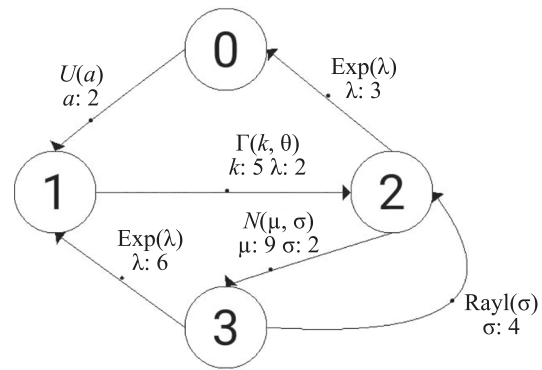


Рис. 10. Граф переходов процесса функционирования системы управления роботом

Fig. 10. Graph of transitions of the robot control system functioning process

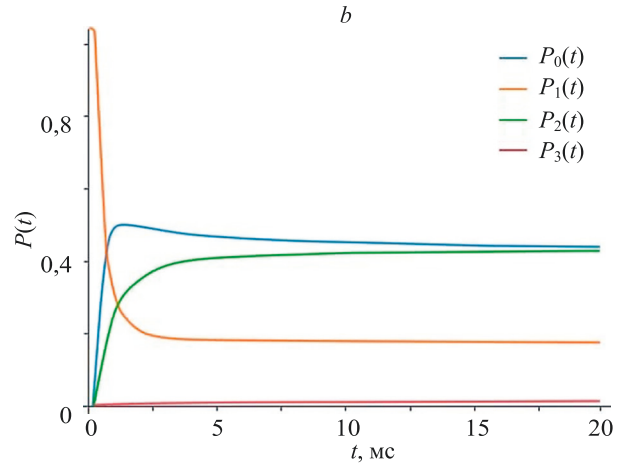
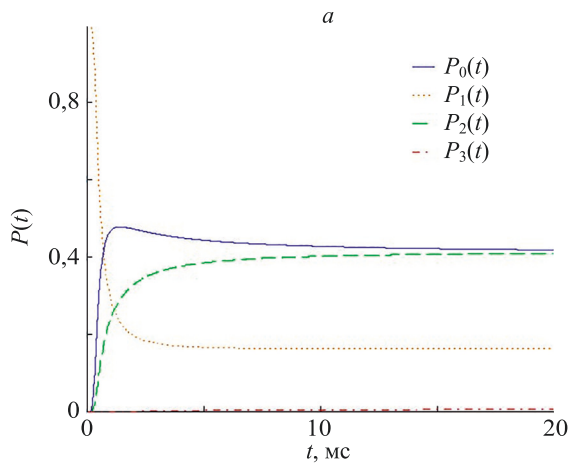


Рис. 11. Графики вероятностей нахождения системы управления роботом в состояниях, полученные в MathCad (a) и в разработанном приложении (b)

Fig. 11. Graphs of probabilities of finding the robot control system in states obtained in MathCad (a) and in the developed application (b)

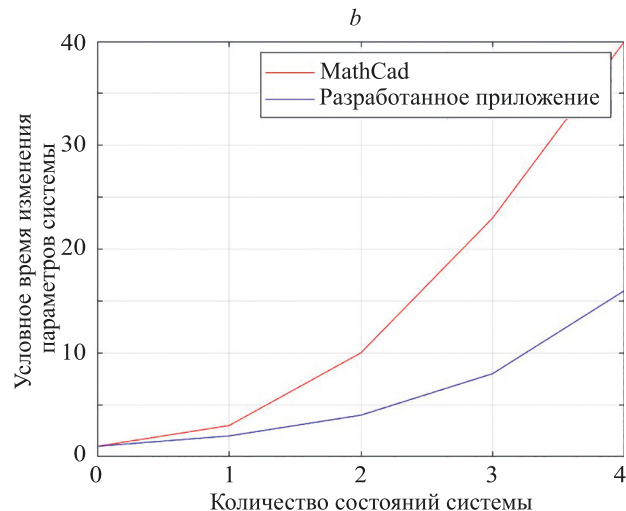
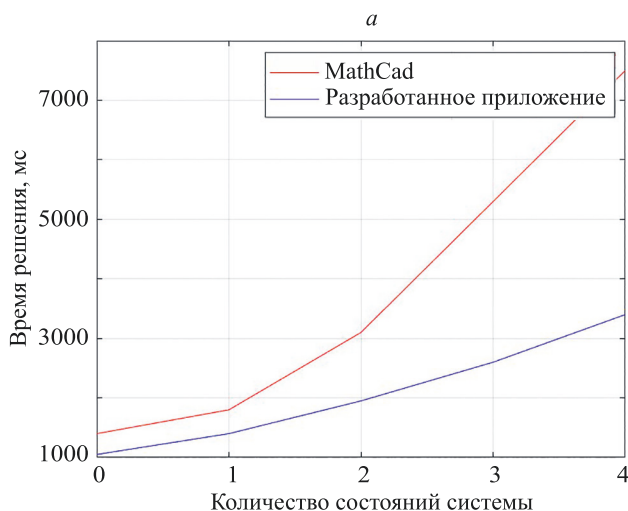


Рис. 12. Сравнительные характеристики разработанного приложения с MathCad: временная сложность (a) и оперативность изменения начальных условий (b)

Fig. 12. Comparative characteristics of the developed application with MathCad: time complexity (a) and speed of changing initial conditions (b)

распределения с числовыми характеристиками k и θ ; Rayl(σ) закон Рэлея с числовой характеристикой σ .

По результатам расчета получены численные значения вероятности нахождения процесса функционирования системы управления роботом для различных состояний (рис. 11, а).

Вычислим значения вероятностей нахождения системы управления роботом во всех состояниях. Результаты представлены на рис. 11, б.

Анализ рис. 11 показал, что результаты, полученные в разработанном приложении, обладают высокой сходимостью с результатами, вычисленными в среде MathCad. При этом, как и в сравнении с системой MATLAB, временная сложность расчета значений вероятностей и условное время изменения параметров системы значительно ниже у приложения (рис. 12).

Заключение

В работе выполнен анализ применения принципа баланса «комплексных вероятностей», основанного на описании стохастического процесса не во временной, а в комплексной области. Результаты позволили разра-

ботать модели нестационарных систем обслуживания с произвольными распределениями вероятностей времени поступления заявок и их обслуживания, учитывающие случайные или детерминированные временные задержки.

Программная реализация принципа баланса позволила получить результаты, которые имеют высокую сходимость с вычислениями, осуществленными в системе MATLAB и среде MathCad. Кроме того, разработанное приложение имеет более высокую производительность за счет своей архитектуры и наличия пользовательского интерфейса, реализующего взаимодействие с оператором. При этом компьютерное моделирование в приложении «Система компьютерного моделирования немарковских процессов» не требует дополнительных знаний языков программирования, навыков работы в специальных системах автоматизированного проектирования, а также необходимости вывода аналитических выражений для получения искомых результатов. Разработанное приложение может быть применено специалистами, занимающимися исследованиями эффективности различных систем.

Литература

1. Сигалов Г.Г., Николаева Г.В., Люперсольский А.М. Влияние параметров надежности на эффективность локальной вычислительной сети с радиальной структурой // Автоматика и вычислительная техника. 1985. № 4. С. 35–42.
2. Горов Г.В., Коган Я.А., Парадизов Н.В. Диффузионно-скачкообразная аппроксимация в однолинейных системах с прерываниями обслуживания и переменным режимом поступления заявок // Автоматика и телемеханика. 1985. № 6. С. 44–51.
3. Бубнов В.П., Сафонов В.И. Разработка динамических моделей нестационарных систем обслуживания. СПб.: Лань, 1999. 64 с.
4. Eremin A.S. A queueing system with determined delay in starting the service // Интеллектуальные технологии на транспорте. 2015. № 4(4). С. 23–26.
5. Смагин В.А., Гусеница Я.Н. К вопросу моделирования одноканальных нестационарных систем с произвольными распределениями моментов времени поступления заявок и длительностей их обслуживания // Труды Военно-космической академии имени А.Ф. Можайского. 2015. № 649. С. 56–63.
6. Смагин В.А., Гусеница Я.Н. О моделировании одноканальных нестационарных систем с произвольными распределениями моментов времени поступления заявок и их обслуживания // Техника средств связи. 2018. № 2(142). С. 199–206.
7. Гусеница Я.Н., Новиков А.Н. Принцип баланса «комплексных вероятностей» при моделировании нестационарных систем обслуживания, представленных циклическим графом состояний // Информатика и космос. 2016. № 3. С. 71–74.
8. Смагин В.А., Гусеница Я.Н. Моделирование одноканальных нестационарных систем обслуживания, представленных циклическим графом состояний // Известия вузов. Приборостроение. 2016. Т. 59. № 10. С. 801–806. <https://doi.org/10.17586/0021-3454-2016-59-10-801-806>
9. Гусеница Я.Н. Принцип баланса «комплексных вероятностей» и его применение для моделирования немарковских процессов: монография. Анапа: Федеральное государственное автономное учреждение «Военный инновационный технополис «ЭРА», 2022. 47 с.
10. Cox D.R. A use of complex probabilities in the theory of stochastic processes // Mathematical Proceedings of the Cambridge Philosophical Society. 1955. V. 51. N 2. P. 313–319. <https://doi.org/10.1017/S0305004100030231>
11. Riordan J. Stochastic Service Systems. New York: Wiley and Sons Inc., 1962. 139 p.
12. Смагин В.А. Вероятностный анализ комплексной переменной // Автоматика и вычислительная техника. 1999. № 5. С. 3–13.

References

1. Sigalov G.G., Nikolaeva G.V., Liupersolskii A.M. Influence of reliability parameters on the efficiency of a local area computer network with a radial structure. *Avtomatika i vychislitel'naja tehnika*, 1985, no. 4, pp. 35–42. (in Russian)
2. Gorov G.V., Kogan A.Ya., Paradizov N.V. Diffusion jumplike approximation in single-channel systems with service interruptions and variable customer arrival rate. *Avtomatika i Telemekhanika*, 1985, no. 6, pp. 44–51. (in Russian)
3. Bubnov V.P., Safonov V.I. *Development of Behavior Models of Non-Stationary Queueing Systems*. St. Petersburg, Lan' Publ., 1999, 64 p. (in Russian)
4. Eremin A.S. A queueing system with determined delay in starting the service. *Intellectual Technologies on Transport*, 2015, no. 4(4), pp. 23–26. (in Russian)
5. Smagin V.A., Gusenitsa Y.N. About modeling of single-channel queueing system with any distribution of time between arriving requirements and any distribution of time between retardation of requirements. *Proceedings of the Mozhaitsky Military Aerospace Academy*, 2015, no. 649, pp. 56–63. (in Russian)
6. Smagin V.A., Gusenitsa Ya.N. On the modeling of single-channel non-stationary systems with arbitrary distributions of incoming requests time and requests servicing. *Means of Communication Equipment*, 2018, no. 2(142), pp. 199–206. (in Russian)
7. Gusenitsa Y., Novikov A. The principle of “complex probabilities” balance, applied to simulation of non-stationary service systems represented by the cyclic graph of states. *Information and Space*, 2016, no. 3, pp. 71–74. (in Russian)
8. Smagin V.A., Gusenitsa Ya.N. Modeling single-channel non-stationary queueing systems presented in the form of a cyclic graph of states. *Journal of Instrument Engineering*, 2016, vol. 59, no. 10, pp. 801–806. (in Russian). <https://doi.org/10.17586/0021-3454-2016-59-10-801-806>
9. Gusenitsa Ya.N. *The Principle of “Complex Probabilities” Balance and Its Application for Modeling Non-Markovian Processes*. Anapa, Voennyj innovatsionnyj tehnopolis «ERA», 2022, 47 p. (in Russian)
10. Cox D.R. A use of complex probabilities in the theory of stochastic processes. *Mathematical Proceedings of the Cambridge Philosophical Society*, 1955, vol. 51, no. 2, pp. 313–319. <https://doi.org/10.1017/S0305004100030231>
11. Riordan J. *Stochastic Service Systems*. New York, Wiley and Sons Inc., 1962, 139 p.
12. Smagin V.A. *Probabilistic analysis of a complex variable*. *Avtomatika i vychislitel'naja tehnika*, 1999, no. 5, pp. 3–13. (in Russian)

13. Smagin V.A. Complex delta function and its information application // *Automatic Control and Computer Sciences*. 2014. V. 48. N 1. P. 10–16. <https://doi.org/10.3103/S0146411614010064>
14. Смагин В.А., Филимоныхин Г.В. О моделировании случайных процессов на основе гипердельта распределения // *Автоматика и вычислительная техника*. 1990. № 1. С. 25–31.
15. Смагин В.А. Немарковские задачи теории надежности. Л.: МО РФ СССР, 1982. 269 с.
16. Ивановский В.С., Гусеница Я.Н., Ширямов О.А. Теоретические основы военной метрологии. Анапа: Федеральное государственное автономное учреждение «Военный инновационный технополис «ЭРА», 2021. 137 с.
17. Ширямов О.А. Стохастическая модель функционирования эталона единицы величины в условиях поступления дополнительных средств измерений на метрологическое обслуживание // *Системы управления, связи и безопасности*. 2018. № 1. С. 95–108.
18. Гусеница Я.Н. Немарковская модель функционирования системы управления подводного робота // *Состояние и перспективы развития современной науки по направлению «Гидроакустические системы обнаружения объектов»*: сборник статей I Всероссийской научно-технической конференции. Анапа: Федеральное государственное автономное учреждение «Военный инновационный технополис «ЭРА», 2021. С. 168–181.
19. Gusenitsa Y., Shiryamov O., Rzhavitin V., Buryj D., Ljvov D. Non-markov general model of the functioning of the robot control system in changing environmental conditions // *Lecture Notes in Electrical Engineering*. 2023. V. 971. P. 38–48. https://doi.org/10.1007/978-3-031-20631-3_5
20. Правоткин И.А. Генерирование изображения на основе текстовых данных с помощью Python на основе библиотеки Pillow // *Приоритетные направления инновационной деятельности в промышленности: сборник научных статей по итогам десятой международной научной конференции*. Ч. 2. М: Общество с ограниченной ответственностью «КОНВЕРТ», 2020. С. 78–79.
21. Таравский Е.А., Сингатуллово И.Ш. Использование Kivy framework для создания мобильного приложения на языке программирования Python // *Концепция развития и эффективного использования научного потенциала общества: сборник статей Международной научно-практической конференции*. В 2-х частях. Ч. 1. Уфа: Общество с ограниченной ответственностью «ОМЕГА САЙНС», 2020. С. 99–101.
22. Ari N., Ustazhanov M. Matplotlib in Python // *Proc. of the 11th International Conference on Electronics, Computer and Computation (ICECCO)*. 2014. P. 1–6. <https://doi.org/10.1109/ICECCO.2014.6997585>
23. Cywiak M., Cywiak D. SymPy // *Multi-Platform Graphics Programming with Kivy*. Apress, Berkeley, CA, 2021. P. 173–190. https://doi.org/10.1007/978-1-4842-7113-1_11
24. Millman K.J., Aivazis M. Python for scientists and engineers // *Computing in Science & Engineering*. 2011. V. 13. N 2. P. 9–12. <https://doi.org/10.1109/MCSE.2011.36>
25. Virtanen P., Gommers R., Oliphant T.E. et al. SciPy 1.0: fundamental algorithms for scientific computing in Python // *Nature Methods*. 2020. V. 17. N 3. P. 261–272. <https://doi.org/10.1038/s41592-019-0686-2>
26. Гусеница Я.Н., Ширямов О.А. Система компьютерного моделирования немарковских процессов. Свидетельство о государственной регистрации программы для ЭВМ № 2022669820. 25.10.2022.
13. Smagin V.A. Complex delta function and its information application. *Automatic Control and Computer Sciences*. 2014, vol. 48, no. 1, pp. 10–16. <https://doi.org/10.3103/S0146411614010064>
14. Smagin V.A., Filimonikhin G.V. On the modeling of random processes based on the hyperdelta distribution. *Avtomatika i vychislitel'naja tehnika*, 1990, no. 1, pp. 25–31. (in Russian)
15. Smagin V.A. *Non-Markovian Problems of Reliability Theory*. Leningrad, Ministry of Defense Soviet Union, 1982, 269 p. (in Russian)
16. Ivanovskii V.S., Gusenitca Ia.N., Shiriamov O.A. *Theoretical Background of the Military Metrolog.* Anapa, Voennyj innovacionnyj tehnopolis «ERA», 2021, 137 p. (in Russian)
17. Shiryamov O.A. Stochastic model of functioning of a measurement standard in conditions when a additional measuring instruments receipt into metrological maintenance. *Systems of Control, Communication and Security*, 2018, no. 1, pp. 95–108. (in Russian)
18. Gusenitca Ia.N. Non-Markovian functional model of an underwater robot control system. *State and prospects for the modern science development in the direction of "Hydroacoustic systems for detecting objects"*. *Collected papers of the 1st All-Russian Scientific and Technical Conference*. Anapa, Voennyj innovacionnyj tehnopolis «ERA», 2021, pp. 168–181. (in Russian)
19. Gusenitsa Y., Shiryamov O., Rzhavitin V., Buryj D., Ljvov D. Non-markov general model of the functioning of the robot control system in changing environmental conditions. *Lecture Notes in Electrical Engineering*, 2023, vol. 971, pp. 38–48. https://doi.org/10.1007/978-3-031-20631-3_5
20. Pravotkin I.A. Generating an image from text data with Python language based on Pillow library. *Priority areas of innovation in industry. Collection of scientific articles following the 10th International Scientific Conference*. Ch. 2. Moscow, KONVERT Publ., 2020, pp. 78–79. (in Russian)
21. Taravskii E.A., Singatullovo I.Sh. Using the Kivy framework to create a mobile application in the Python programming language. *The concept of development and efficient use of the society scientific potential. Collected papers of the International Scientific and Practical Conference (in two parts)*. Ch. 1. Ufa, Omega science Publ., 2020, pp. 99–101. (in Russian)
22. Ari N., Ustazhanov M. Matplotlib in Python. *Proc. of the 11th International Conference on Electronics, Computer and Computation (ICECCO)*, 2014, pp. 1–6. <https://doi.org/10.1109/ICECCO.2014.6997585>
23. Cywiak M., Cywiak D. SymPy. *Multi-Platform Graphics Programming with Kivy*. Apress, Berkeley, CA, 2021, pp. 173–190. https://doi.org/10.1007/978-1-4842-7113-1_11
24. Millman K.J., Aivazis M. Python for scientists and engineers. *Computing in Science & Engineering*, 2011, vol. 13, no. 2, pp. 9–12. <https://doi.org/10.1109/MCSE.2011.36>
25. Virtanen P., Gommers R., Oliphant T.E. et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nature Methods*, 2020, vol. 17, no. 3, pp. 261–272. <https://doi.org/10.1038/s41592-019-0686-2>
26. Gusenitca Ia.N., Shiriamov O.A. Computer simulation system for non-Markovian processes. *Certificate of registration of computer software 2022669820*. 25.10.2022. (in Russian)

Авторы

Гусеница Ярослав Николаевич — кандидат технических наук, начальник научно-исследовательского отдела, Военный инновационный технополис «ЭРА», Анапа, 353456, Российская Федерация, orcid.org/0000-0003-4757-5066, yaromir226@gmail.com

Ширямов Олег Анатольевич — кандидат технических наук, старший научный сотрудник, Военно-космическая академия А.Ф. Можайского, Санкт-Петербург, 197198, Российская Федерация, <https://orcid.org/0000-0002-8985-8251>, shiryamov_ua@mail.ru

Authors

Yaroslav N. Gusenitsa — PhD, Head of Department, Military Innovative Technopolis “ERA”, Anapa, 353456, Russian Federation, orcid.org/0000-0003-4757-5066, yaromir226@gmail.com

Oleg A. Shiryamov — PhD, Senior Researcher, Mozhaisky Military Aerospace Academy, Saint Petersburg, 197198, Russian Federation, <https://orcid.org/0000-0002-8985-8251>, shiryamov_ua@mail.ru

Статья поступила в редакцию 17.11.2022
Одобрена после рецензирования 29.11.2022
Принята к печати 20.01.2023

Received 17.11.2022
Approved after reviewing 29.11.2022
Accepted 20.01.2023



Работа доступна по лицензии
Creative Commons
«Attribution-NonCommercial»

doi: 10.17586/2226-1494-2023-23-1-161-168

УДК 519.816

Метод объективизации значений весовых коэффициентов для принятия решений в многокритериальных задачах

Денис Сергеевич Соловьев✉

Тамбовский государственный университет имени Г.Р. Державина, Тамбов, 392036, Российская Федерация
solovjevdenis@mail.ru✉, <https://orcid.org/0000-0001-6613-3218>

Аннотация

Предмет исследования. При определении весовых коэффициентов в многокритериальных задачах для одинаковых исходных данных с целью оценки важности критериев на основе существующих качественных и количественных подходов возникает противоречие понятию «единственность решения». Это приводит к снижению степени доверия к принятым решениям. Таким образом, требуется определить степень объективности используемых весовых коэффициентов. Целью исследования является объективизация значений весовых коэффициентов для задач принятия решений. **Метод.** В работе предложено сочетание качественного и количественного подходов для определения весовых коэффициентов с заданной согласованностью. Сформирована матрица весовых коэффициентов (количественный подход), которой сопоставлена матрица рангов (качественный подход). Для получения заданного значения коэффициента согласованности с помощью матрицы рангов решена задача оптимизации. **Основные результаты.** Применение предложенного метода продемонстрировано на примере решения многокритериальной задачи выбора наилучшей альтернативы. Расчет значений весовых коэффициентов осуществлен с помощью разработанного программного обеспечения на языке Python. Решение сведено к однокритериальной задаче на основе максиминного подхода с использованием найденных значений весовых коэффициентов. В результате решение задачи с заданной согласованностью обеспечило объективность результата и повысило доверие к принятому решению. **Практическая значимость.** Предложенный метод может быть использован при оценке важности критериев без необходимости участия лица, принимающего решение.

Ключевые слова

метод, объективность, согласованность, весовые коэффициенты, принятие решений, оптимизация, многокритериальная задача

Ссылка для цитирования: Соловьев Д.С. Метод объективизации значений весовых коэффициентов для принятия решений в многокритериальных задачах // Научно-технический вестник информационных технологий, механики и оптики. 2023. Т. 23, № 1. С. 161–168. doi: 10.17586/2226-1494-2023-23-1-161-168

The objectification method of the weight coefficients for decision-making in multicriteria problems

Denis S. Solovjev✉

Derzhavin Tambov State University, Tambov, 392036, Russian Federation
solovjevdenis@mail.ru✉, <https://orcid.org/0000-0001-6613-3218>

Abstract

The contradiction in the concept of “solution uniqueness” arises when determining weight coefficients in multicriteria problems for the same initial data in order to assess the criteria importance based on existing qualitative and quantitative approaches. This leads to a decrease in the degree of confidence in the decisions made. Thus, it is required to determine the objectivity degree of the weighting coefficients used. The objectification of the weight coefficients for decision-making problems is the purpose of the study. The article proposes a combination of qualitative and quantitative approaches to determine weight coefficients with a given consistency. The weight coefficients matrix is formed

(quantitative approach). This matrix is correlated with the rank matrix (qualitative approach). The optimization problem is solved to obtain a given consistency coefficient using the rank matrix. The proposed method application is demonstrated by the example of solving the problem of choosing the best alternative in multicriteria problem. The calculation of the weight coefficients is carried out using the developed software in the Python. The solution is reduced to a single-objective problem based on the maximin approach using the found weight coefficients. Thus, solving the problem with a given consistency ensures the result objectivity and increases the decision confidence. The proposed method can be used in assessing the criteria importance without the need for the participation of the decision maker.

Keywords

method, objectivity, consistency, weight coefficients, decision-making, optimization, multicriteria problem

For citation: Solovjev D.S. The objectification method of the weight coefficients for decision-making in multicriteria problems. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 2023, vol. 23, no. 1, pp. 161–168 (in Russian). doi: 10.17586/2226-1494-2023-23-1-161-168

Введение

Определение весовых значений коэффициентов для оценки важности критериев — одна из ключевых проблем анализа данных в многокритериальных задачах оптимизации при разработке систем поддержки принятия решений [1]. Методология определения важности критериев много лет являются предметом интенсивных исследований и научных дискуссий. Большинство подходов к определению весовых коэффициентов для критериев можно разделить на качественные и количественные. Качественные подходы основаны на определении весовых коэффициентов с использованием информации от лиц, принимающих решения, или экспертов, включенных в процесс принятия решений. В качественных подходах лица, принимающие решения, влияют на процесс принятия решений. Качественные подходы реализованы в таких методах как: экспертные оценки [2], анализ иерархий [3] и др. Данные методы различаются количеством участников процесса принятия решения и способом формирования окончательных весов критериев. В отличие от качественных подходов, количественные подходы определяют значения весовых коэффициентов для критериев на основе анализа данных, присутствующих в исходной матрице решений. Примерами количественных подходов являются методы: равнозначный [4]; энтропийный [5]; стандартного отклонения [6], основанный на эффектах удаления [7]; корреляции [8]; потери влияния [9]; комплексный [10]; угловой [11], основанный на коэффициенте Джини [12]; статистической дисперсии [13] и др. Количественные подходы устраняют лицо, принимающее решение, от оценки важности критериев и различаются алгоритмами анализа данных. Использование данных подходов приводит к различающимся значениям весовых коэффициентов, что противоречит понятию «единственность решения» для одинаковых исходных данных. В свою очередь остается открытым вопрос определения степени объективности сформированных значений весовых коэффициентов для критериев, которые напрямую зависят от выбора конкретных участников процесса принятия решения (в качественных подходах) и алгоритмов анализа данных в исходной матрице решений (в количественных подходах), а также их количества, в результате чего снижается степень доверия к принятым решениям. Поскольку согласованность — один из основных показателей объективности, то разработка метода для расчета весовых коэффициентов с

заданной согласованностью — выход из сложившейся ситуации.

Цель работы — объективизация значений весовых коэффициентов для принятия решений в многокритериальных задачах.

Материалы и методы

Предложен метод расчета значений весовых коэффициентов для принятия решений в многокритериальных задачах. Метод отличается использованием количественных подходов оценки при условии, что ранговые значения весовых коэффициентов, полученные с помощью качественного подхода, обеспечивают заданную согласованность и статистическую значимость.

Пусть имеется множество альтернатив $\mathbf{A} = \{A_1, A_2, \dots, A_M\}$, значения которых определены множеством критериев $\mathbf{K} = \{K_1, K_2, \dots, K_N\}$ согласно матрице решений:

$$\mathbf{k} = \begin{pmatrix} k_{1,1} & k_{1,2} & \dots & k_{1,N} \\ k_{2,1} & k_{2,2} & \dots & k_{2,N} \\ \dots & \dots & \dots & \dots \\ k_{M,1} & k_{M,2} & \dots & k_{M,N} \end{pmatrix}_{M \times N}, \quad (1)$$

где $k_{m,n}$ — значение m -й альтернативы по n -му критерию.

В результате использования множества методов $\mathbf{F} = \{F_1, F_2, \dots, F_H\}$, реализующих количественный подход, сформируем матрицу значений весовых коэффициентов:

$$\mathbf{\omega} = \begin{pmatrix} \omega_{1,1} & \omega_{1,2} & \dots & \omega_{1,N} \\ \omega_{2,1} & \omega_{2,2} & \dots & \omega_{2,N} \\ \dots & \dots & \dots & \dots \\ \omega_{H,1} & \omega_{H,2} & \dots & \omega_{H,N} \end{pmatrix}_{H \times N}, \quad (2)$$

где $\omega_{h,n}$ — значение n -го критерия по h -му методу.

В каждой h -й строке матрицы (2) найдем перестановку индексов $p_h(1), p_h(2), \dots, p_h(N)$ для $(1, 2, \dots, N)$, после которой элементы данной строки будут расположены в порядке неубывания в виде:

$$\forall h = 1, 2, \dots, H: \omega_{h,p_h(1)} \leq \omega_{h,p_h(2)} \leq \dots \leq \omega_{h,p_h(N)}, \quad (3)$$

где H — количество методов для принятия решений.

В полученных строках (3) для каждого элемента найдем число его дублирующих значений:

$$t_{h,p_h(n)} = \sum_{l=1}^N \begin{cases} 1, & \text{если } \omega_{h,p_h(n)} = \omega_{h,p_h(l)} \wedge l \neq n \\ 0, & \text{иначе} \end{cases}, \quad (4)$$

а также суммарное число дублей:

$$T_h = \sum_{n=1}^N t_{h,p_h(n)}, \quad (5)$$

где l — индекс критерия ($l = 1, 2, \dots, N$).

Если суммарное число дублей $T_h = 0$, то осуществим переход от количественной оценки критериев к качественной с помощью их расстановки в порядке возрастания значимости (ранжирования). Весовому коэффициенту в (3), имеющему минимальное значение, присвоим первое место (ранг) в расстановке, а последующим в порядке увеличения значения — места, возрастающие между собой на единицу. Таким образом, чем большее значение имеет весовой коэффициент, тем выше его место (ранг), занятое в расстановке. Элементам в h -й строке сопоставим ранги согласно:

$$r_{h,p_h(n)} = p_h(n), \quad (n = 1, 2, \dots, N). \quad (6)$$

Если $T_h \neq 0$, то получим такую перестановку индексов $p_h^*(1), p_h^*(2), \dots, p_h^*(G_h)$ для $(1, 2, \dots, G_h)$, после которой элементы h -й строки будут располагаться в порядке возрастания:

$$\omega_{h,p_h^*(1)} < \omega_{h,p_h^*(2)} < \dots < \omega_{h,p_h^*(G_h)}, \quad (7)$$

при этом $G_h < N$. Символ «*» обозначает однозначную идентификацию перестановок, так как $p_h(1), p_h(2), \dots, p_h(N)$ обеспечивает неубывание (3), а $p_h^*(1), p_h^*(2), \dots, p_h^*(G_h)$ — возрастание (7) в h -й строке матрицы весовых коэффициентов (2).

Определим количество уникальных значений $\omega_{h,p_h(n)}$ в h -й строке (3):

$$G_h = 1 + \sum_{n=2}^N \begin{cases} 1, & \text{если } \omega_{h,p_h(n)} > \omega_{h,p_h(n-1)} \\ 0, & \text{иначе} \end{cases}. \quad (8)$$

Каждому уникальному элементу h -й строки (7) сопоставим количество его повторов:

$$z_{h,p_h^*(g)} = \sum_{n=1}^N \begin{cases} 1, & \text{если } \omega_{h,p_h^*(g)} = \omega_{h,p_h(n)} \\ 0, & \text{иначе} \end{cases}. \quad (9)$$

В связи с тем, что в выражении (3) имеются повторяющиеся значения весовых коэффициентов (присутствуют значения $z_{h,p_h^*(g)}$ не равные единице), то для элемента $\omega_{h,p_h^*(g)}$ в (7) вычислим среднее значение места (ранга) $r_{h,p_h^*(g)}^*$ как отношение суммы порядковых номеров в (3) к количеству дублей (9) согласно:

$$r_{h,p_h^*(g)}^* = \left(\sum_{n=1}^N \begin{cases} n, & \text{если } \omega_{h,p_h^*(g)} = \omega_{h,p_h(n)} \\ 0, & \text{иначе} \end{cases} \right) / z_{h,p_h^*(g)}. \quad (10)$$

При выполнении условия:

$$\omega_{h,p_h^*(g)} = \omega_{h,p_h(n)}, \quad (11)$$

элементам h -й строки (3) сопоставим полученные ранги (10) следующим образом:

$$r_{h,p_h(n)} = r_{h,p_h^*(g)}^* \quad (n = 1, 2, \dots, N), \quad (g = 1, 2, \dots, G_h) \quad (12)$$

Вместо матрицы значений весовых коэффициентов (2) сформируем матрицу рангов:

$$\mathbf{r} = \begin{pmatrix} r_{1,1} & r_{1,2} & \dots & r_{1,N} \\ r_{2,1} & r_{2,2} & \dots & r_{2,N} \\ \dots & \dots & \dots & \dots \\ r_{H,1} & r_{H,2} & \dots & r_{H,N} \end{pmatrix}_{H \times N}. \quad (13)$$

Для матрицы рангов (13) определим коэффициент согласованности Кендалла W , величину которого найдем в диапазоне от 0 до 1, причем от значения 0,7 и выше согласованность будем считать высокой. Значимость $\chi^2_{\text{расч}}$ для коэффициента согласованности W рассчитаем согласно критерию Пирсона. Табличное значение критерия Пирсона $\chi^2_{\text{расч}}(\alpha_{\text{зад}}, \nu)$ определим уровнем значимости $\alpha_{\text{зад}}$ и числом степеней свободы $\nu = N - 1$.

Если $\chi^2_{\text{расч}} > \chi^2_{\text{табл}}(\alpha_{\text{зад}}, \nu)$, то с вероятностью $(1 - \alpha_{\text{зад}})$ коэффициент согласованности признается значимым, и возможно сформировать окончательные значения весовых коэффициентов важности критериев:

$$\tilde{\omega} = (\tilde{\omega}_1, \tilde{\omega}_2, \dots, \tilde{\omega}_N), \quad (14)$$

где n -ое значение коэффициента $\tilde{\omega}_n$ определяется как отношение суммы рангов n -го критерия к общей сумме рангов всех N критериев у всех H методов количественного подхода:

$$\tilde{\omega}_n = \left(\sum_{h=1}^H r_{h,n} \right) / \left(\sum_{h=1}^H \sum_{n=1}^N r_{h,n} \right), \quad (15)$$

что обеспечивает выполнение условия нормировки (сумма весовых коэффициентов должна равняться единице).

Так как применение всех методов из множества \mathbf{F} не гарантирует обеспечение заданных значений согласованности $W_{\text{зад}}$ и уровня значимости $\alpha_{\text{зад}}$, то необходимо сформулировать задачу оптимизации. Для этого множество методов \mathbf{F} расширим до мультимножества следующим образом:

$$\tilde{\mathbf{F}} = \{b_1 F_1, b_2 F_2, \dots, b_H F_H\}, \quad (16)$$

где b_h — кратность вхождения метода F_h в мультимножество $\tilde{\mathbf{F}}$.

Кратность вхождения b_h принимает следующие значения:

$$b_h = \begin{cases} 1, & \text{если } F_h \in \tilde{\mathbf{F}} \\ 0, & \text{иначе} \end{cases}, \quad (17)$$

которые определяют наличие «1» или отсутствие «0» строки h в матрице (2).

Таким образом, задача оптимизации будет звучать следующим образом. Необходимо найти такие кратности вхождения методов $\mathbf{b}^* = (b_1^*, b_2^*, \dots, b_H^*)$ из

мультимножества $\tilde{\mathbf{F}}$ и соответствующие им весовые коэффициенты $\tilde{\mathbf{w}}^* = (\tilde{w}_1^*, \tilde{w}_2^*, \dots, \tilde{w}_N^*)$, для которых значение отклонения расчетного W^* от заданного $W_{\text{зад}}$ минимальное:

$$\Delta W^* = \min(W^* - W_{\text{зад}}), \quad (18)$$

при уравнениях связи (1)–(17) и ограничениях:

$$0 < W_{\text{зад}} < 1, \quad (19)$$

$$W^* \geq W_{\text{зад}}, \quad (20)$$

$$\chi^2_{\text{расч}} > \chi^2_{\text{табл}}(\alpha_{\text{зад}}, \nu). \quad (21)$$

Задача (18) представляет собой случай задачи бинарного нелинейного программирования, когда необходимо выбрать или отказаться от выбора отдельных методов для получения весовых коэффициентов с заданной степенью согласованности. Гарантированное оптимальное решение такой задачи получим при использовании метода сканирования.

Экспериментальная часть

Рассмотрим применение предложенного метода на примере принятия решения в многокритериальной задаче выбора наилучшей альтернативы.

Пусть имеется множество критериев $\mathbf{K} = \{K_1, K_2, K_3, K_4\}$, где критерии: K_1 — равномерность распределения толщины гальванического покрытия на поверхности детали (безразмерное значение от 0 до 1); K_2 — производительность гальванической ванны (значение больше 0); K_3 — электроэнергия, затрачиваемая на нанесение гальванического покрытия (значение больше 0); K_4 — стоимость реализации гальванического процесса (значение больше 0). Критерии K_1 и K_2 стремятся к максимуму, а критерии K_3 и K_4 — к минимуму.

Дано множество альтернатив $\mathbf{A} = \{A_1, A_2, A_3, A_4, A_5, A_6, A_7\}$, где гальванические ванны: A_1 — с анодами, работающими различное время [14]; A_2 — с анодами, работающими в циклическом режиме [15]; A_3 — с отключаемыми при реверсировании тока анодами [16]; A_4 — с защитными катодами [17]; A_5 — с токо-непроводящими экранами [18]; A_6 — с биполярными электродами [19]; A_7 — с фигурными анодами [20].

Полученные значения на основании матрицы решений (1) представлены в табл. 1.

Пусть дано множество методов $\mathbf{F} = \{F_1, F_2, F_3, F_4, F_5, F_6, F_7, F_8, F_9, F_{10}\}$, где методы: F_1 — равнозначный; F_2 — энтропийный; F_3 — стандартного отклонения; F_4 — основанный на эффектах удаления; F_5 — корреляции; F_6 — потери влияния; F_7 — комплексный; F_8 — угловой; F_9 — основанный на коэффициенте Джини; F_{10} — статистической дисперсии.

Многокритериальную задачу выбора наилучшей альтернативы m^* сведем к однокритериальной с использованием максиминного подхода:

$$m^* = \operatorname{argmax}_m \min_n \tilde{w}_n \bar{k}_{m,n}, \quad (22)$$

где $\bar{k}_{m,n}$ — нормализованное значение m -й альтернативы по n -му критерию.

Нормализацию значений $k_{m,n}$ осуществим по методу Вейтендорфа [21]. Весовые коэффициенты $\tilde{\mathbf{w}}^*$ для (22) определим с помощью решения задачи (18) при заданном значении согласованности $W_{\text{зад}} = 0,85$ согласно (19) и уровне значимости $\alpha_{\text{зад}} = 0,05$ с помощью разработанного программного обеспечения на языке Python, использующего модули `scipy.stats` и `pymcdm.weights` [22].

Результаты и их обсуждение

Нормализованная матрица решений на основании полученных результатов из табл. 1, представлена в табл. 2.

Сформируем значения матрицы весовых коэффициентов (2), используя выбранное множество методов \mathbf{F} (табл. 3).

В матрице весовых коэффициентов рассматриваемые критерии \mathbf{K} имеют наибольшие значения следующее количество раз: K_1 — 2, K_2 — 1, K_3 — 4, K_4 — 2 и $K_1 \sim K_2 \sim K_3 \sim K_4$ — 1.

Применение предлагаемого метода ранжирования весовых коэффициентов формирует значения матрицы рангов \mathbf{r} (13), которая представлена в табл. 4.

Из табл. 4 следует, что методы F_2 , F_8 и F_9 , а также F_3 и F_4 обеспечивают одинаковую ранжировку весовых коэффициентов.

Таблица 1. Матрица решений

Table 1. Decision matrix

Индекс альтернативы m	Критерии			
	$k_{m,1}$	$k_{m,2}$, 1/ч	$k_{m,3}$, Вт·ч	$k_{m,4}$, руб.
1	0,86845	0,833	126,3	10994,19
2	0,86939	0,727	358,7	9212,17
3	0,85274	0,856	382,9	11456,55
4	0,82814	0,800	284,4	9372,79
5	0,80668	0,500	220,2	10512,34
6	0,86132	0,593	152,2	5567,23
7	0,77248	0,941	290,1	16903,02

Таблица 2. Нормализованная матрица решений

Table 2. Normalized decision matrix

Индекс альтернативы m	Нормализованные критерии			
	$\bar{k}_{m,1}$	$\bar{k}_{m,2}$	$\bar{k}_{m,3}$	$\bar{k}_{m,4}$
1	0,990	0,755	1,000	0,521
2	1,000	0,515	0,094	0,678
3	0,828	0,807	0,000	0,480
4	0,574	0,680	0,384	0,664
5	0,353	0,000	0,634	0,564
6	0,917	0,211	0,899	1,000
7	0,000	1,000	0,362	0,000

Таблица 3. Матрица весовых коэффициентов

Table 3. Weight coefficients matrix

Индекс метода h	Весовые коэффициенты			
	$\omega_{h,1}$	$\omega_{h,2}$	$\omega_{h,3}$	$\omega_{h,4}$
1	0,250	0,250	0,250	0,250
2	0,006	0,150	0,502	0,342
3	0,000	0,000	0,028	0,972
4	0,053	0,277	0,313	0,357
5	0,376	0,187	0,232	0,205
6	0,720	0,191	0,036	0,053
7	0,067	0,413	0,257	0,263
8	0,047	0,221	0,393	0,338
9	0,045	0,222	0,411	0,322
10	0,284	0,247	0,290	0,179

Таблица 4. Матрица рангов

Table 4. Rank matrix

Индекс метода h	Ранги			
	$r_{h,1}$	$r_{h,2}$	$r_{h,3}$	$r_{h,4}$
1	2,5	2,5	2,5	2,5
2	1	2	4	3
3	1	2	3	4
4	1	2	3	4
5	4	1	3	2
6	4	3	1	2
7	1	4	2	3
8	1	2	4	3
9	1	2	4	3
10	3	2	4	1

Решение задачи (20) получим среди всевозможных сочетаний без повторений C_H^d для ($d = 2, 3, \dots, H$) с суммарным количеством элементов 1013.

Значения согласованности и критерия Пирсона для возможных сочетаний показаны на рис. 1.

Количество выполнений ограничений (20) и (21) представлено на рис. 2.

На рис. 1 видно, что с увеличением количества методов d : снижается среднее значение коэффициента согласованности Кендалла, что объясняется добавлением

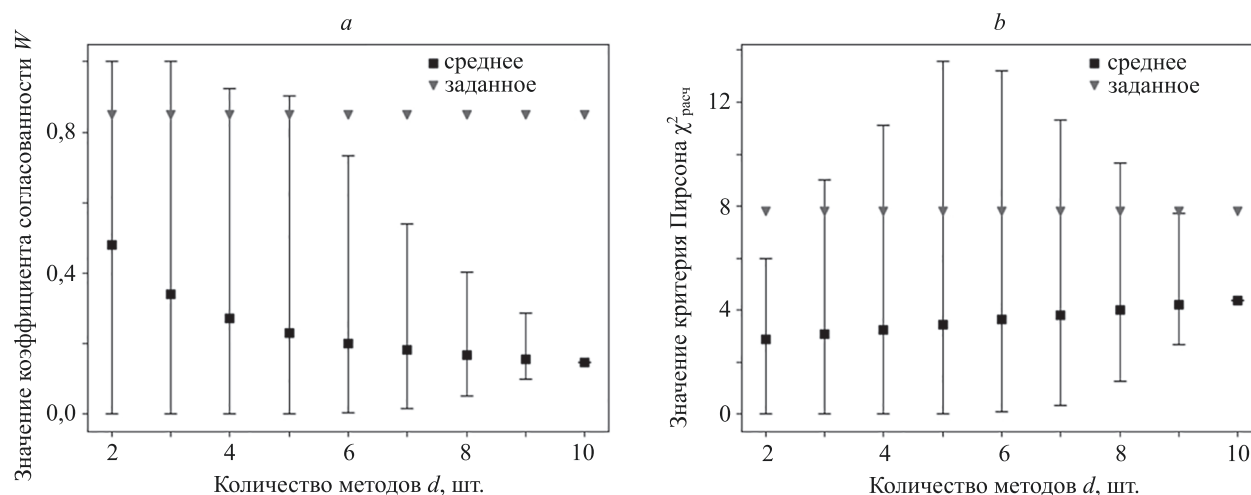


Рис. 1. Значение коэффициента согласованности W (a) и критерия Пирсона $\chi^2_{\text{расч}}$ (b) для d -го сочетания

Fig. 1. The consistency coefficient W (a) and Pearson criterion χ^2_{calc} (b) for the d -th combination

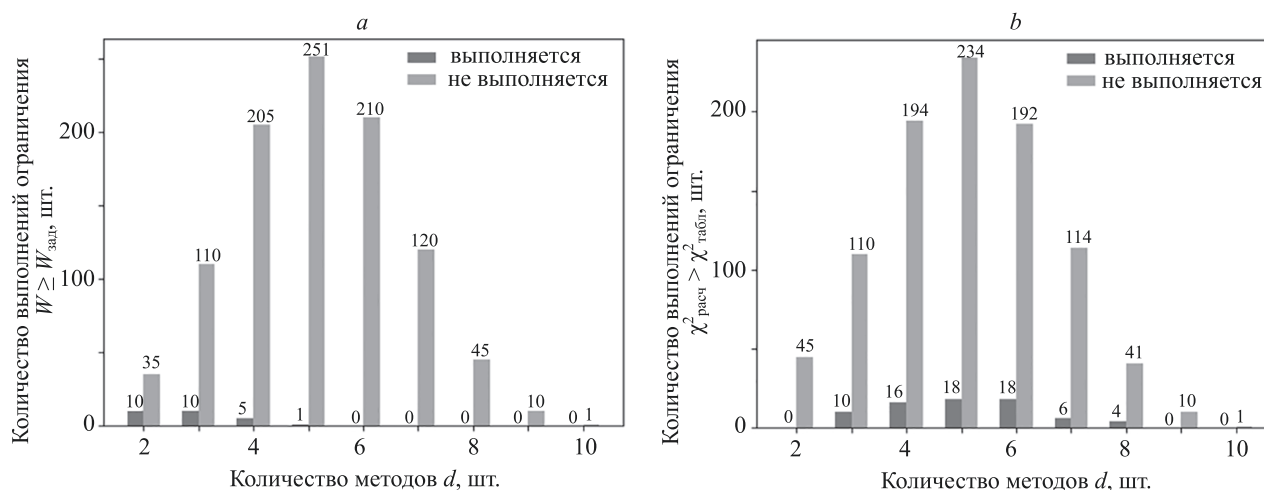


Рис. 2. Количество выполнений ограничения $W \geq W_{\text{зад}}$ (a) и $\chi^2_{\text{расч}} > \chi^2_{\text{табл}}(\alpha_{\text{зад}}, \nu)$ (b)

Fig. 2. Number of constraint executions $W \geq W_{\text{given}}$ (a) and $\chi^2_{\text{calc}} > \chi^2_{\text{table}}(\alpha_{\text{given}}, \nu)$ (b)

в матрицу (2) весовых коэффициентов и, как следствие, в матрицу (13) рангов, полученных с использованием методов, значения которых резко отличаются от остальных; увеличивается среднее значение значимости коэффициента согласованности по критерию Пирсона, что объясняется увеличением количества слагаемых числителя в (15).

Для различных сочетаний, начиная с $d = 6$, минимальные и максимальные значения выборок коэффициента согласованности и его значимости уменьшаются.

На рис. 2 показано, что: ограничение (20) выполняется только 26 раз при ($d = 2, 3, 4, 5$), что составляет 2,57 % от общего количества сочетаний методов; ограничение (21) выполняется только 72 раза при ($d = 3, 4, 5, 6, 7, 8$), что составляет 7,11 % от общего количества сочетаний методов. Получается, что количество используемых методов d должно быть больше 3 и не превышать 5, что сопоставимо с количеством критериев N .

Таким образом, область допустимых решений задачи оптимизации (18) находится среди 16 вариантов при

($d = 3, 4, 5$), что составляет 1,58 % от общего количества сочетаний методов. В результате решения задачи оптимизации (18) получено минимальное отклонение расчетного коэффициента согласованности от заданного $\Delta W^* = 0,05$ при найденном значении $W^* = 0,9$. Значение W^* статистически значимо, поскольку обеспечивает величину критерия $\chi^2_{\text{расч}} = 10,8$, которое превышает $\chi^2_{\text{табл}}(\alpha_{\text{зад}}, \nu) = 7,81$. Полученное значение ΔW^* обеспечивается следующими значениями кратностей вхождения методов $\mathbf{b}^* = (0, 0, 1, 1, 0, 0, 1, 1, 0)$, формирующих весовые коэффициенты $\tilde{\mathbf{w}}^* = (0,1; 0,2; 0,35; 0,35)$. Значения нормализованной матрицы решений, с учетом найденных весовых коэффициентов $\tilde{\mathbf{w}}^*$, приведены в табл. 5.

Согласно данным из табл. 5 задача выбора наилучшей альтернативы (22) имеет решение $m^* = \arg\max \{0,099; 0,033; 0,000; 0,057; 0,000; 0,042; 0,000\} = m^* = 1$, соответствующее гальванической ванне с анодами, работающими различное время.

Таблица 5. Нормализованная матрица решений с учетом весовых коэффициентов

Table 5. Normalized decision matrix with weight coefficients

Индекс альтернативы m	Нормализованные критерии с учетом весовых коэффициентов			
	$\tilde{\omega}_1 \bar{k}_{m,1}$	$\tilde{\omega}_2 \bar{k}_{m,2}$	$\tilde{\omega}_3 \bar{k}_{m,3}$	$\tilde{\omega}_4 \bar{k}_{m,4}$
1	0,099	0,151	0,350	0,182
2	0,100	0,103	0,033	0,237
3	0,083	0,161	0,000	0,168
4	0,057	0,136	0,134	0,233
5	0,035	0,000	0,222	0,197
6	0,092	0,042	0,315	0,350
7	0,000	0,200	0,127	0,000

Заключение

Весовые коэффициенты являются калибровочным инструментом в многокритериальных задачах принятия решений, и качество их определения напрямую влияет на получаемый результат. Каждый из качественных и количественных подходов определения значений весовых коэффициентов имеет свои преимущества и недостатки. В предлагаемом подходе качественные методы определения значений коэффициентов использованы

для корректировки весов критериев, полученных с помощью количественных методов. Решение задачи оптимизации выбора среди количественных подходов обеспечило получение значений весовых коэффициентов с заданной согласованностью и статистической значимостью. Последнее обстоятельство повысило объективность сформированных значений весовых коэффициентов для критериев и, как следствие, степень доверия к принятым решениям.

Литература

- Ионов М.В., Болгова Е.В., Звартау Н.Э., Авдонина Н.Г., Балахонцева М.А., Ковальчук С.В., Конради А.О. Внедрение системы поддержки принятия решений для повышения качества медицинских данных пациентов с артериальной гипертензией // Научно-технический вестник информационных технологий, механики и оптики. 2022. Т. 22. № 1. С. 217–222. <https://doi.org/10.17586/2226-1494-2022-22-1-217-222>
- Зубкова Т.М., Тагирова Л.Ф., Тагиров В.К. Прототипирование адаптивных пользовательских интерфейсов прикладных программ с использованием методов искусственного интеллекта // Научно-технический вестник информационных технологий, механики и оптики. 2019. Т. 19. № 4. С. 680–688. <https://doi.org/10.17586/2226-1494-2019-19-4-680-688>
- Repetski E.J., Sarkani S., Mazzuchi T. Applying the analytic hierarchy process (AHP) to expert documents // International Journal of the Analytic Hierarchy Process. 2022. V. 14. N 1. <https://doi.org/10.13033/ijahp.v14i1.919>
- Машунин К.Ю., Машунин Ю.К. Векторная оптимизация с равнозначными и приоритетными критериями // Известия Российской академии наук. Теория и системы управления. 2017. № 6. С. 80–99. <https://doi.org/10.7868/S0002338817060075>
- Lotfi F.H., Fallahnejad R. Imprecise Shannon's entropy and multi attribute decision making // Entropy. 2010. V. 12. N 1. P. 53–62. <https://doi.org/10.3390/e12010053>
- Wang Y.M., Luo Y. Integration of correlations with standard deviations for determining attribute weights in multiple attribute decision making // Mathematical and Computer Modelling. 2010. V. 51. N 1–2. P. 1–12. <https://doi.org/10.1016/j.mcm.2009.07.016>
- Keshavarz-Ghorabae M., Amiri M., Zavadskas E.K., Turskis Z., Antucheviciene J. Determination of objective weights using a new method based on the removal effects of criteria (MEREC) // Symmetry. 2021. V. 13. N 4. P. 525. <https://doi.org/10.3390/sym13040525>
- Krishnan A.R., Kasim M.M., Hamid R., Ghazali M.F. A modified CRITIC method to estimate the objective weights of decision criteria // Symmetry. 2021. V. 13. N 6. P. 973. <https://doi.org/10.3390/sym13060973>
- Zavadskas E.K., Podvezko V. Integrated determination of objective criteria weights in MCDM // International Journal of Information Technology & Decision Making. 2016. V. 15. N 2. P. 267–283. <https://doi.org/10.1142/S0219622016500036>

References

- Ionov M.V., Bolgova E.V., Zvartau N.E., Avdonina N.G., Balakhontseva M.A., Kovalchuk S.V., Konradi A.O. Implementation of a clinical decision support system to improve the medical data quality for hypertensive patients. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 2022, vol. 22, no. 1, pp. 217–222. (in Russian). <https://doi.org/10.17586/2226-1494-2022-22-1-217-222>
- Zubkova T.M., Tagirova L.F., Tagirov V.K. Prototyping of adaptive user application programming interfaces by artificial intelligence methods. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 2019, vol. 19, no. 4, pp. 680–688. (in Russian). <https://doi.org/10.17586/2226-1494-2019-19-4-680-688>
- Repetski E.J., Sarkani S., Mazzuchi T. Applying the analytic hierarchy process (AHP) to expert documents. *International Journal of the Analytic Hierarchy Process*, 2022, vol. 14, no. 1. <https://doi.org/10.13033/ijahp.v14i1.919>
- Mashunin K.Y., Mashunin Y.K. Vector optimization with equivalent and priority criteria. *Journal of Computer and Systems Sciences International*, 2017, vol. 56, no. 6, pp. 975–996. <https://doi.org/10.1134/S1064230717060119>
- Lotfi F.H., Fallahnejad R. Imprecise Shannon's entropy and multi attribute decision making. *Entropy*, 2010, vol. 12, no. 1, pp. 53–62. <https://doi.org/10.3390/e12010053>
- Wang Y.M., Luo Y. Integration of correlations with standard deviations for determining attribute weights in multiple attribute decision making. *Mathematical and Computer Modelling*, 2010, vol. 51, no. 1–2, pp. 1–12. <https://doi.org/10.1016/j.mcm.2009.07.016>
- Keshavarz-Ghorabae M., Amiri M., Zavadskas E. K., Turskis Z., Antucheviciene J. Determination of objective weights using a new method based on the removal effects of criteria (MEREC). *Symmetry*, 2021, vol. 13, no. 4, pp. 525. <https://doi.org/10.3390/sym13040525>
- Krishnan A.R., Kasim M.M., Hamid R., Ghazali M.F. A modified CRITIC method to estimate the objective weights of decision criteria. *Symmetry*, 2021, vol. 13, no. 6, pp. 973. <https://doi.org/10.3390/sym13060973>
- Zavadskas E.K., Podvezko V. Integrated determination of objective criteria weights in MCDM. *International Journal of Information Technology & Decision Making*, 2016, vol. 15, no. 2, pp. 267–283. <https://doi.org/10.1142/S0219622016500036>

10. Podvezko V., Zavadskas E.K., Podvezko A. An extension of the new objective weight assessment methods CILOS and IDOCRIW to fuzzy MCDM // *Economic Computation and Economic Cybernetics Studies and Research*. 2020. N 2. P. 59–75. <https://doi.org/10.24818/18423264/54.2.20.04>
11. Shuai D., Zongzhun Z., Yongji W., Lei L. A new angular method to determine the objective weights // *Proc. of the 24th Chinese Control and Decision Conference (CCDC)*. 2012. P. 3889–3892. <https://doi.org/10.1109/CCDC.2012.6244621>
12. Li G., Chi G. A new determining objective weights method-gini coefficient weight // *Proc. of the First International Conference on Information Science and Engineering*. 2009. P. 3726–3729. <https://doi.org/10.1109/ICISE.2009.84>
13. Rao R.V., Patel B.K. A subjective and objective integrated multiple attribute decision making method for material selection // *Materials & Design*. 2010. V. 31. N 10. P. 4738–4747. <https://doi.org/10.1016/j.matdes.2010.05.014>
14. Соловьев Д.С., Соловьева И.А. Разработка и исследование системы оптимального управления гальваническими процессами в ваннах со многими анодами // *Мехатроника, автоматизация, управление*. 2022. Т. 23. № 9. С. 462–472. <https://doi.org/10.17587/mau.23.462-472>
15. Solovjev D.S., Solovjeva I.A., Konkina V.V., Litovka Y.V. Improving the uniformity of the coating thickness distribution during electroplating treatment of products using multi anode baths // *Materials Today: Proceedings*. 2019. V. 19. N 5. P. 1895–1898. <https://doi.org/10.1016/j.matpr.2019.07.036>
16. Соловьев Д.С., Потлов А.Ю., Литовка Ю.В. Снижение неравномерности распределения толщины гальванического покрытия с использованием отключаемых анодных секций при реверсировании тока // *Теоретические основы химической технологии*. 2019. Т. 53. № 1. С. 102–111. <https://doi.org/10.1134/S0040357118060155>
17. Solovjev D.S., Solovjeva I.A., Konkina V.V. Software development for the optimal parts location in the bath space with the purpose to reduce the non-uniformity of the coating thickness // *Proc. of the 6th International Conference on Industrial Engineering (ICIE 2020)*. 2021. P. 85–93. https://doi.org/10.1007/978-3-030-54817-9_10
18. Пчелинцева И.Ю., Литовка Ю.В. Система автоматизированного управления процессом нанесения гальванического покрытия в ванне с токопроводящим экраном // *Мехатроника, автоматизация, управление*. 2022. Т. 23. № 4. С. 188–196. <https://doi.org/10.17587/mau.23.188-196>
19. Karimian N., Hashemi P., Afkhami A., Bagheri H. The principles of bipolar electrochemistry and its electroanalysis applications // *Current Opinion in Electrochemistry*. 2019. V. 17. P. 30–37. <https://doi.org/10.1016/j.coelec.2019.04.015>
20. Volgin V.M., Lyubimov V.V., Gnidina I.V., Kabanova T.B., Davydov A.D. Effect of anode shape on uniformity of electrodeposition onto resistive substrates // *Electrochimica Acta*. 2017. V. 230. P. 382–390. <https://doi.org/10.1016/j.electacta.2017.02.015>
21. Gardziejczyk W., Zabicki P. Normalization and variant assessment methods in selection of road alignment variants — case study // *Journal of Civil Engineering and Management*. 2017. V. 23. N 4. P. 510–523. <https://doi.org/10.3846/13923730.2016.1210223>
22. Соловьев Д.С., Соловьева И.А. Определение весовых коэффициентов с заданной согласованностью в многокритериальных задачах принятия решений. Свидетельство о регистрации программы для ЭВМ RU2022665895. 23.08.22.
10. Podvezko V., Zavadskas E.K., Podvezko A. An extension of the new objective weight assessment methods CILOS and IDOCRIW to fuzzy MCDM. *Economic Computation and Economic Cybernetics Studies and Research*, 2020, no. 2, pp. 59–75. <https://doi.org/10.24818/18423264/54.2.20.04>
11. Shuai D., Zongzhun Z., Yongji W., Lei L. A new angular method to determine the objective weights. *Proc. of the 24th Chinese Control and Decision Conference (CCDC)*, 2012, pp. 3889–3892. <https://doi.org/10.1109/CCDC.2012.6244621>
12. Li G., Chi G. A New Determining Objective Weights Method-Gini Coefficient Weight. *Proc. of the First International Conference on Information Science and Engineering*, 2009, pp. 3726–3729. <https://doi.org/10.1109/ICISE.2009.84>
13. Rao R.V., Patel B.K. A subjective and objective integrated multiple attribute decision making method for material selection. *Materials & Design*, 2010, vol. 31, no. 10, pp. 4738–4747. <https://doi.org/10.1016/j.matdes.2010.05.014>
14. Solovjev D.S., Solovjeva I.A. Development and research of the optimal control system of electroplating processes in multi-anode baths. *Mekhatronika, Avtomatizatsiya, Upravlenie*, 2022, vol. 23, no. 9, pp. 462–472. (in Russian). <https://doi.org/10.17587/mau.23.462-472>
15. Solovjev D.S., Solovjeva I.A., Konkina V.V., Litovka Y.V. Improving the uniformity of the coating thickness distribution during electroplating treatment of products using multi anode baths. *Materials Today: Proceedings*, 2019, vol. 19, no. 5, pp. 1895–1898. <https://doi.org/10.1016/j.matpr.2019.07.036>
16. Solovjev D.S., Potlov A.Y., Litovka Y.V. Reduction of nonuniformity in the thickness of a galvanic coating using disableable anode sections under current reversal conditions. *Theoretical Foundations of Chemical Engineering*, 2019, vol. 53, no. 1, pp. 97–106. <https://doi.org/10.1134/S004057951806012X>
17. Solovjev D.S., Solovjeva I.A., Konkina V.V. Software development for the optimal parts location in the bath space with the purpose to reduce the non-uniformity of the coating thickness. *Proceedings of the 6th International Conference on Industrial Engineering (ICIE 2020)*, 2021, pp. 85–93. https://doi.org/10.1007/978-3-030-54817-9_10
18. Pchelintseva I.Yu., Litovka Yu.V. Automated control system for the process of electroplating in a bath with a non-conductive of electric current screen. *Mekhatronika, Avtomatizatsiya, Upravlenie*, 2022, vol. 23, no. 4, pp. 188–196. (in Russian). <https://doi.org/10.17587/mau.23.188-196>
19. Karimian N., Hashemi P., Afkhami A., Bagheri H. The principles of bipolar electrochemistry and its electroanalysis applications. *Current Opinion in Electrochemistry*, 2019, vol. 17, pp. 30–37. <https://doi.org/10.1016/j.coelec.2019.04.015>
20. Volgin V.M., Lyubimov V.V., Gnidina I.V., Kabanova T.B., Davydov A.D. Effect of Anode Shape on Uniformity of Electrodeposition onto Resistive Substrates. *Electrochimica Acta*, 2017, vol. 230, pp. 382–390. <https://doi.org/10.1016/j.electacta.2017.02.015>
21. Gardziejczyk W., Zabicki P. Normalization and variant assessment methods in selection of road alignment variants — case study. *Journal of Civil Engineering and Management*, 2017, vol. 23, no. 4, pp. 510–523. <https://doi.org/10.3846/13923730.2016.1210223>
22. Solovjev D.S., Solovjeva I.A. Determination of weight coefficients with a given consistency in multicriteria decision-making problems. *Computer program registration certificate RU2022665895*, 2022.

Автор

Соловьев Денис Сергеевич — кандидат технических наук, доцент, доцент, Тамбовский государственный университет имени Г.Р. Державина, Тамбов, 392036, Российская Федерация, 57201777025@mail.ru, <https://orcid.org/0000-0001-6613-3218>, solovjevdenis@mail.ru

Author

Denis S. Solovjev — PhD, Associate Professor, Associate Professor, Derzhavin Tambov State University, Tambov, 392036, Russian Federation, 57201777025@mail.ru, <https://orcid.org/0000-0001-6613-3218>, solovjevdenis@mail.ru

Статья поступила в редакцию 01.11.2022
Одобрена после рецензирования 15.12.2022
Принята к печати 31.01.2023

Received 01.11.2022
Approved after reviewing 15.12.2022
Accepted 31.01.2023



Работа доступна по лицензии
Creative Commons
«Attribution-NonCommercial»

doi: 10.17586/2226-1494-2023-23-1-169-177

УДК 621.376.4; 621.396.944; 621.396.945

Оценка границ применимости и методов модуляции ближнепольной магнитной связи

Алексей Юрьевич Гришенцев¹✉, Вячеслав Александрович Горошков²,
Роман Ильич Чернов³

^{1,2,3} Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация

¹ AGrishentsev@yandex.ru✉, <https://orcid.org/0000-0003-1373-0670>

² gorosvia@ya.ru, <https://orcid.org/0000-0001-9950-5778>

³ aeijo@mail.ru, <https://orcid.org/0000-0001-9361-1238>

Аннотация

Предмет исследования. Одним из актуальных направлений исследований в области беспроводной связи является разработка ближнепольных магнитных систем и средств передачи сообщений в средах или сквозь среды, значительно поглощающие электромагнитное поле. К таким средам относятся: вода, грунт, строения. В проводящих средах ослабление магнитного поля возрастает с увеличением частоты. Для организации каналов связи сквозь проводящую среду, например морскую воду, применяют электромагнитное излучение крайне низких и сверхнизких частот от 3 до 300 Гц. Применение излучения в таких диапазонах частот сопряжено с большими размерами передающих и приемных антенн. Метод ближнепольной связи позволяет существенно уменьшить габариты приемных и излучающих антенн и потребляемую передатчиком мощность. Существенным ограничением ближнепольной длинноволновой связи является невысокая скорость передачи сообщений и небольшая, до десятков метров, дальность связи. **Метод.** Принцип действия предложенной системы связи основан на использовании магнитной компоненты электромагнитного поля. Передающим элементом такой системы служит катушка индуктивности с сердечником. Прием осуществляется датчиком магнитного поля в виде магнита, закрепленного на торсионном подвесе. Магнит совмещен с зеркалом, отражающим лазерный луч. Под действием внешнего магнитного поля магнит совершает крутильные движения, которые приводят к изменению угла отражения лазерного луча от зеркальной поверхности магнита. Отраженный сигнал регистрируется линейным фотоприемником. Оценка ослабления и затухания магнитного поля при передаче излучения из диэлектрической в проводящую среду выполнена на основе решения уравнений Максвелла. Разработаны трехпозиционная бинарная фазовая манипуляция и модифицированная трехпозиционная бинарная фазовая манипуляция, обеспечивающие противоположное расположение сигнальных символов, высокую информационную плотность сообщений, локализацию энергии излучаемого сигнала в низкочастотной области и повышение дальности связи. **Основные результаты.** Эксперименты показали, что в результате использования модифицированного вида манипуляции в сравнении с трехпозиционной бинарной фазовой манипуляцией удалось увеличить дальность связи на 10 % при неизменной высокой надежности доставки сообщений. Модельные оценки, выполненные в соответствии с предложенной методикой расчета ослабления и затухания магнитного поля при распространении в слоистых средах, подтверждены экспериментальными измерениями. **Практическая значимость.** Результаты исследований могут найти применение при решении задач локального развертывания защищенных систем ближнепольной связи при обеспечении беспроводной коммуникации через среды, поглощающие электромагнитное поле.

Ключевые слова

магнитная связь, бинарная фазовая манипуляция, передача сообщений, магнитное экранирование, распространение радиоволн

Ссылка для цитирования: Гришенцев А.Ю., Горошков В.А., Чернов Р.И. Оценка границ применимости и методов модуляции ближнепольной магнитной связи // Научно-технический вестник информационных технологий, механики и оптики. 2023. Т. 23, № 1. С. 169–177. doi: 10.17586/2226-1494-2023-23-1-169-177

Assessment of the limits of applicability and methods of modulation of near-field magnetic coupling

Alexey Yu. Grishentsev¹✉, Viacheslav A. Goroshkov², Roman I. Chernov³

^{1,2,3} ITMO University, Saint Petersburg, 197101, Russian Federation

¹ AGrishentsev@yandex.ru✉, <https://orcid.org/0000-0003-1373-0670>

² gorosvia@ya.ru, <https://orcid.org/0000-0001-9950-5778>

³ aeijo@mail.ru, <https://orcid.org/0000-0001-9361-1238>

Abstract

The development of near-field magnetic systems and means of transmitting messages through media that significantly absorb the electromagnetic field is one of the topical areas of research in the field of wireless communication. These lossy media include water, soil, buildings. The attenuation of the magnetic field in conducting media increases with increasing frequency. To organize communication channels through a conductive medium such as sea water electromagnetic radiation of extremely low frequencies and ultra-low frequencies from 3 Hz to 300 Hz is used. Traditional communication methods due to electromagnetic radiation in that frequency ranges require large sizes of transmitting and receiving antennas. The near-field communication method makes possible significant reduction both the dimensions of receiving and emitting antennas and the transmitter power consumption. A significant limitation of near-field long-wave communication is the low bandwidth and small, up to tens of meters, communication range. The operating principle of the proposed communication system is based on the use of the magnetic component of an electromagnetic field. Transmitting element in proposed system is a solenoid with a magnetic core. Receiving magnetic field sensor is a magnet fixed on a torsion suspension. The magnet is combined with a mirror reflecting the laser beam. Rotation of the magnet under the action of an external magnetic field leads to a change in the angle of reflection of the laser beam from the mirror surface of the magnet. The reflected signal is recorded by a linear photodetector. The attenuation of the magnetic field during the transmission of radiation from a dielectric to a conducting medium was evaluated with the solution of Maxwell's equations. A three-position binary phase shift keying and a modified three-position binary phase shift keying are developed and substantiated. The proposed solutions provide the opposite arrangement of signal symbols, high message information density, localization of the emitted signal energy in low-frequency region and an increase in communication range. Experiments had shown that usage of modified keying type shown an increase the communication range by 10 % with the same reliability of message delivery in comparison with three-position binary phase keying. The estimates of the weakening and attenuation of the magnetic field during propagation in layered media obtained from the simulation are confirmed by experimental measurements. The results of research could be used in solving problems of local deployment of secure near-field communication systems through media that absorb an electromagnetic field.

Keywords

magnetic coupling, binary phase manipulation, message transmission, magnetic shielding, radio wave propagation

For citation: Grishentsev A.Yu., Goroshkov V.A., Chernov R.I. Assessment of the limits of applicability and methods of modulation of near-field magnetic coupling. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 2023, vol. 23, no. 1, pp. 169–177 (in Russian). doi: 10.17586/2226-1494-2023-23-1-169-177

Введение

Одно из актуальных направлений исследований в области систем беспроводной связи — разработка методов и средств передачи сообщений в средах или сквозь среды, значительно поглощающие электромагнитное поле. В частности, подобные решения востребованы для организации беспроводной передачи информации сквозь толщу воды, в том числе морской, грунта, бетонных сооружений и экранирующих конструкций. Магнитное экранирование реализуется за счет разнообразных кожухов и экранов, изготовленных из материалов, магнитное сопротивление которых меньше, чем магнитное сопротивление внешней среды [1]. В проводящих средах ослабление магнитного поля возрастает с увеличением частоты и происходит за счет индуцированных токов Фуко [2]. Полное экранирование магнитного поля обеспечивают сверхпроводники. Для организации каналов связи¹ сквозь проводящую среду, например, морскую воду, обычно применяют крайне

низкие частоты в диапазоне от 3 до 30 Гц и сверхнизкие частоты — от 30 до 300 Гц. В некоторых случаях используют частоты менее 3 Гц. Для организации связи на сверхнизких и крайне низких частотах с возможностью передачи сообщений объектам, находящимся под водой, применяют протяженные антенно-фидерные системы, габаритные размеры которых могут варьироваться от сотен метров до единиц километров. Питание таких антенно-фидерных систем требует значительной энергии, исчисляемой сотнями киловатт и мегаваттами в зависимости от реализации и режимов работы. Излучаемые радиоволны могут несколько раз огибать естественный волновод, образованный поверхностью Земли и ионосферой, прежде чем будут поглощены, и их энергия трансформируется в тепло [3]. При таком виде передачи информации не обеспечивается локальность развертывания и действия систем связи. Альтернативные методы связи за счет электромагнитного излучения с локальным действием, в том числе за счет магнитной компоненты электромагнитного поля, предлагают отечественные разработчики. Так,

¹ Рекомендация МСЭ-R V.431-7* Номенклатура диапазонов частот и длин волн, используемых в электросвязи [Электронный ресурс]. URL: https://www.itu.int/dms_pubrec/

[itu-r/rec/v/R-REC-V.431-7-200005-S!!PDF-R.pdf](https://www.itu.int/rec/v/R-REC-V.431-7-200005-S!!PDF-R.pdf) (дата обращения: 03.12.2022).

например, АО «Научно-производственное предприятие «Радиосвязь» предлагает комплекс ближнепольной магнитной связи¹ для решения задач в интересах Министерства обороны Российской Федерации в промышленности, медицине и др. Зарубежные коллективы и отдельные исследователи предлагают свои решения в области магнитной связи, например в [4] представлена магнитоиндукционная система связи, обеспечивающая скорость передачи сообщений 100–300 бит/с и дальность связи до 400 м. В работе [5] рассмотрена магнитная система связи, действующая в средах воздух-вода (пресная), с дальностью действия до 30 м. В [6] выполнен обзор систем магнитоиндукционной ближнепольной связи, отмечены ее особенности и преимущества.

Цель настоящей работы — разработка методов поддержки и обеспечения системы связи за счет магнитной компоненты электромагнитного поля.

Предложен вид модифицированной трехпозиционной бинарной фазовой манипуляции, обеспечивающий противоположное расположение сигнальных символов и высокую информационную плотность сообщений. Данный вид является новизной для системы магнитной связи, в которой использован в качестве детектора магнитооптический датчик.

В рамках работы представлен метод оперативной оценки ослабления и затухания магнитного поля при распространении излучения из диэлектрической среды в проводящую. Отличительные особенности разработанной системы связи: локальность действия и развертывания; относительно небольшое энергопотребление. Например, экспериментальная установка имеет объем приемного устройства не более 0,25 дм³, объем передающего устройства (усилитель и магнитоиндукционный излучатель) — не более 4,5 дм³. При мощности излучателя 100 Вт обеспечивается дальность связи в воздушной среде до 25 м, в средах воздух-вода (вода морская составляет 95 % от длины всей трассы) до 16 м. При этом дальность действия можно увеличивать за счет наращивания мощности излучателя и использования излучателей с более «узкой» диаграммой направленности.

Принцип действия магнитной системы связи

Система магнитной связи образована передатчиком и приемником. В свою очередь передатчик состоит из модуля преобразования бинарных сообщений в модулированный сигнал, усилителя модулированного сигнала и магнитного излучателя, выполненного в виде двояной катушки индуктивности с сердечником. Приемник включает датчик магнитного поля, оптический усилитель, фотоприемник, систему детектирования на микроконтроллере и интерфейс, выдающую полученные сообщения в бинарном виде [7]. В качестве датчика магнитного поля использован магнит с зеркальной поверхностью на торсионном подвесе (не витой и нера-

стяжимой кевларовой нити). Падающий на зеркальную поверхность лазерный луч изменяет свое положение (угол отражения) в зависимости от положения магнита. Отраженный от магнита лазерный луч проходит систему зеркал, позволяющую в относительно компактном объеме добиться значительного изменения в положении луча. Таким образом, реализуется усиление отраженного сигнала. Отраженный лазерный луч регистрируется фотоприемной линейкой, которая отслеживает смещение пятна от лазерного луча. С фотоприемной линейки аналоговый сигнал поступает на микроконтроллер, выполняющий функцию аналого-цифрового преобразования и детектирования информационного сообщения.

Метод оценки границ применения и методы модуляции в системе ближнепольной магнитной связи

Для оценки границ применимости системы магнитной связи и особенностей распространения электромагнитной волны сквозь неоднородные среды в качестве источника излучения используем единичный виток с током (рис. 1). Магнитная индукция витка с током на торцах эквивалентна индукции на торцах индукционного излучателя. Запишем уравнение, выражающее приращение вектора магнитной индукции $d\mathbf{B}$, обусловленное протеканием тока i , через участок проводника $d\mathbf{l}$:

$$d\mathbf{B} = \frac{\mu_0 \mu i}{4\pi} \frac{[d\mathbf{l} \times \mathbf{r}]}{|\mathbf{r}|^3}, \quad (1)$$

где $\mu_0 = 1,257 \cdot 10^{-6}$ Гн/м — магнитная постоянная; μ — относительная магнитная проницаемость среды; \mathbf{r} — радиус-вектор, проведенный из участка проводника $d\mathbf{l}$ в данную точку пространства. На чертеже (рис. 1) приведен виток с током в прямоугольной $Oxyz$ системе координат. R — радиус единичного витка, который направлен в сторону элементарного сонаправленного с током i участка $d\mathbf{l}$; φ — угол, отложенный от оси Oy и определяющий положение участка $d\mathbf{l}$; $\mathbf{B}(x,y,z)$ — вектор магнитной индукции в точке с координатами (x,y,z) .

Учитывая осевую симметрию (рис. 1) магнитного поля витка с током, отметим, что аналогичной симметрией будет обладать и многovitковый соленоид.

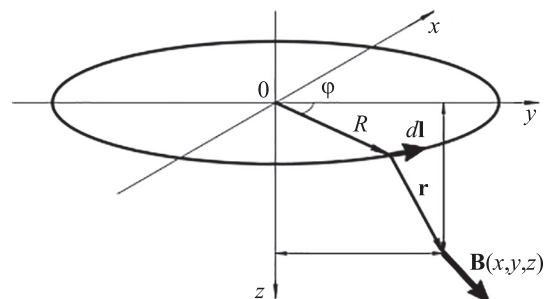


Рис. 1. Единичный виток с током в прямоугольной системе координат

Fig. 1. Single circular current loop in Cartesian coordinate system

¹ Системы ближнепольной магнитной связи. Акционерное общество «Научно-производственное предприятие «Радиосвязь» [Электронный ресурс]. URL: <https://krtz.su/node/286> (дата обращения: 21.11.2022).

Запишем решение (1) для частного случая в плоскости yOz , предложенное в работе [8], в виде:

$$\begin{cases} B_x(0,y,z) = \frac{\mu_0 \mu R z}{4\pi} \int_0^{2\pi} \frac{\sin(\varphi)}{(R^2 + y^2 + z^2 - 2Ry \cos(\varphi))^{3/2}} d\varphi \\ B_y(0,y,z) = -\frac{\mu_0 \mu R z}{4\pi} \int_0^{2\pi} \frac{\cos(\varphi)}{(R^2 + y^2 + z^2 - 2Ry \cos(\varphi))^{3/2}} d\varphi \\ B_z(0,y,z) = \frac{\mu_0 \mu R}{4\pi} \int_0^{2\pi} \frac{R - y \cos(\alpha)}{(R^2 + y^2 + z^2 - 2Ry \cos(\varphi))^{3/2}} d\varphi \end{cases} \quad (2)$$

Решение уравнений Максвелла в общем виде позволяет найти электромагнитное поле в любой точке пространства [9]. Численное решение обычно реализуется методами конечных разностей, конечных или граничных элементов. Для современных программных средств, реализующих подобные методы, для которых разрабатывается система магнитной связи, не всегда доступно применение таких решений в полевых условиях. Для решения данной задачи выведем выражение для оценки ослабления и затухания магнитного поля при прохождении из диэлектрика в проводник и распространения по проводнику.

Рассмотрим случай плоской электромагнитной волны (рис. 2). Выберем систему координат $(Oxyz)$ таким образом, что направления векторов: \mathbf{E} и \mathbf{H} — напряженности электрического и магнитного полей, а также и $\mathbf{S} = [\mathbf{E} \times \mathbf{H}]$ — вектора Умова–Пойтинга, совпадают с базисными векторами системы координат. На рис. 2 представлены обозначения для первой и второй сред: ϵ_1, ϵ_2 — диэлектрические проницаемости; μ_1, μ_2 — магнитные проницаемости; γ_1, γ_2 — удельные электропроводности.

Тогда из уравнений Максвелла для плоской электромагнитной волны $\text{rot} \mathbf{H} = \gamma \mathbf{E}$ и $\text{rot} \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t} = -\mu \frac{\partial \mathbf{H}}{\partial t}$, следует:

$$\begin{cases} -\frac{\partial H_y}{\partial z} = \gamma E_x \\ \frac{\partial E_x}{\partial z} = -\mu \frac{\partial H_y}{\partial t} \end{cases}, \quad (3)$$

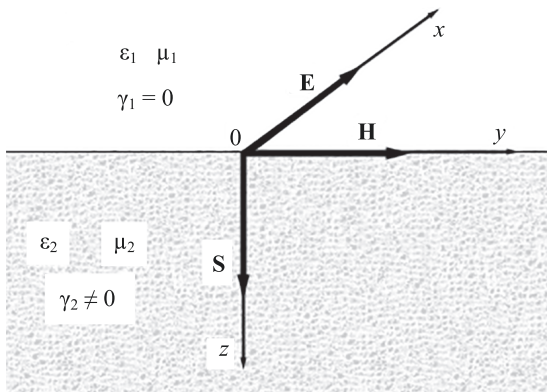


Рис. 2. Распространение плоской электромагнитной волны через границу раздела двух сред

Fig. 2. Propagation of a plane electromagnetic wave through the interface between two media

где t — время.

Решение системы уравнений (3) известно из работы [10]:

$$\begin{cases} \dot{H}_y = \dot{H}_0 e^{-kz} e^{-jkz} e^{-j\omega t} \\ \dot{E}_x = \frac{1}{\gamma} (1+j) k \dot{E}_0 e^{-kz} e^{-jkz} e^{-j\omega t} \end{cases}, \quad (4)$$

где \dot{H}_0 и \dot{E}_0 — комплексные напряженности магнитного и электрического полей при $z=0$; $k = \sqrt{\frac{\omega \mu \gamma}{2}}$. Множитель e^{-kz} определяет затухание, а e^{-jkz} — фазу волны в среде от z .

На основании решения (3), определим отношение между падающими ($\dot{E}_{\phi 1}, \dot{H}_{\phi 1}$) из диэлектрика на поверхность проводника волнами, отраженными ($\dot{E}_{\psi 1}, \dot{H}_{\psi 1}$) и преломленными ($\dot{E}_{\phi 2}, \dot{H}_{\phi 2}$) волнами, прошедшими в проводящую среду. Примем, что: $Z_2 = \frac{\dot{E}_x}{\dot{H}_y} = (1+j) \sqrt{\frac{\omega \mu_2}{2\gamma_2}}$ — комплексное волновое сопротивление проводящей среды; $Z_1 = \sqrt{\frac{\mu_1}{\epsilon_1}}$ — волновое сопротивление для падающей и отраженной волн в диэлектрике. Тогда запишем:

$$\begin{cases} \dot{E}_{\phi 2} = \frac{2Z_2}{Z_1 + Z_2} \dot{E}_{\phi 1} \\ \dot{H}_{\phi 2} = \frac{2Z_1}{Z_1 + Z_2} \dot{H}_{\phi 1} \end{cases}. \quad (5)$$

На достаточно большом расстоянии, виток с током можно считать точечным источником электромагнитного поля. Используя уравнения (2), (4), (5) запишем общее выражение оценки затухания и ослабления магнитного поля при прохождении из диэлектрика в проводящую среду и движении по проводящей среде в предположении, что источник магнитного поля мал по сравнению с рассматриваемыми расстояниями z . Тогда получим, что в дальней зоне витка с током:

$$\dot{H}_{\phi 2} = \frac{2Z_1}{Z_1 + Z_2} e^{-zk} \frac{1}{z^3} \dot{H}_{\phi 1}. \quad (6)$$

Решим следующую задачу: разработаем метод манипуляции магнитного сигнала. При разработке способа манипуляции [10, 11] представлены следующие требования и рекомендации:

- 1) с учетом крайне низкой частоты передачи сообщений необходимо выбрать вид манипуляции, который обеспечит высокую информационную плотность манипулированного сообщения;
- 2) желательно использовать вид манипуляции, который сможет работать без явной синхронизации передатчика и приемника и при некотором изменении несущей частоты;
- 3) необходимо учесть конструктивные особенности приемного устройства для снижения вероятности ошибки при приеме сообщений.

В результате анализа различных видов манипуляции с учетом требований и рекомендаций, в качестве осно-

вы выберем способ бинарной фазовой манипуляции (БФМ). БФМ доработан в вид — трехпозиционной бинарной фазовой манипуляции (ТБФМ), задаваемой выражением:

$$x_{\text{ТБФМ}}(t) = \begin{cases} A_{\max} \sin\left(2\pi f\left(t - \left\lfloor \frac{t}{T} \right\rfloor T\right)\right), & s[n] = 1 \\ -A_{\max} \sin\left(2\pi f\left(t - \left\lfloor \frac{t}{T} \right\rfloor T\right)\right), & s[n] = 0 \end{cases}, \quad (7)$$

где $\omega = 2\pi f$ — несущая частота; $n = 1, \dots, N$ — индекс дискретного сообщения $s[n]$, выдаваемого каналным кодером; T — длительность одного сигнального символа; A_{\max} — амплитуда сигнального символа; $\left\lfloor \frac{t}{T} \right\rfloor$ — округление в меньшую сторону. Пример манипуляции приведен на рис. 3.

Разработанный вид ТБФМ удовлетворяет требованиям и рекомендациям (1)–(3), но в то же время содержит точки с резкой сменой фазы, обозначенные красными окружностями. Видно резкое изменение фазы, которое образует устранимый разрыв первого рода, что,

$$x_{\text{ТБФМ}}(t) = \begin{cases} -0,5 - 0,5\sin(\omega t - 0,5\pi), & a(t) < N, s[a(t)] = 0, s[b(t)] = 0 \\ -\sin(0,5\omega t + \pi(a(t) \bmod 2)), & a(t) < N, s[a(t)] = 0, s[b(t)] = 1 \\ +\sin(0,5\omega t + \pi(a(t) \bmod 2)), & a(t) < N, s[a(t)] = 1, s[b(t)] = 0 \\ +0,5 + 0,5\sin(\omega t - 0,5\pi), & a(t) < N, s[a(t)] = 1, s[b(t)] = 1 \\ -0,5 - 0,5\sin(\omega t - 0,5\pi), & s\left[1 + \left\lfloor \frac{t}{T} \right\rfloor\right] = 0, t \leq \frac{T}{2} \\ +0,5 + 0,5\sin(\omega t - 0,5\pi), & s\left[1 + \left\lfloor \frac{t}{T} \right\rfloor\right] = 1, t \leq \frac{T}{2} \\ -0,5 - 0,5\sin(\omega t - 0,5\pi), & a(t) = N, s[a(t)] = 0 \\ +0,5 + 0,5\sin(\omega t - 0,5\pi), & a(t) = N, s[a(t)] = 1 \end{cases}, \quad (8)$$

где $a(t) = 1 + \left\lfloor \frac{t}{T} - 0,5 \right\rfloor$; $b(t) = 2 + \left\lfloor \frac{t}{T} - 0,5 \right\rfloor$.

Основные результаты

В плоскости yOz из выражения (2) компонента вектора магнитной индукции $B_x(0, y, z)$. Отметим, что в любой плоскости, параллельной оси Oz , нормальная компонента вектора магнитной индукции равна нулю

в свою очередь, в соответствии с теоремой Дирихле [12, 13], приводит к значительному расширению спектра Фурье сигнала $x_{\text{ТБФМ}}(t)$. Отметим, что разрыв происходит в критически значимой области для обеспечения прохождения пятном лазерного луча нейтрального положения на фотоприемной линейке. Установлено, что затухание магнитного сигнала с ростом частоты возрастает. Спектр сигнала, содержащего разрывы, возрастает настолько, что значимые гармоники превышают частотный диапазон, регистрируемый датчиком. В результате вероятность недостижения пятном лазерного луча нейтрального положения значительно возрастает, следовательно, возрастает вероятность ошибки при приеме сообщений.

Для устранения перечисленных недостатков предложен модифицированный вид трехпозиционной бинарной фазовой манипуляции (МТБФМ), основанный на сшивке гармонических функций $\sin(\omega t)$ и $\sin(0,5\omega t)$ в точке равенства значений функций и первой производной. Данные условия обеспечили концентрацию энергии сигнала в низкочастотной области.

(рис. 1). Исходя из этого, решение трехмерной задачи (1) сводится к решению плоской задачи (2), которая в силу центральной симметрии относительно начала координат приводит к необходимости поиска решения в одной четверти плоскости. В результате расчета (2), для значений радиуса витка $R = 0,05$ м и тока в витке 1 кА, построим модуль вектора магнитной индукции в плоскости yOz (рис. 4, а) и по осям Oy и Oz (рис. 4, б).

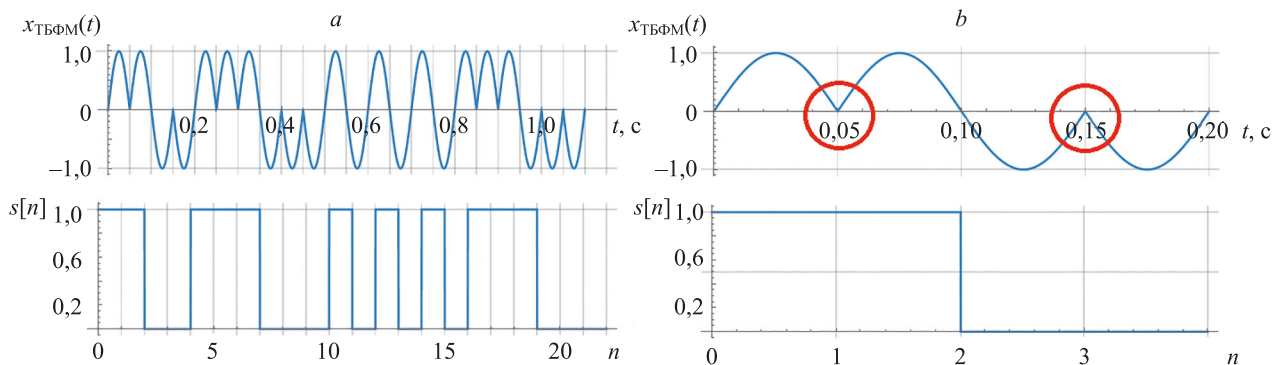


Рис. 3. Трехпозиционная бинарная фазовая манипуляция $x_{\text{ТБФМ}}(t)$ и бинарный сигнал сообщения $s[n]$ длительностью сообщения: 22 символа (а); четыре символа (б)

Fig. 3. Three-position binary phase shift keying $x_{\text{ТБФМ}}(t)$ and binary message signal $s[n]$

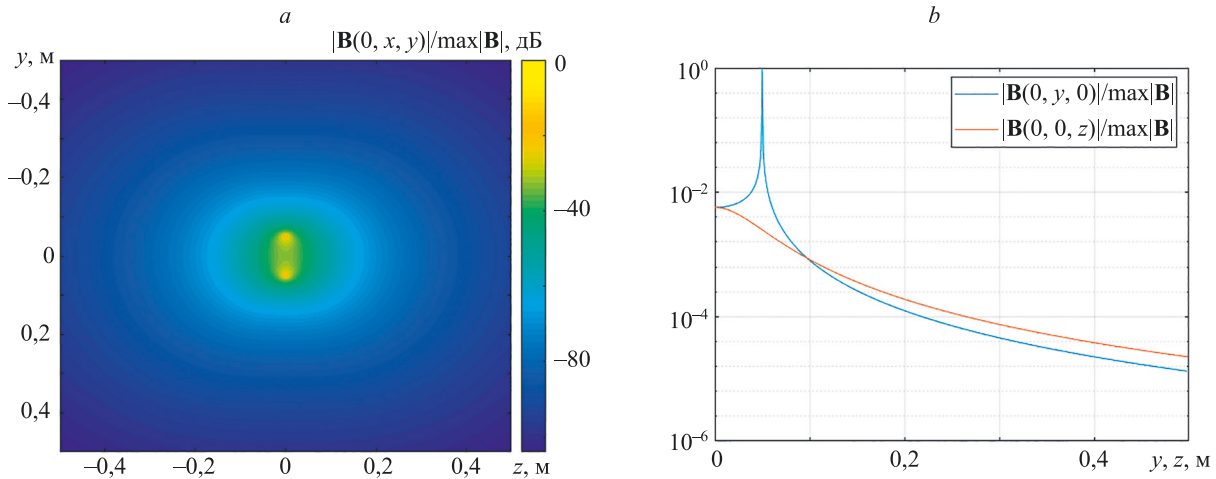


Рис. 4. Нормированный модуль вектора магнитной индукции в плоскости $y0z$ (a); график нормированного модуля вектора магнитной индукции по осям $0y$ и $0z$ (b). На рис. 4, b по вертикальной оси указана безразмерная величина

Fig. 4. Distribution of the normalized modulus of the magnetic induction vector in the $y0z$ plane (a); graph of the normalized modulus of the magnetic induction vector along the axes $0y$ and $0z$ (b). On fig. 4, b the dimensionless value is indicated along the vertical axis

Полученная картина поля (рис. 4, b) позволила сделать вывод, что при расстояниях от центра витка, больших чем $4R$, соотношения между векторными компонентами $B_y(x, y, z)$ и $B_z(x, y, z)$ приобретают определенную тенденцию, которая сохраняется далее. Таким образом, дальней зоной для единичного витка можно считать расстояния, большие чем $4R$, а из рис. 4, a видно, что при достаточно больших расстояниях от единичного витка его можно рассматривать как точечный источник.

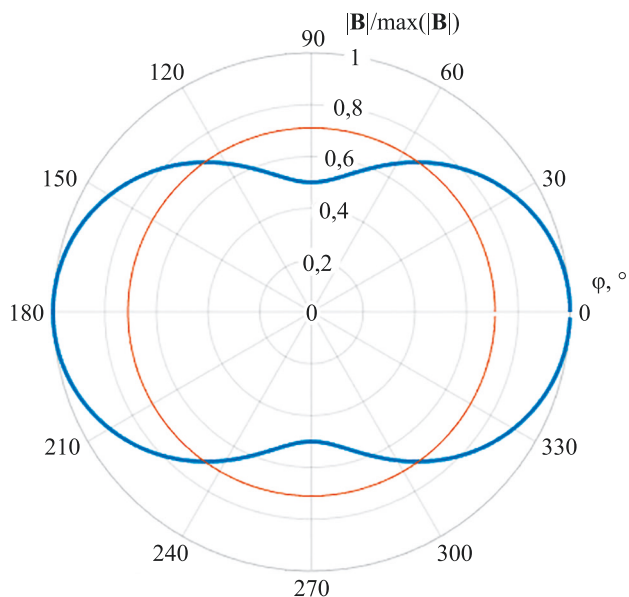


Рис. 5. Диаграмма направленности магнитного излучателя в дальней зоне ($100R$), красной линией обозначен уровень ослабления в 0,71 раз (-3 дБ), синей — модуль нормированного вектора магнитной индукции

Fig. 5. The radiation pattern of the magnetic emitter in the far zone ($100R$), the red line indicates the attenuation level of 0.71 times (-3 dB), the blue line indicates the module of the normalized magnetic induction vector

Диаграмма направленности единичного витка, как излучателя магнитного поля в полярных координатах, приведена на рис. 5. Ось единичного витка соответствует $\varphi = 0^\circ$.

На основании полученных выражений (2) и (6) произведем оценку затухания и ослабления индукции магнитного поля в направлении оси $0z$, при прохождении из диэлектрика в проводящую среду (морскую воду, почву) (рис. 2) и в воздухе¹. При построении графиков (рис. 6) в качестве источника магнитного поля используем единичный виток, расположенный на расстоянии R от границы раздела, при условии, что точка начала координат расположена на границе раздела сред.

Практическая реализация системы магнитной связи выполнена с использованием в качестве излучателя электромагнита в виде соленоида [14]. Соленоид состоит из двух катушек, каждая по 600 витков, проводящих синусоидальный ток с амплитудой 5 А, равный 6 кА·витков. Система в воздушной среде при скорости передачи 20 бит/с при экспериментальных исследованиях показала уверенный прием сообщений на расстоянии 25 м, что соответствует 135 дБ затухания (рис. 6). Таким образом, при заданном режиме работы излучателя и скорости передачи 20 бит/с, расстояние уверенного приема оценено: для морской воды около 16 м; для глинистой почвы — 17 м.

В случае магнитного экранирования для оценки ослабления магнитной составляющей электромагнитного поля предлагается использовать уже имеющиеся модели, например [1, 9], предоставляющие относительно простые методы оценки ослабления магнитного поля при различных формах и материалах экранов.

¹ Рекомендация МСЭ-R P.527-4 Электрические характеристики земной поверхности. Серия Р Распространение радиоволн. Электронная публикация Женева: МСЭ-R, 2018 [Электронный ресурс]. URL: https://www.itu.int/dms_pubrec/itu-r/rec/p/R-REC-P.527-4-201706-I!!PDF-R.pdf (дата обращения: 10.12.2022).

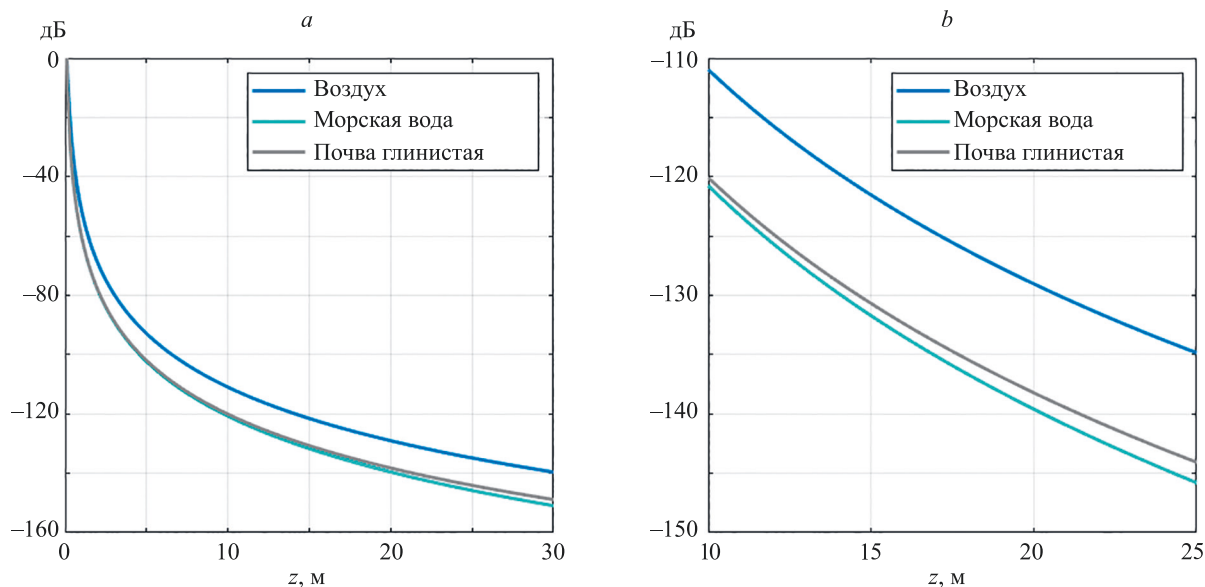


Рис. 6. График ослабления магнитного поля (а) и фрагмент графика в увеличенном масштабе (b)

Fig. 6. Attenuation of magnetic field induction in air, sea water and clay soil

Из (8) видно, что сшивка функций $f_1(t) = \pm 0,5 \pm 0,5\sin(\omega t - 0,5\pi)$ и $f_2(t) = \sin(0,5\omega t)$ происходит в точках равенства значений гармонических функций. Данные точки одновременно являются минимумами и максимумами гармонических функций \cos и \sin , где

все нечетные производные равны нулю, при: $f_1^{(k)}(t) = f_2^{(k)}(t) = 0$, при $k = 1, 3, 5, \dots$

Сравнительный анализ спектров МТБФМ и ТБФМ (рис. 7) показал, что в результате модификации удалось существенно снизить амплитуду высокочастотных ком-

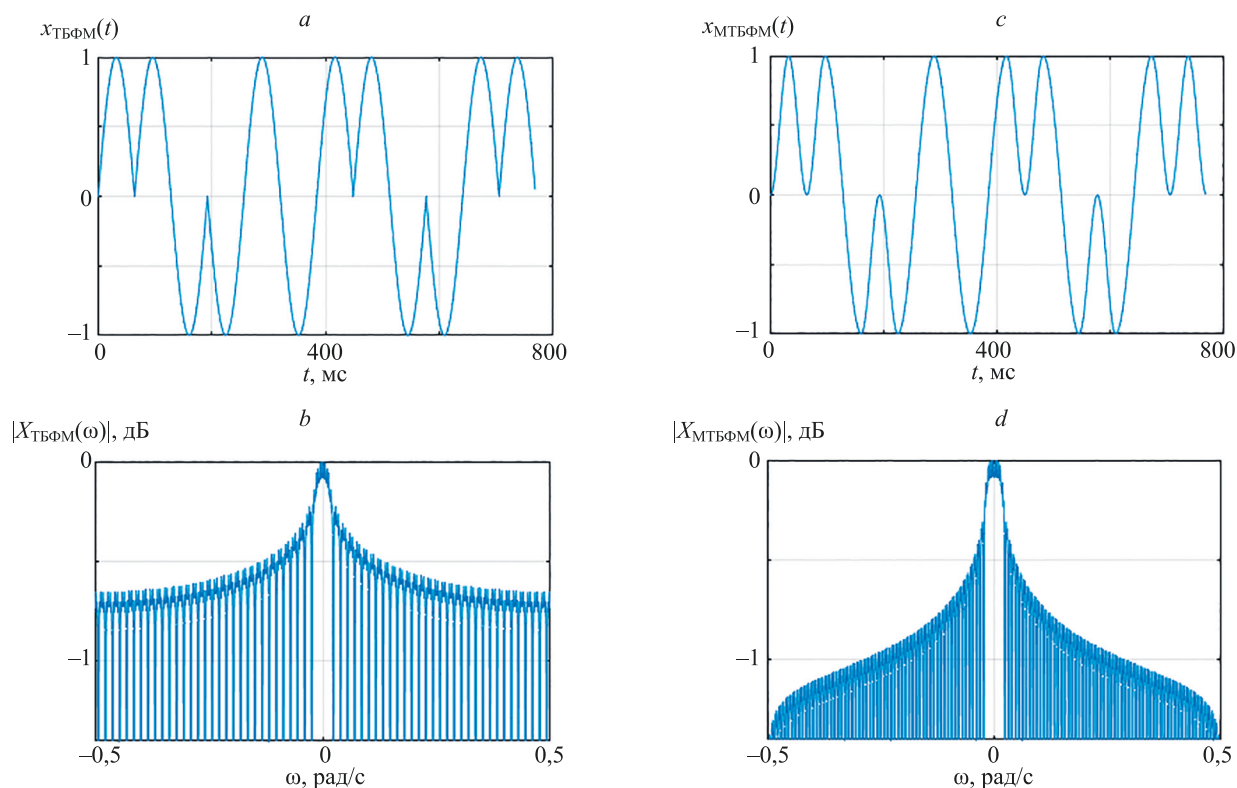


Рис. 7. Сравнение сигналов: трехпозиционной бинарной фазовой манипуляции $x_{\text{ТБФМ}}(t)$ (а), модифицированной трехпозиционной бинарной фазовой манипуляции $x_{\text{МТБФМ}}(t)$ (с) и их нормированных по частоте модулей спектров $|X_{\text{ТБФМ}}(\omega)|$ (b) и $|X_{\text{МТБФМ}}(\omega)|$ (d)

Fig. 7. Comparison of signals: $x_{\text{TBPSK}}(t)$ — three-position binary phase shift keying (a) with $x_{\text{MTBPSK}}(t)$ — modified three-position binary phase shift keying (c) and their frequency-normalized spectrum modulus $|X_{\text{TBPSK}}(\omega)|$ (b) and $|X_{\text{MTBPSK}}(\omega)|$ (d)

понент. Из (8) видно, что манипулированный сигнал образован гармоническими функциями двух несущих частот: ω и $0,5\omega$. В результате использования МТБФМ удалось повысить надежность системы магнитной связи за счет того, что лазерный луч проходит нейтральное положение на фотополоске. Из выражений (7)–(8) получено, что разработанные методы манипуляции используют противоположные сигналы, которые обеспечивают максимальное евклидово расстояние $2\sqrt{E}$ среди сигналов с равной энергией E . Данный результат определяет максимально возможную помехоустойчивость при бинарной манипуляции сигнальными символами равных энергий.

Вычисление МТБФМ (8) можно выполнить средствами любого микроконтроллера, имеющего аппаратные или программные функции для арифметики с плавающей точкой и способного вычислять гармонические функции. Альтернативным вариантом может быть использование вычисленных табличных значений гармонических функций в заданных точках.

Полученная МТБФМ удовлетворяет всем требованиям и рекомендациям к способу манипуляции бинарного сигнала для разрабатываемой системы магнитной связи.

Заключение

В ходе исследований разработан метод оценки затухания и ослабления магнитной компоненты электро-

магнитного поля при прохождении из диэлектрика в проводящую среду. На основе выведенного выражения (6) произведена сравнительная оценка ослабления и затухания магнитного поля при распространении в слоистых средах: воздух-воздух; воздух-морская вода; воздух-почва. Выполнены экспериментальные исследования, подтверждающие адекватность численных расчетов.

Разработаны два вида манипуляции цифрового сигнала: трехпозиционная бинарная фазовая манипуляция и модифицированная трехпозиционная бинарная фазовая манипуляция. Оба вида манипуляций позволяют без явной синхронизации производить передачу сообщений за счет возвращения чувствительного элемента датчика и лазерного луча в нейтральное положение после передачи каждого бита. За счет разработанного модифицированного вида манипуляции значительно сужен спектр манипулированного сообщения и уменьшены амплитуды высших гармоник в спектре сигнала, что приводит к концентрации энергии сигнала в низкочастотной области и способствует возвращению датчика магнитного поля в нейтральное положение после передачи каждого бита.

Практический эксперимент показал, что в результате использования модифицированного вида манипуляции удалось при неизменной высокой надежности доставки сообщений увеличить дальность связи на 10 % по сравнению с использованием трехпозиционной бинарной фазовой манипуляции.

Литература

1. Аполлонский С.М. Справочник по расчету электромагнитных экранов. Л.: Энергоатомиздат, 1988. 224 с.
2. Аполлонский С.М. Дифференциальные уравнения математической физики в электротехнике. СПб.: Питер, 2012. 352 с.
3. Яковлев О.И., Якубов В.П., Урядов В.П., Павел'ев А.Г. Распространение радиоволн: М.: УРСС ЛЕНАНД, 2009. 496 с.
4. Sojdehei J.J., Wrathall P.N., Dinn D.F. Magneto-inductive (MI) communications // MTS/IEEE Oceans 2001. An Ocean Odyssey: Conference Proceedings. V. 1. 2001. P. 513–519. <https://doi.org/10.1109/OCEANS.2001.968775>
5. Bogie I.S. Conduction and magnetic signalling in the sea a background review // Radio and Electronic Engineer. 1972. V. 42. N 10. P. 447–452. <https://doi.org/10.1049/ree.1972.0076>
6. Власов А.А., Родионов А.Ю. Перспективы использования систем подводной коммуникации на основе магнитной индукции (обзор) // Вестник инженерной школы ДВФУ. 2021. № 2(47). С. 36–49. <https://doi.org/10.24866/2227-6858/2021-2-5>
7. Гришенцев А.Ю., Коробейников А.Г., Горошков В.А., Чернов Р.И., Тихомиров А.В., Козин О.В. Разработка и моделирование магнитооптического датчика градиента магнитного поля с торсионным подвесом чувствительного элемента // Журнал радиоэлектроники [электронный журнал]. 2021. № 11. С. 8. <https://doi.org/10.30898/1684-1719.2021.11.4>
8. Перов Н.С., Грановский С.А., Стрелков Н.В., Шапаева Т.Б., Макарова Л.А., Шапаев Б.А. Изучение постоянного магнитного поля. Численное моделирование и эксперимент. М.: Изд-во МГУ, 2017. 23 с.
9. Демирчян К.С., Нейман Л.Р., Коровкин Н.В., Чечурин В.Л. Теоретические основы электротехники. В 3-х томах. Т. 3 / 4-е изд. СПб.: Питер, 2006. 377 с.
10. Прокис Дж. Цифровая связь: пер. с англ. М.: Радио и связь, 2000. 800 с.
11. Скляр Б. Цифровая связь. Теоретические основы и практическое применение / 2-е изд. М.: Вильямс, 2017. 1100 с.

References

1. Apollonskii S.M. *Handbook for the Calculation of Electromagnetic Screens*. Leningrad, Jenergoatomizdat Publ., 1988, 224 p. (in Russian)
2. Apollonskii S.M. *Differential Equations of Mathematical Physics in Electrical Engineering*. St. Petersburg, Piter Publ., 2012. 352 p. (in Russian)
3. Yakovlev O.I., Yakubov V.P., Uryadov V.P., Pavel'ev A.G. *Radiowaves Spread*. Moscow, Lenand Publ., 2009, 496 p. (in Russian)
4. Sojdehei J.J., Wrathall P.N., Dinn D.F. Magneto-inductive (MI) communications. *MTS/IEEE Oceans 2001. An Ocean Odyssey. Conference Proceedings. V. 1*, 2001, pp. 513–519. <https://doi.org/10.1109/OCEANS.2001.968775>
5. Bogie I.S. Conduction and magnetic signalling in the sea a background review. *Radio and Electronic Engineer*, 1972, vol. 42, no. 10, pp. 447–452. <https://doi.org/10.1049/ree.1972.0076>
6. Vlasov A., Rodionov A. Prospects for the usage of underwater communication systems based on magnetic induction (review). *FEFU: School of Engineering Bulletin*, 2021, no. 2(47), pp. 36–49. (in Russian). <https://doi.org/10.24866/2227-6858/2021-2-5>
7. Grishentsev A.Y., Korobeynikov A.G., Goroshkov V.A., Chernov R.I., Tikhomirov A.V., Kozin O.V. Development and modeling of a magneto-optical magnetic field gradient sensor with a torsion suspension of the sensor element. *Journal of Radio Electronics*, 2021, no. 11, pp. 8. (in Russian). <https://doi.org/10.30898/1684-1719.2021.11.4>
8. Perov N.S., Granovskii S.A., Strelkov N.V., Shapaeva T.B., Makarova L.A., Shapaev B.A. The Study of the Constant Magnetic Field. *Numerical Modeling and Experiment*. Moscow, MSU Publ., 2017, 23 p. (in Russian)
9. Demirchian K.S., Neiman L.R., Korovkin N.V., Chечурин V.L. *Theoretical Basics of Electrical Engineering. V. 3*. St. Petersburg, Piter Publ., 2006, 377 p. (in Russian)
10. Proakis J.G. *Digital Communications*. New York, McGraw-Hill Book Co., 1989.
11. Sklar B. *Digital Communications*. Pearson Education, 2009, 1164 p.

12. Ильин В.А., Позняк Э.Г. Основы математического анализа. В 2 частях. Ч. 1 / 5-е изд. М.: Наука, Физматлит, 1998. 616 с.
13. Письменный Д.Т. Конспект лекций по высшей математике. М.: Айрис-Пресс, 2010. 608 с.
14. Калантаров П.Л., Цейтлин Л.А. Расчет индуктивностей. Справочная книга. Л.: Энергоатомиздат, 1986. 488 с.

Авторы

Гришенцев Алексей Юрьевич — доктор технических наук, доцент, доцент, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, [sc 56321138400](https://orcid.org/0000-0003-1373-0670), <https://orcid.org/0000-0003-1373-0670>, AGrishentsev@yandex.ru

Горошков Вячеслав Александрович — аспирант, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, <https://orcid.org/0000-0001-9950-5778>, gorosvia@ya.ru

Чернов Роман Ильич — инженер, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, <https://orcid.org/0000-0001-9361-1238>, aeijo@mail.ru

12. Ilin V.A., Pozniak E.G. *Basics of Mathematical Analysis. Part 1*. Moscow, Nauka, Fizmatlit Publ., 1998, 616 p. (in Russian)
13. Pismennyi D.T. *Higher Mathematics: Lecture Notes*. Moscow, Ajris-Press Publ., 2010, 608 p. (in Russian)
14. Kalantarov P.L., Tseitlin L.A. *Inductance Calculation. Reference Book*. Leningrad, Jenergoatomizdat Publ., 1986, 488 p. (in Russian)

Authors

Alexey Yu. Grishentsev — D.Sc., Associate Professor, Associate Professor, ITMO University, Saint Petersburg, 197101, Russian Federation, [sc 56321138400](https://orcid.org/0000-0003-1373-0670), <https://orcid.org/0000-0003-1373-0670>, AGrishentsev@yandex.ru

Viacheslav A. Goroshkov — PhD Student, ITMO University, Saint Petersburg, 197101, Russian Federation, <https://orcid.org/0000-0001-9950-5778>, gorosvia@ya.ru

Roman I. Chernov — Engineer, ITMO University, Saint Petersburg, 197101, Russian Federation, <https://orcid.org/0000-0001-9361-1238>, aeijo@mail.ru

Статья поступила в редакцию 22.11.2022
Одобрена после рецензирования 03.01.2023
Принята к печати 26.01.2023

Received 22.11.2022
Approved after reviewing 03.01.2023
Accepted 26.01.2023



Работа доступна по лицензии
Creative Commons
«Attribution-NonCommercial»

doi: 10.17586/2226-1494-2023-23-1-178-191

УДК 533.6

Численно-аналитическое моделирование пропульсивного крыла и фюзеляжа аэротакси

Павел Викторович Булат¹✉, Антон Александрович Курнухин²,
Николай Васильевич Продан³

^{1,2,3} Севастопольский государственный университет, Севастополь, 299053, Российская Федерация

^{1,2,3} Балтийский государственный технический университет «ВОЕНМЕХ» им. Д.Ф. Устинова, Санкт-Петербург, 190005, Российская Федерация

¹ pavelbulat@mail.ru✉, <https://orcid.org/0000-0003-0099-9953>

² anton.kurnukhin@outlook.com, <https://orcid.org/0000-0003-4851-6594>

³ kolinti@mail.ru, <https://orcid.org/0000-0002-6383-039X>

Аннотация

Предмет исследования. Рассмотрена проблема создания аэродинамических профилей, создающих тягу при отборе воздуха с верхней поверхности крыла. Профили имеют тяговый участок за точкой отсоса пограничного слоя, давление на котором превосходит давление в окружающей среде. В последние 15–20 лет в мире активно исследуется концепция так называемого пропульсивного крыла, имеющего пониженное или нулевое сопротивление за счет отсоса пограничного слоя с верхней его поверхности. Такое крыло позволяет в несколько раз уменьшить аэродинамическое сопротивление самолета за счет ламинаризации пограничного слоя и минимизации дефекта скорости, связанного с вязким трением в пограничном слое, в следе за самолетом. Предложена методика численного моделирования профилей для пропульсивного крыла, построенных методом решения обратной задачи аэродинамики. Проектируемые профили имеют максимальную строительную высоту, оптимальное сочетание коэффициентов подъемной силы C_y и тяги C_T , создаваемой за счет отбора воздуха с поверхности крыла. Методика корректно предсказывает точку ламинарно-турбулентного перехода, поскольку характеристики профилей напрямую зависят от протяженности ламинарного участка. Исследована компоновка летательного аппарата, построенного по схеме пропульсивного летящего крыла сверхмалого удлинения с применением разработанных профилей. **Метод.** Проектирование аэродинамических профилей выполнено путем решения обратной задачи аэродинамики с последующей доводкой геометрии при помощи алгоритмов глобальной оптимизации. Расчеты проведены с использованием γ - Re_θ Transition Shear Stress Transport модели турбулентности Лэнгтри–Ментера, в которой существуют соотношения для критерия перемежаемости, что позволяет моделировать ламинарно-турбулентный переход. **Основные результаты.** Расчеты показали, что разработанные профили позволяют создать планер летательного аппарата с максимальным коэффициентом подъемной силы $C_{y_{max}}$, который превосходит на взлете и посадке данную силу у механизированного крыла с выпущенным закрылком. В горизонтальном полете C_y в три раза больше, чем у типичного крыла. Крыло с разработанными профилями имеет высокий пропульсивный коэффициент полезного действия за счет близости давления и скорости на тяговом участке профилей и внешнего течения. Одновременно тяговая поверхность пропульсивного крыла превосходит по площади в несколько раз площадь сопла или суммарное покрытие авиационных винтов. **Практическая значимость.** Разработанные профили и интегрированная аэродинамическая компоновка летательного аппарата успешно сочетаются с принципами построения распределенной силовой установки, а также создают устойчивость к повышенной турбулентности атмосферы при вертикальном взлете и посадке с экономичным горизонтальным полетом. Профили имеют важное преимущество по сравнению с традиционной механизацией крыла, так как не имеют подвижных частей, а увеличение или уменьшение подъемной силы регулируется изменением расхода отсасываемого воздуха.

Ключевые слова

аэродинамический профиль, математическое моделирование, оптимизация, пропульсивное крыло, численный эксперимент, энергетические методы увеличения подъемной силы

Благодарности

Работа выполнена при финансовой поддержке Министерства науки и высшего образования Российской Федерации в ходе реализации проекта «Фундаментальные основы механики, систем контроля и управления

© Булат П.В., Курнухин А.А., Продан Н.В., 2023

беспилотных авиационных систем с формообразующими конструкциями, глубоко интегрированными с силовыми установками, и уникальными свойствами, не применяемыми сегодня в пилотируемой авиации», № FEFM-2020-0001.

Ссылка для цитирования: Булат П.В., Курнухин А.А., Продан Н.В. Численно-аналитическое моделирование пропульсивного крыла и фюзеляжа аэротакси // Научно-технический вестник информационных технологий, механики и оптики. 2023. Т. 23, № 1. С. 178–191. doi: 10.17586/2226-1494-2023-23-1-178-191

Numerical and analytical modeling of the propulsive wing and fuselage of an air taxi

Pavel V. Bulat¹, Anton A. Kurnukhin², Nikolay V. Prodan³

^{1,2,3} Sevastopol State University, Sevastopol, 299053, Russian Federation

^{1,2,3} Baltic State Technical University “VOENMEH” named after D.F. Ustinov, Saint Petersburg, 190005, Russian Federation

¹ pavelbulat@mail.ru, <https://orcid.org/0000-0003-0099-9953>

² anton.kurnukhin@outlook.com, <https://orcid.org/0000-0003-4851-6594>

³ kolinti@mail.ru, <https://orcid.org/0000-0002-6383-039X>

Abstract

The problem of creating propulsive airfoils is considered. Such airfoils have a slit through which the boundary layer is sucked out. Located just behind this gap, a specially profiled section of the airfoil creates a propulsive thrust. The thrust is created due to an abrupt change in the pressure profile on the slit through which the boundary layer is sucked. In the last 15–20 years, the concept of a so-called propulsive wing with reduced or zero aerodynamic drag due to the suction of the boundary layer from its upper surface has been actively studied in the world. Such a wing makes it possible to reduce the aerodynamic drag of the aircraft by several times due to boundary layer laminarization and minimizing the velocity defect associated with viscous friction in the boundary layer, in the wake of the aircraft. The paper proposes a method for numerical modeling of airfoils for a propulsive wing constructed by solving the inverse problem of aerodynamics. The designed airfoils have a maximum construction height, an optimal combination of the lifting force coefficient C_l and the thrust coefficient C_T , created by air suction from the wing surface. The developed technique correctly predicts the point of the laminar-turbulent transition, since the characteristics of the airfoils directly depend on the length of the laminar section. The layout of an aircraft built according to the scheme of a propulsive flying wing of ultra-small aspect ratio using the developed airfoils has been studied. The design of aerodynamic profiles was carried out by solving the inverse problem of aerodynamics with subsequent refinement of geometry using global optimization algorithms. Calculations were carried out using the Langtry–Menter turbulence γ - Re_θ Transition Shear Stress Transport model, in which there are relations for the intermittency criterion, makes it possible to simulate a laminar-turbulent transition. Calculations have shown that the developed airfoils make it possible to create an aircraft airframe with a maximum lift coefficient $C_{l_{max}}$ which exceeds the $C_{l_{max}}$ of a mechanized wing with a flap released during takeoff and landing. In horizontal flight, the C_l is three times larger than that of a typical wing. The wing with the developed profiles has a high propulsive efficiency due to the proximity of pressure and velocity in the thrust section of the airfoils and external flow. At the same time, the thrust surface of the propulsive wing exceeds the nozzle area or the total coverage of aircraft propellers by several times. The developed airfoils and integrated aerodynamic layout of the aircraft are well combined with the principles of building a distributed power plant, and allow you to combine immunity to increased atmospheric turbulence during vertical takeoff and landing with economical horizontal flight. Airfoils have an important advantage over traditional wing mechanization because they have no moving parts, and the increase or decrease in lift is regulated by changing the flow rate of the sucked air.

Keywords

aerodynamic profile, mathematical modeling, optimization, propulsive wing, numerical experiment, power aerodynamic

Acknowledgment

The work was carried out with the financial support of the Ministry of Science and Higher Education of the Russian Federation during the implementation of the project “Fundamentals of mechanics, control and control systems of unmanned aircraft systems with form-forming structures deeply integrated with power plants and unique properties not used today in manned aviation”, No. FEFM-2020-0001.

For citation: Bulat P.V., Kurnukhin A.A., Prodan N.V. Numerical and analytical modeling of the propulsive wing and fuselage of an air taxi. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 2023, vol. 23, no. 1, pp. 178–191 (in Russian). doi: 10.17586/2226-1494-2023-23-1-178-191

Введение

В работе описано тестирование методики численно-го расчета с учетом ламинарно-турбулентного перехода и процесс проектирования пропульсивных профилей и планера летательного аппарата класса аэротакси.

Интерес к пропульсивным профилям возник в авиации 15–20 лет назад, когда NASA запустило проект

Subsonic Fixed Wing, в котором были сформулированы технологические барьеры производства самолетов, подлежащие преодолению в три этапа: $N + 1$, $N + 2$, $N + 3$. В рамках проекта Subsonic Fixed Wing к 2025 году (поколение $N + 3$) должны быть решены следующие задачи: снижение уровня шума на –71 дБ, сокращение выбросов оксидов азота NO_x на 80 % и общее снижение энергопотребления на 60 % [1]. Для



Рис. 1. Проекты NASA «летающее крыло» HBWB (a) и D8 с фюзеляжем в форме восьмерки с утопленными в хвостовую часть вентиляторами для достижения эффекта «утилизации» пограничного слоя (b)

Fig. 1. NASA “flying wing” projects HBWB (a) and D8 with a figure-of-eight fuselage with fans recessed into the tail section to achieve the effect of “utilization” of the boundary layer on (b)

достижения данных задач необходимы значимые успехи в области аэродинамики, двигателей и интеграции планера и силовой установки. Среди прочих проектов особый интерес представляют самолет NASA Hybrid Blended-Wing-Body (HBWB)¹, построенный по схеме «летающее крыло» (рис. 1, a) с распределенной силовой установкой (PCU), и самолет D8 (рис. 1, b) с фюзеляжем в форме восьмерки и утопленными в хвостовую часть вентиляторами. В обоих проектах вентиляторы поглощают и ускоряют рабочее тело за самолетом², что имеет цель — снижение сопротивления.

Эти два проекта продемонстрировали преимущество по сравнению с серией более традиционных по аэродинамике проектов Boeing SUGAR [2]. В проекте HBWB использована PCU с двумя турбогенераторами и множеством электровентиляторов, расположенных вдоль всей верхней задней поверхности несущего фюзеляжа. Применение гибридной турбоэлектрической установки и утопленных вентиляторов, «утилизирующих» пограничный слой в несущей схеме «летающее крыло», активно исследуется в течение последних 10 лет и демонстрирует существенное снижение расхода топлива и шума, по сравнению с самолетами обычной схемы [3–5].

Активные исследования в области интеграции PCU [6] и способов «утилизации» пограничного слоя при большом разнообразии разработанных решений [7] постепенно свелись к концепции пропульсивных, т. е. создающих тягу, аэродинамических профилей.

Концепция пропульсивного крыла берет свое начало в разработке аэродинамических профилей с ламинарным потоком и со ступенчатым повышением давления, которые были впервые предложены Гриффитом в 1940-х годах (рис. 2).

Секция крыла Гриффита спроектирована достаточно широкой и с благоприятным градиентом давления вдоль большей части верхней поверхности (рис. 2). Данное условие привело к сохранению ламинарного характера течения на данном участке и, следовательно, к снижению сопротивления трения обшивки.

Всасывающая щель расположена немного выше по потоку от задней кромки крыла, которая использована для обеспечения разрывного увеличения давления на верхней части аэродинамического профиля. В результате на хвостовой части самолета давление оказывалось больше, чем на передней. Эта разница привела к возникновению статической тяги, которая практически удалила значительную часть сопротивления трения обшивки.

Концепция пропульсивного крыла в дальнейшем была развита Голдшмидом, который использовал аналогичный профиль для дирижабля, но добавил выброс рабочего тела из задней кромки. Эксперименты в аэро-

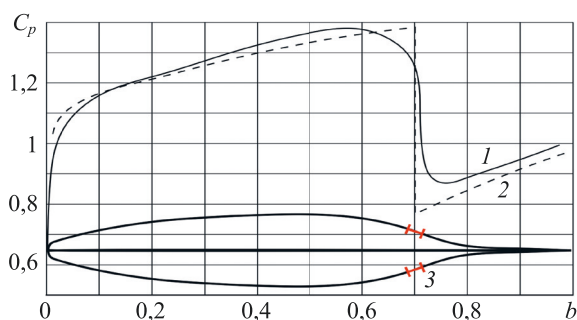


Рис. 2. Характеристики профиля Гриффита [8] при распределении безразмерного коэффициента давления C_p по безразмерной длине хорды b : экспериментальный (кривая 1) и теоретический (кривая 2) графики давления; щель для отбора воздуха с поверхности профиля (кривая 3)

Fig. 2. Characteristics of the Griffith airfoil [8] in the distribution of the dimensionless pressure coefficient C_p along the dimensionless chord length b : experimental (curve 1) and theoretical (curve 2) pressure plots; air sampling slot from the airfoil surface (curve 3)

¹ [Электронный ресурс]. Режим доступа: http://www.nasa.gov/content/hybrid-wing-body-goes-hybrid/#.U96an2M0B_5 (дата обращения: 22.12.2022).

² [Электронный ресурс]. Режим доступа: <http://www.nasa.gov/content/the-double-bubble-d8-0> (дата обращения: 22.12.2022).

динамической трубе [9] и летные эксперименты [10] показали, что избыточное давление на хвостовой части дирижабля полностью компенсирует сопротивление трения, а струя, выбрасываемая через хвостовое сопло, создает дополнительную тягу. В результате была предложена концепция грузового самолета, груз в котором был полностью размещен в очень толстом крыле, составленном из профилей Гриффита/Голдшмида [11].

Перечисленные проекты были ориентированы на сравнительно низкие скорости $u_\infty = 300\text{--}400$ км/ч, при которых явления сжимаемости еще не проявляются. Однако более поздние работы, например [12], показали, что концепция пропульсивного профиля применима вплоть до чисел Маха $M = 0,75$.

Цель работы — разработка метода проектирования и численного моделирования планера летательного аппарата с вертикальным взлетом и посадкой, обладающего высокими аэродинамическими характеристиками, большими внутренними объемами для размещения коммерческого груза и топлива, устойчивого к порывам ветра и возмущениям атмосферы на режимах взлета и посадки. Работа является продолжением [13].

Постановка задачи

Сопротивление летательного аппарата состоит из индуктивного сопротивления, связанного с созданием подъемной силы, и паразитного. Первая составляющая образуется за счет скоса вектора скорости потока \mathbf{u} за крылом (рис. 3, а) относительно скорости невозмущенного потока u_∞ , при этом уменьшается горизонтальная составляющая u_x , что, в соответствии с законом сохранения импульсов, эквивалентно образованию силы сопротивления. Вторая составляющая складывается из вязкого трения о поверхность и уменьшения скорости Δu в следе ω_s , образуемом из оторвавшегося от задней кромки летательного аппарата пограничного слоя

[14]. Коэффициент трения C_f о поверхность зависит от типичных чисел Рейнольдса (табл. 1) и типов пограничных слоев — турбулентного или ламинарного (рис. 2, б). Для беспилотных летательных аппаратов малой размерности эта зависимость особенно выражена.

Концепция «утилизации» пограничного слоя подразумевает увеличение его энергии за счет вентилятора. Подсасывающий эффект препятствует переходу ламинарного течения в турбулентное вверх по течению до вентилятора. За счет подведения мощности к потоку устраняется дефект профиля скорости Δu в следе за летательным аппаратом. В результате снижения сопротивления уменьшается общий потребный запас топлива на борту и размеры самолета существенно уменьшаются (рис. 4).

Проблема, препятствующая применению такого решения на современных самолетах — отсутствие отработанного способа его технической реализации. Использование традиционных осевых вентиляторов не подходит, поскольку вентиляторы чувствительны к неравномерности течения внутри пограничного слоя и плохо работают при высоком противодавлении [15, 16]. Перспективный способ реализации отбора воздуха через щель — поперечный или трансверсальный (Cross Flow Fan, CFF) вентилятор (рис. 5).

Отбор при помощи CFF пограничного слоя обеспечивает ламинарный характер течения воздуха на верхней поверхности, что снижает сопротивление трения. Выброс отобранного с поверхности воздуха через щелевое сопло устраняет дефект профиля скорости в следе за крылом. Это эквивалентно устранению паразитного сопротивления. Повышенное давление на участке профиля от щели отбора до сопла создает положительную тягу. В работах [17, 18] впервые в авиации рассмотрен вентилятор с поперечным потоком.

В работе [19] продемонстрирована возможность использования CFF для создания эффекта суперцирку-

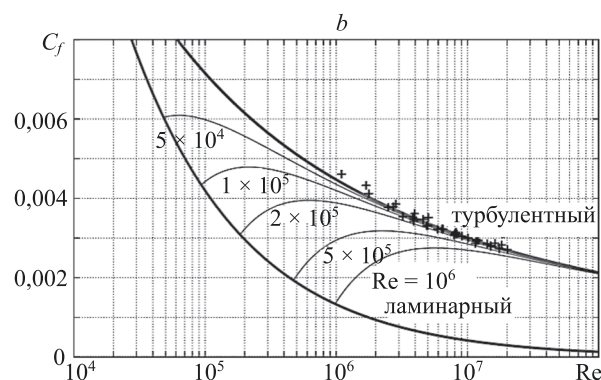
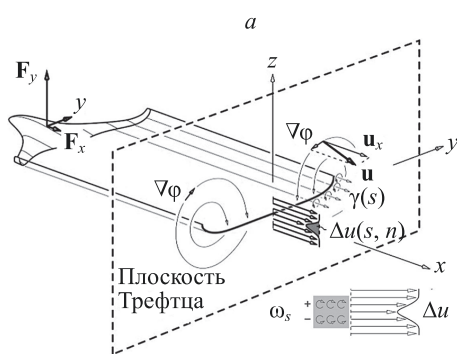


Рис. 3. Механизм образования сопротивления, связанного с подъемной силой, за счет скоса вектора скорости \mathbf{u} за летательным аппаратом, и паразитного сопротивления, возникающего за счет дефекта профиля скорости $\Delta u(s, n)$ в вязком следе ω_s за пограничным слоем (а), а также зависимость величины коэффициента вязкого трения C_f от числа Рейнольдса Re для ламинарного и турбулентного течений (б).

F_x, F_y — составляющие результирующей силы, действующей на летательный аппарат, $\nabla\phi$ — градиент потенциала скорости, $\gamma(s)$ — интенсивность вихря

Fig. 3. The mechanism of formation of resistance associated with the lift force due to the velocity vector \mathbf{u} bevel behind the aircraft, and the parasitic resistance arising due to the defect of the velocity profile Δu in the viscous wake ω_s behind the boundary layer (а), as well as the dependence of the coefficient value viscous friction C_f versus Reynolds number Re for laminar and turbulent flow (б).

F_x, F_y — components of the resultant force acting on the aircraft, $\nabla\phi$ — velocity potential gradient, $\gamma(s)$ — vortex intensity

Таблица 1. Типичные числа Рейнольдса Re для летательных аппаратов различной размерности

Table 1. Typical Reynolds numbers Re for aircraft of various dimensions

Число Re^*	Объект и его размерность
$1 \cdot 10^5 - 3 \cdot 10^5$	Малые беспилотные летательные аппараты и спортивные авиамodelи
$2 \cdot 10^5 - 6 \cdot 10^5$	Малые тактические беспилотные летательные аппараты и большие масштабные модели летательных аппаратов
$6 \cdot 10^5 - 3 \cdot 10^6$	Тактические беспилотные летательные аппараты и пилотируемые планеры
$2 \cdot 10^6 - 8 \cdot 10^6$	Большие беспилотные летательные аппараты и авиация общего назначения
$5 \cdot 10^6 - 2 \cdot 10^6$	Пилотируемые самолеты

Примечание. * вычисляется по длине средней аэродинамической хорды крыла

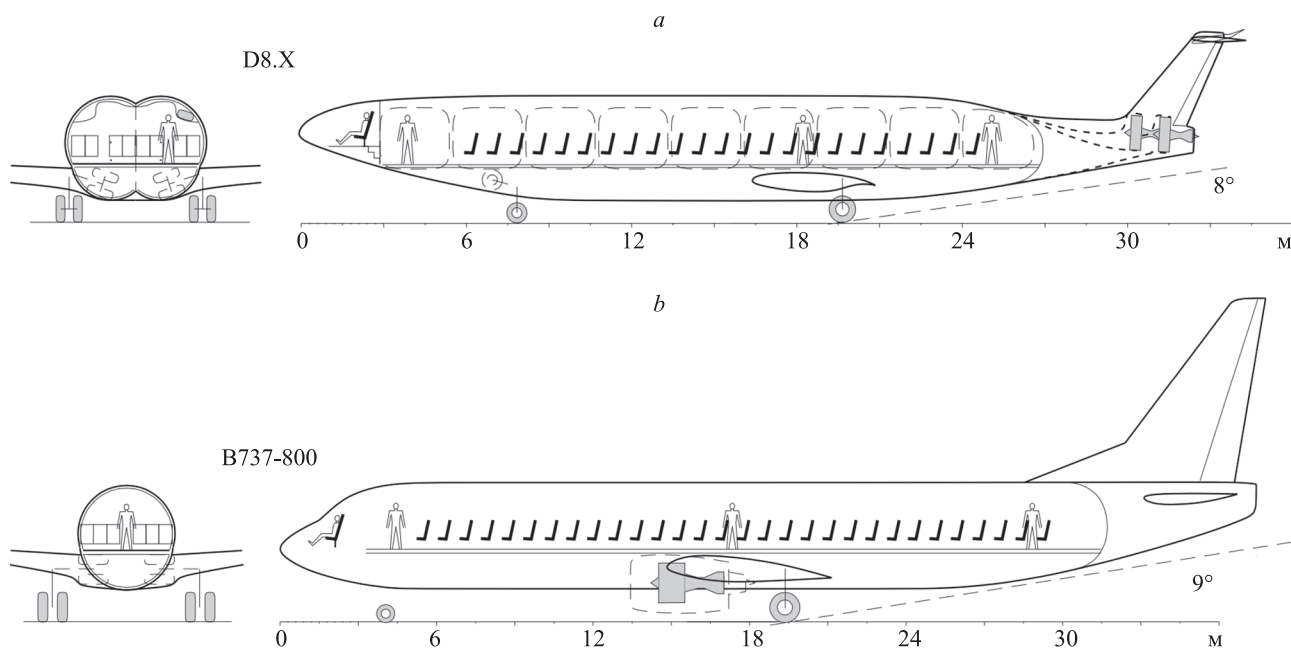


Рис. 4. Сравнение проекта самолета D8.X, оснащенного вентиляторной хвостовой установкой с «утилизацией» пограничного слоя (a), и равного по вместимости обычного самолета Boeing 737-800 (b)

Fig. 4. Comparison of the design of the D8.X aircraft equipped with a fan tail unit with boundary layer “recycling” (a), and equal in capacity to a conventional Boeing 737-800 aircraft (b)

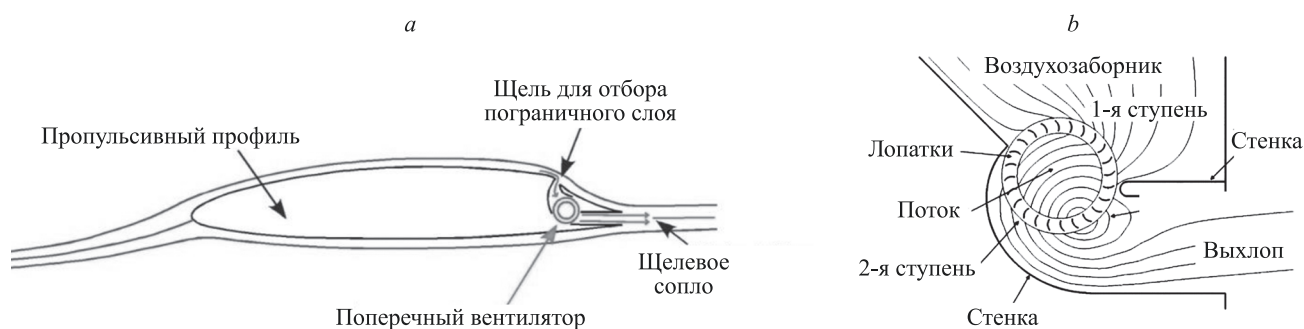


Рис. 5. Концепция пропульсивного профиля с отбором пограничного слоя с верхней поверхности и выбросом его через заднюю кромку (a) при помощи поперечного (трансверсального) вентилятора (b)

Fig. 5. The concept of a propulsion airfoil with the selection of the boundary layer from the upper surface and its ejection through the trailing edge (a) using a transverse (transversal) fan (b)

ляции, в [20] представлена концепция регионального самолета с вертикальным взлетом и посадкой с толстым крылом и CFF. Отметим, что коэффициент полезного действия CFF достигает величины порядка 80 %. Данное значение меньше, чем у осевых вентиляторов,

но оно компенсируется удобством компоновки, меньшей чувствительностью к неоднородности течения в пограничном слое [21, 22] и большому противодавлению [23]. CFF может применяться не только для отсоса пограничного слоя с верхней поверхности крыла, но

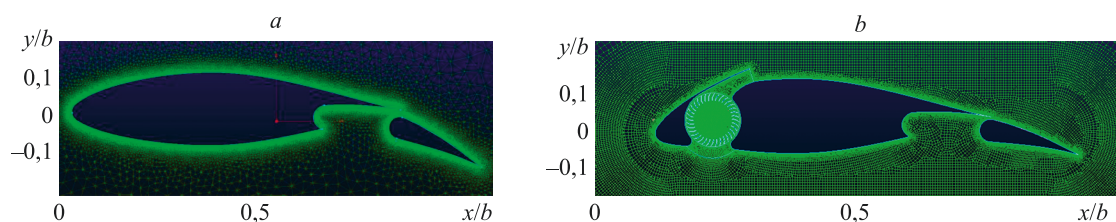


Рис. 6. Модификация профиля NACA 65(3)-221 (а) путем установки CFF для сдува пограничного слоя с верхней части крыла (b).
 x/b — относительная длина профиля, y/b — относительная толщина профиля

Fig. 6. Modification of the NACA 65(3)-221 airfoil (a) by installing a CFF to blow off the boundary layer from the top of the wing (b).
 x/b — relative airfoil length, y/b — relative airfoil thickness

и для сдува (рис. 6, b). В этом случае воздух может забираться из области высокого давления в нижней части профиля, тогда CFF играет роль энергетического предкрылка [24].

Уменьшение размеров и скорости приводит к упрощению задачи обеспечения ламинарного характера течения на большей части поверхности крыла [25] (рис. 7). Иначе говоря, применяя пропульсивные профили, можно обеспечить целиком ламинарное обтекание летательного аппарата и значительно уменьшить паразитное сопротивление.

В настоящей работе исследованы перспективные аэродинамические схемы для нового класса транспортных летательных аппаратов — аэротакси. Особенности аэротакси: малая размерность; дальность не более 200–500 км; комфортное время полета — не более 1 ч. Отсюда следует требование по скорости — не более 400–500 км/ч. Предусмотрен вертикальный взлет и посадка, в том числе в стесненных городских условиях. Аэротакси должно обладать высокой надежностью и устойчивостью к горизонтальным и вертикальным порывам ветра. Устойчивость к порывам ветра обратно

пропорциональна площади крыла и квадрату удлинения крыла. В связи с этим крыло должно быть маленьким и иметь небольшое удлинение.

Модели и методы

Исследуем необычную аэродинамическую компоновку аэротакси (рис. 8). Фюзеляж набран пропульсивными профилями. Воздухозаборник, расположенный «на спине», служит для отбора воздуха с верхней поверхности, ламинаризации потока и питания двигателей. Воздух из воздухозаборника по внутренним каналам поступает в тоннельные вентиляторы и выбрасывается через щелевое сопло в хвостовой части фюзеляжа. Под щелевым соплом расположена отклоняемая вниз «ласта», служащая для управления вектором тяги и выполняющая функции руля высоты. Консоли крыла малого удлинения также сформированы пропульсивными профилями. Они не имеют подвижных частей и щелевого сопла. Отбираемый с поверхности консолей воздух поступает в общий канал тоннельных вентиляторов, расположенных в фюзеляже.

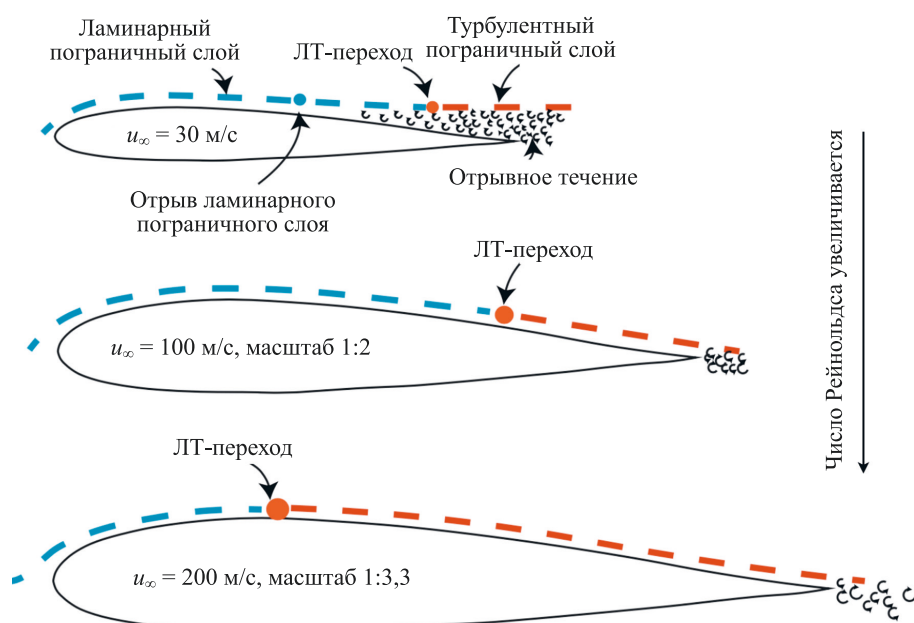


Рис. 7. Характер течения в пограничном слое и положение точки ламинарно-турбулентного (ЛТ) перехода в зависимости от масштаба крыла и числа Рейнольдса

Fig. 7. The nature of the flow in the boundary layer and the position of the laminar-turbulent transition point depending on the scale of the wing and the Reynolds number

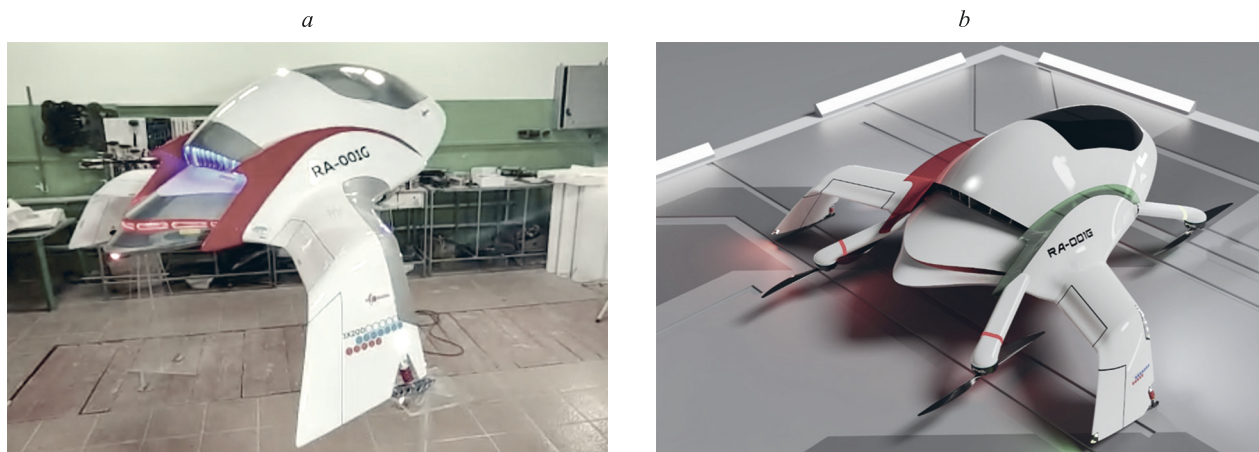


Рис. 8. Макет (a) и компьютерная модель разрабатываемого двухместного аэротакси (b)

Fig. 8. Model (a) and computer model of the developed double air taxi (b)

Пропульсивный профиль (рис. 9, a) фюзеляжа аэротакси был разработан в работе [1] на основе профиля Гриффита/Голдшмида (рис. 9, b) и исследован в [26]. Он отличается от исходного профиля наличием щелевого сопла в задней кромке, модифицированным тяговым участком за щелью отбора воздуха, измененным носиком и более плоской нижней поверхностью. Профиль оптимизирован на максимальные аэродинамическое качество при угле атаки $\alpha = 2^\circ$ протяженность ламинарного участка обтекания при оптимальном расходе отбираемого через щель воздуха.

На рис. 10 представлена поверхность планера, набранная разработанными профилями. Форма фюзеляжа и его мидель получены расчетным путем в результате оптимизации распределения давления в поперечных сечениях. В качестве критерия оптимальности использовано условие максимальной близости формы распределения давления к эллиптической. Таким образом, оптимизация профиля и распределение давления в поперечных сечениях проведены независимо друг от друга по методике, описанной в работе [27].

Крыло на рис. 10 показано условно, так как задача его проектирования в настоящей работе не ставилась. Исследован только профиль крыла.

При разработке профиля крыла за основу выбран оптимизированный профиль Eppler 550, доработка которого выполнена [28] с использованием алгоритма COBYLA (рис. 11, a). Спроектированный профиль (рис. 11, b), является основой для создания пропульсивного крыла, который создан и оптимизирован на максимальное аэродинамическое качество и наибольшую площадь ламинарного обтекания с применением алгоритмов, описанных в [29].

Пропульсивный профиль создан следующим образом. В соответствии с рекомендациями [30] принято, что оптимальное расположение щели для отбора воздуха — на верхней поверхности. Это условие связано с тем, что в данной точке происходит отрыв потока без отбора воздуха, и она является критической.

Оптимальное распределение давления вдоль дужки профиля должно быть примерно постоянным с разрывом на участке, где расположена щель. В каждом

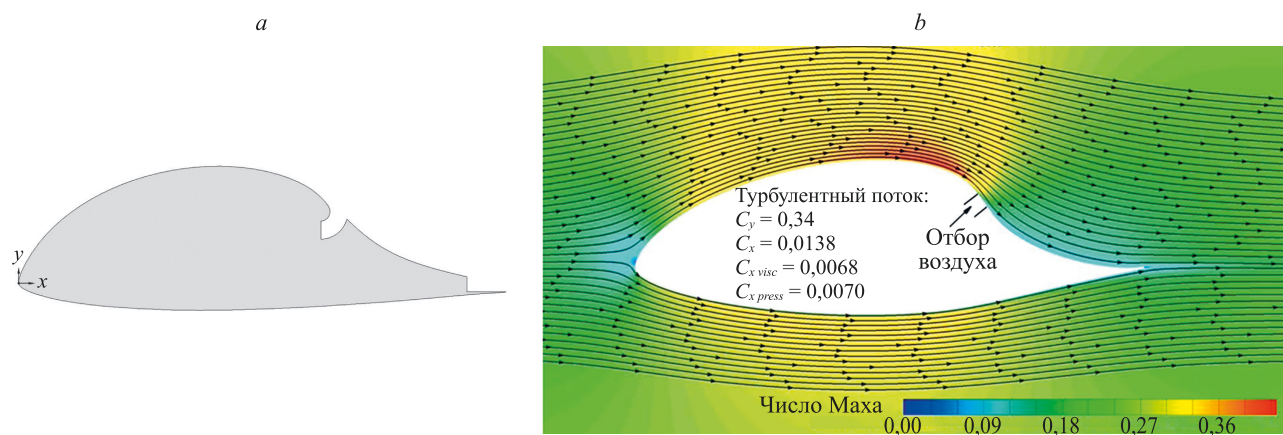


Рис. 9. Пропульсивный профиль для фюзеляжа аэротакси (a) и исходный профиль Гриффита/Голдшмида (b); C_x и C_y — коэффициенты общего сопротивления и подъемной силы; $C_{x\text{ visc}}$ и $C_{x\text{ press}}$ — коэффициенты сопротивления вязкого трения и давления

Fig. 9. Propulsion airfoil for the fuselage of an air taxi (a) and the original Griffith/Goldschmid airfoil (b); C_d and C_l — coefficients of drag and lifting force; $C_{d\text{ visc}}$ and $C_{d\text{ press}}$ — drag coefficients of viscous friction and pressure

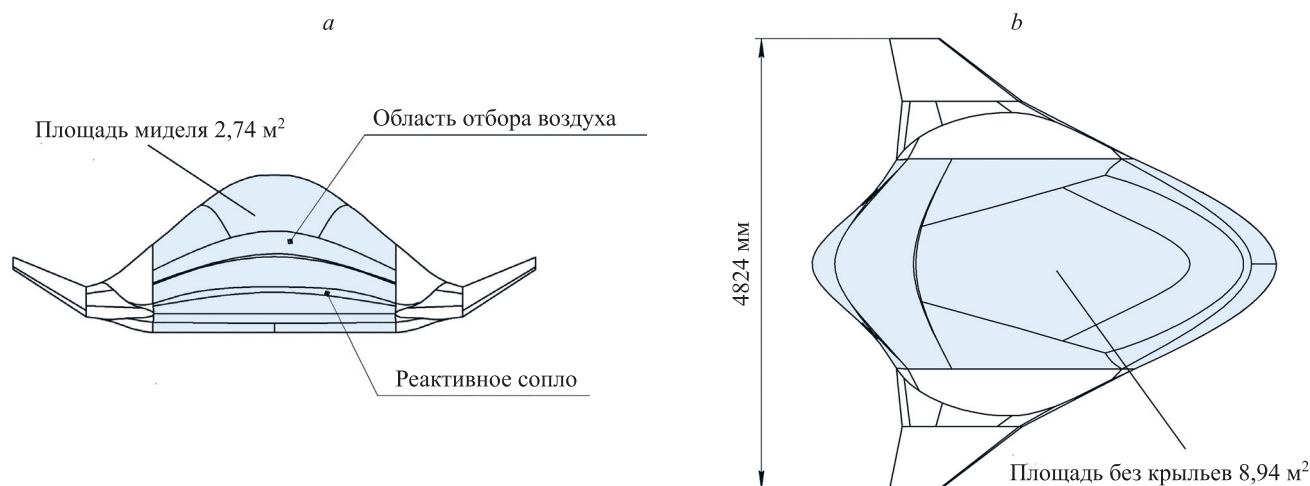


Рис. 10. Форма расчетной области фюзеляжа, набранная пропульсивными профилями: вид сзади (a) и сверху (b)

Fig. 10. Estimated area of the fuselage assembled by propulsion profiles in rear (a) and top (b) views

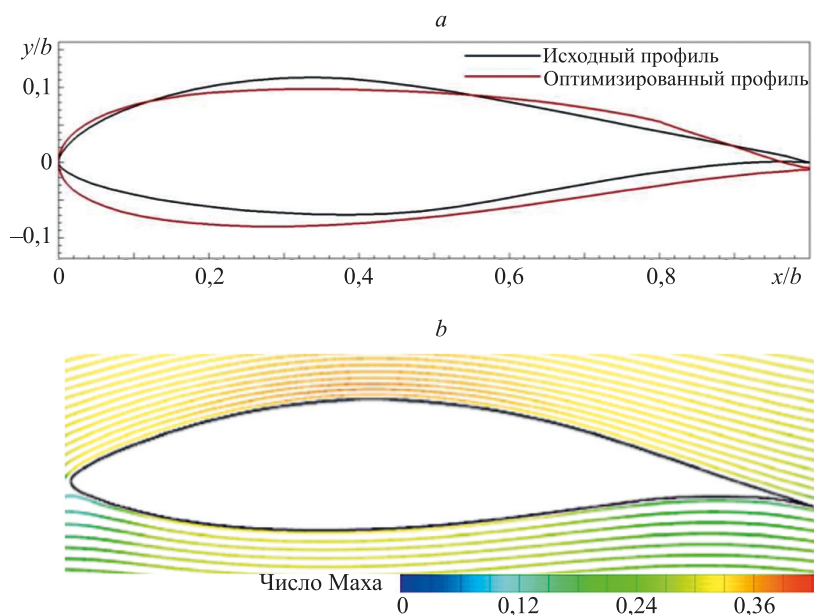


Рис. 11. Оптимизированный профиль Eppler 550 с коэффициентом подъемной силы $C_y = 0,33$ (a) и спроектированный для крыла аэротакси профиль с $C_y = 0,8$ (b)

Fig. 11. Eppler 550 optimized airfoil with lift coefficient $C_l = 0.33$ (a) and airfoil designed for air taxi wing with $C_l = 0.8$ (b)

расчете форма профиля оптимизировалась для обеспечения максимального аэродинамического качества. Для расчетов геометрии профиля использован открытый онлайн-сервис XFOIL [31]. Этот алгоритм отличается от принятого в работе [28], авторы которой постепенно перемещали щель от сечения максимальной высоты к хвостовому сечению профиля и подбирали расход воздуха через щель, обеспечивающий безотрывное течение. Сравнение результатов работы двух алгоритмов приведено ниже.

Для корректного математического моделирования обтекания летательного аппарата в диапазоне чисел Рейнольдса Re от 10^5 до $5 \cdot 10^6$ необходимо учитывать факт, что часть поверхности обтекает ламинарным потоком, а часть турбулентным.

В настоящей работе моделирование ЛТ-перехода выполнено с использованием модели γ - Re_θ Лэнгтри–

Ментера в сочетании с моделью турбулентности Transition Shear Stress Transport (Transition SST) [21, 32]. Модель Лэнгтри–Ментера основана на двух уравнениях переноса (перемежаемости и критерия начала перехода в терминах импульса) и числа Рейнольдса. Перемежаемость турбулентности γ представляет собой долю времени, в течение которого в пограничном слое присутствуют турбулентные пульсации. В ламинарном потоке $\gamma = 0$ внешний поток считается турбулентным с $\gamma = 1$. Показано, что модель хорошо моделирует ЛТ-переход для двумерного случая [33, 34], однако, например для стреловидных крыльев, результаты гораздо скромнее. При тестировании модель турбулентности Transition SST смогла правильно предсказать местоположение точки перехода и наличие отрывных пузырьков в ламинарном пограничном слое. Хорошие результаты показала также модель Спаларта–Аллараса

[35] при условии, что известна точка ЛТ-перехода. Формулировки моделей Transition SST и Спаларта–Аллмараса приведены в работе [36]. Коэффициент вязкости принят равным $\mu = 0,000018375 \text{ кг/(м}\cdot\text{с)}$, размер хорды профиля $L = 1 \text{ м}$. Высоту слоя на поверхности профиля, в котором задано сгущение сетки, определим по формуле

$$\Delta y = Ly^+ \sqrt{74} \text{Re}_L^{-13/14}. \quad (1)$$

В формуле (1) параметр y^+ принят равным $y^+ = 2$ для модели Спаларта–Аллмараса и $y^+ = 0,5$ — для Transition SST. Ранг размеров ячеек в глобальной области 10^{-1} , в пограничном слое — 10^{-2} , на стенках — 10^{-3} . Расчеты выполнены до сеточной сходимости. Основные расчеты проведены для скорости полета $u_\infty = 100 \text{ м/с}$.

Результаты и обсуждение

Выполним сравнение исходного [28] и разработанного профилей. В работе [37] проведена настройка модели турбулентности Transition SST для разрабатываемых профилей в диапазоне углов атаки $\alpha = 0\text{--}12^\circ$ путем сравнения с экспериментальными результатами результатов расчетов профиля № 5 Miley M06-13-128. Осуществим расчеты распределения параметра перемежаемости γ для профилей с рис. 9, а и рис. 11, b, результаты которых представлены на рис. 12.

При расчете профиля для фюзеляжа расход отбираемого воздуха принят равным значению, при котором коэффициент подъемной силы переставал расти. Данные взяты из результатов работы [1]. На рис. 12, а видно, что на всей верхней поверхности профиля имеется тонкий слой, окрашенный в сине-зеленые цвета, что свидетельствует о ламинарном характере течения. На

нижней поверхности точка ЛТ-перехода расположена в зависимости от угла атаки α на расстоянии $0,55\text{--}0,6L$ от носика профиля. Таким образом, спроектированный профиль оказался лучше профиля Гриффита/Голдшмида (рис. 9, b), для которого в работе [28] приведены следующие значения для точек ЛТ-перехода: верхняя — $0,63L$, нижняя — $0,39L$.

Расчеты крылового профиля (рис. 12, b) показали, что он обладает высокими несущими свойствами и низким сопротивлением трения даже без отбора воздуха с поверхности. Так, на верхней дужке точка ЛТ-перехода располагается на расстоянии $0,6L$ от носика профиля. При низких числах Рейнольдса примерно в этой точке происходит отрыв потока от поверхности профиля (рис. 13). При $\text{Re} > 5 \cdot 10^5$ течение носит безотрывный характер.

У спроектированного в работе [28] профиля при отсутствии отсоса воздуха с его поверхности течение сопровождается отрывом потока во всем диапазоне скоростей вплоть до чисел Маха $M = 0,7$, кроме того, профиль не создает подъемной силы (рис. 14, a). При оптимальном отборе воздуха с верхней поверхности (рис. 14, b) течение безотрывное и ламинарное на всей протяженности верхней дужки от носика профиля до щели отбора. Коэффициент подъемной силы достигает величины $C_y = 0,6$, тяговый участок создает тягу C_T , превосходящую силу профильного сопротивления C_{xp}

$$C_{x \text{ press}} = C_{xp} - C_T < 0,$$

но уступающую силе вязкого трения $C_{x \text{ visc}}$. Таким образом, общее сопротивление C_x положительное и не компенсируется тягой C_T . Аналогичные результаты демонстрирует и профиль Гриффита/Голдшмида с рис. 9, b, тяговый участок создает тягу больше профильного со-

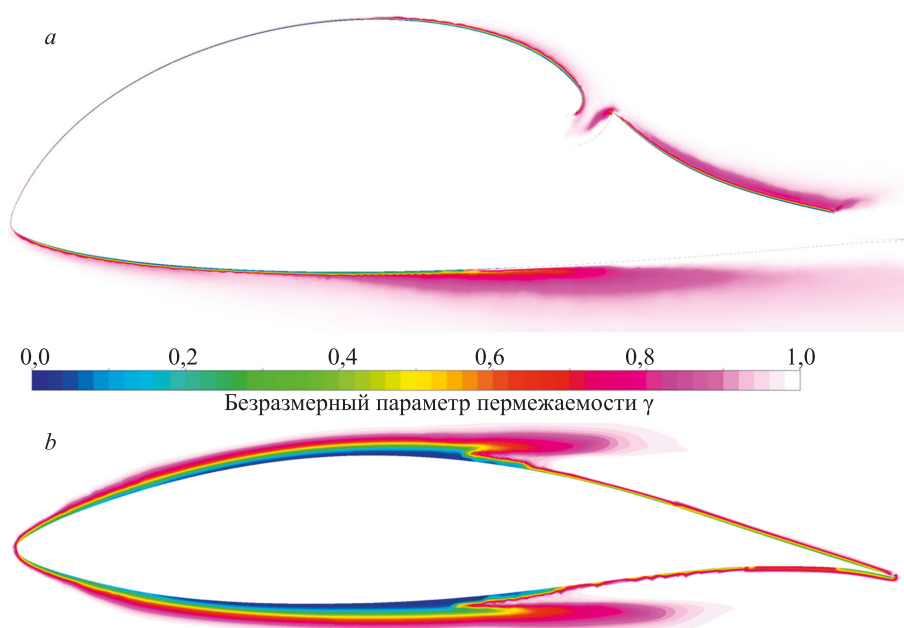


Рис. 12. Результаты расчетов параметра перемежаемости γ в пограничном слое пропульсивного профиля фюзеляжа (a) и исходного профиля, на основании которого спроектирован профиль крыла (b), $\text{Re} = 10^6$

Fig. 12. The results of calculations of the intermittency parameter γ in the boundary layer of the propulsive fuselage airfoil (a) and the initial airfoil, on the basis of which the wing profile was designed (b), $\text{Re} = 10^6$

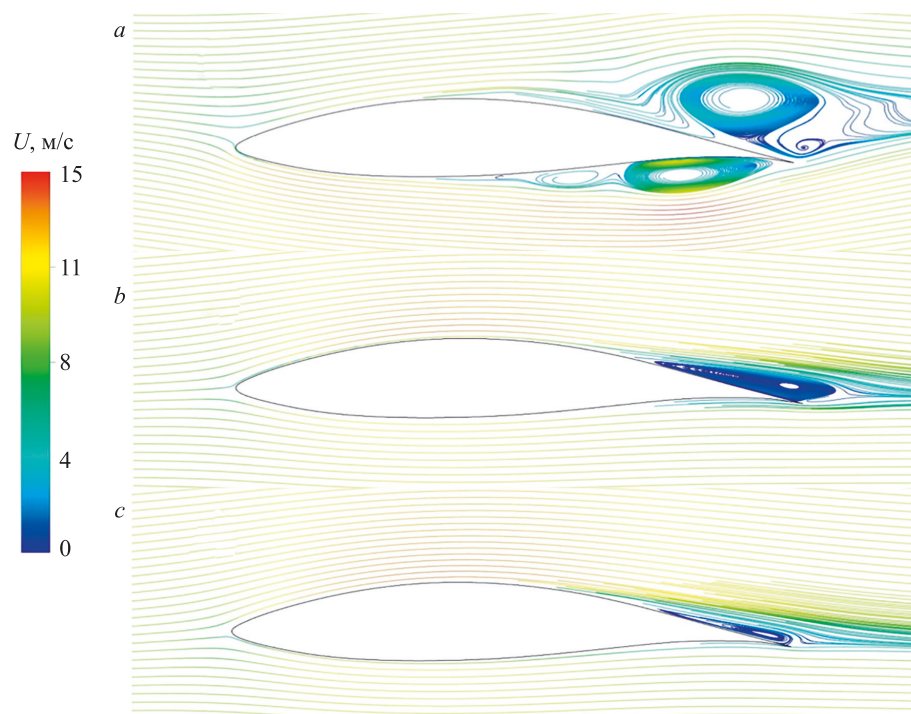


Рис. 13. Результаты расчета отрыва потока при малых числах Рейнольдса $Re = 10^5$: ламинарное течение (a); модели Transition SST (b) и Спаларта–Аллараса (c)

Fig. 13. Flow separation calculation results at low Reynolds numbers $Re = 10^5$: laminar flow (a), Transition SST model (b), Spalart–Allmaras model (c)

противления, но общее сопротивление положительное (рис. 15). Имеется оптимальный расход отбираемого с поверхности профиля воздуха, при котором сопротивление вязкого трения минимальное. Объясняется это тем, что ламинарный участок течения при оптимальной скорости отбора воздуха достигает своей максимальной протяженности, а при дальнейшем росте скорости отбора увеличивается сопротивление в ламинарном пограничном слое, так как растет его скорость.

Из сравнения данных, приведенных на рис. 14, с видно, что разработанный для фюзеляжа профиль на всех рабочих режимах создает тягу, которая превосходит силу вязкого трения, в отличие от профиля Гриффита/Голдшмида. Таким образом, данный профиль с полным основанием можно называть пропульсивным, т. е. создающим тягу. Он по коэффициенту подъемной силы превышает профиль Гриффита/Голдшмида (рис. 9, b). В рабочем диапазоне для режимов набора высоты и горизонтального полета углов атаки $\alpha = 0-8^\circ$ секция фюзеляжа (рис. 10), набранная разработанными профилями, создает подъемную силу $C_y = 0,4-0,55$. Сравнение с профилем Гриффита/

Голдшмида приведено в табл. 2. Если трактовать равнодействующую сил давления на тяговом участке фюзеляжа как часть силы тяги силовой установки, то можно ввести аэродинамическое качество, которое для данного фюзеляжа $K = 9,96$.

Профиль, разработанный для крыла (рис. 11, b), отличается высокими несущими свойствами и низким сопротивлением, но при числах Рейнольдса $Re \approx 10^5$ его обтекание сопровождается отрывом потока, поэтому на его основе был разработан пропульсивный профиль (рис. 15).

В результате профиль получен чуть более толстым, по сравнению с исходным, и с более плоской верхней поверхностью. Аналогичным образом, решая обратную задачу аэродинамики, в работе [28] спроектирован профиль (рис. 15, b), альтернативный профилю, приведенному на рис. 14. Как видно, он более тонкий, так как рассчитан на скорости полета вплоть до $M = 0,7$. Крыловой профиль на рис. 15, a оптимизирован на скорости полета $u_\infty = 50-100$ м/с, поэтому получен более толстым. За счет большей строительной высоты тяговый участок короче, чем у профиля из работы [28].

Таблица 2. Коэффициент подъемной силы C_y при различных углах атаки α

Table 2. Lift coefficient C_l at different angles of attack α

Угол атаки, град.	Фюзеляж	Профиль с рис. 9, a	Профиль с рис. 9, b
0	0,400	1,20	0,534
2	0,485	1,47	0,771
4	0,550	1,75	0,505

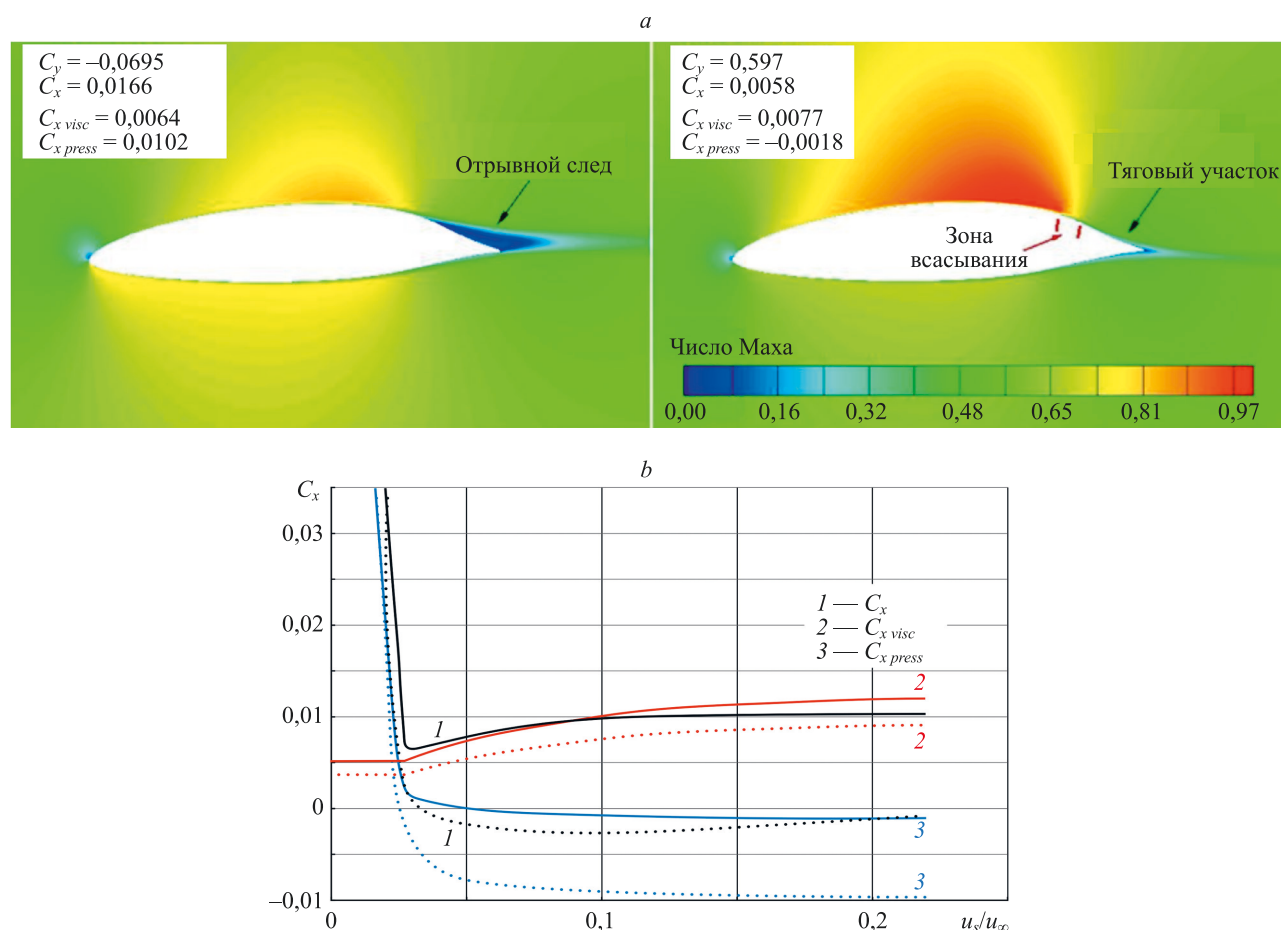


Рис. 14. Результаты расчета профиля Гриффита/Голдшмида [28], $M = 0,7$, $Re > 10^6$, при наличии: отрывного следа (a) и отсоса пограничного слоя, обеспечивающего безотрывное течение в области восстановления давления (тяговый участок) (b); сравнение коэффициентов сопротивления профиля Гриффита/Голдшмида (рис. 9, b) (сплошные линии) с разработанным профилем для фюзеляжа (рис. 9, a) (пунктирные линии) (c).

C_x и C_y — коэффициенты общего сопротивления и подъемной силы; $C_{x\text{ visc}}$ и $C_{x\text{ press}}$ — коэффициенты сопротивления вязкого трения и давления (2)

Fig. 14. Results of calculation of the Griffith/Goldschmid airfoil [28], $M = 0.7$, $Re > 10^6$, in the presence of: a tear-off trace (a) and a boundary layer suction providing a gap-free flow in the pressure recovery region (traction section) (b); comparison of the Griffith/Goldschmid airfoil drag coefficients (Fig. 9, b) (solid lines) with the developed airfoil for the fuselage (Fig. 9, a) (dashed lines) (c).

C_d and C_l — coefficients of total drag and lift; $C_{d\text{ visc}}$ and $C_{d\text{ press}}$ — coefficients of viscous friction resistance and pressure (2)

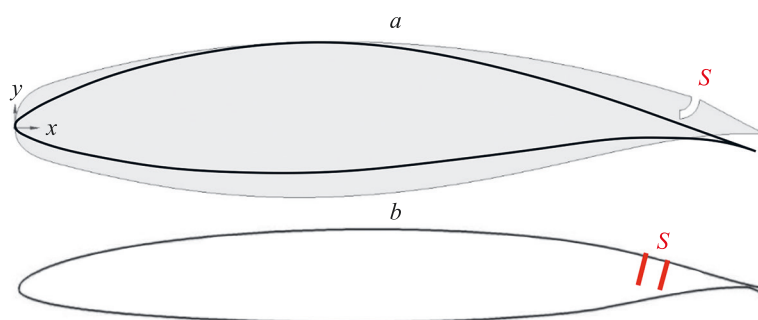


Рис. 15. Спроектированный пропульсивный профиль для крыла (a) в сравнении с исходным профилем (рис. 9, b) и профилем из работы [28].

S — щель для отбора воздуха с поверхности профиля

Fig. 15. The designed propulsion airfoil for the wing (a) in comparison with the original profile (Fig. 9, b) and the airfoil from [28], the letter "S" marks the slot for air extraction from the airfoil surface

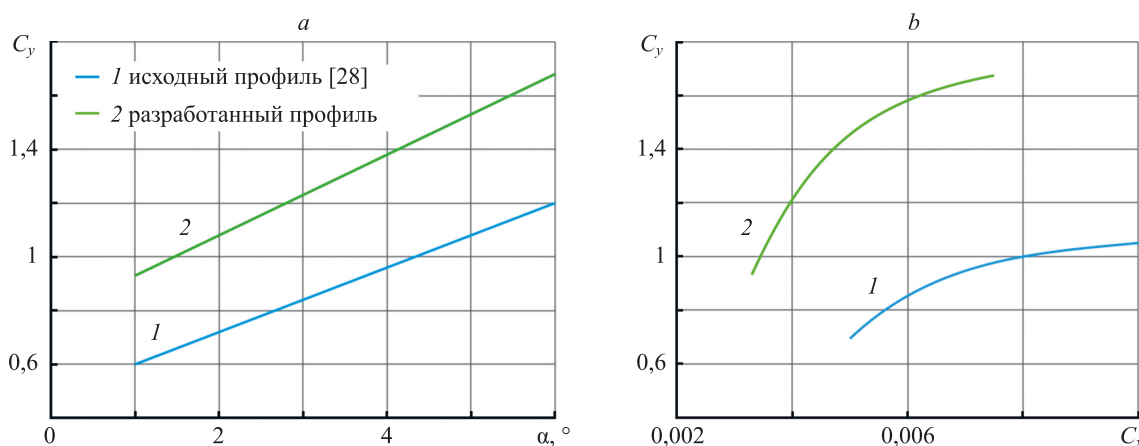


Рис. 16. Сравнение характеристик коэффициента подъемной силы пропульсивных профилей крыла в зависимости от угла атаки (а) и коэффициента сопротивления (C_x) (b)

Fig. 16. Comparison of the characteristics of the lift coefficient of the propulsive wing profiles, depending on the angle of attack (a) and the drag coefficient (C_x) (b)

На рис. 16 приведено сравнение аэродинамических характеристик профилей на рабочих углах атаки $\alpha = 0-6^\circ$. В работе [28] приведены соответствующие данные эксперимента, близкие к полученным расчетным данным.

Заключение

В настоящей работе численными методами исследованы пропульсивные профили и аэродинамическая компоновка летательного аппарата класса «аэротакси». Полученные результаты показали высокое значение аэродинамического качества ($K = 9,96$) для несущего фюзеляжа с большой строительной высотой и внутрен-

ними объемами. Разработанные профили при отборе воздуха с верхней поверхности создают тягу, превышающую лобовое сопротивление во всем диапазоне скоростей. Для крыла созданы два варианта профилей с отбором воздуха и без, имеющие высокие коэффициенты подъемной силы, от 1 до 1,7 в рабочем диапазоне углов атаки, что позволяет в горизонтальном полете ограничиться крылом малой площади и удлинения. Компоновка с несущим фюзеляжем, высота и ширина которого только примерно в 1,5 раза меньше длины, а также крылом малого удлинения обладает высокой устойчивостью к порывам ветра и повышенной турбулентности у земли. Это качество имеет особое значение для обеспечения гарантированного безопасного взлета и посадки.

Литература

1. Felder J.L., Tong M.T., Chu J. Sensitivity of mission energy consumption to turboelectric distributed propulsion design assumptions on the N3-X hybrid wing body aircraft // Proc. of the 48th AIAA/ASME/SAE/ASEE Joint Propulsion Conference and Exhibit. 2012. <https://doi.org/10.2514/6.2012-3701>
2. Bradley M., Droney C. Subsonic Ultra Green Aircraft Research: Phase I Final Report. NASA/CR2011-216847. 2011. 193 p.
3. Felder J.L., Kim H.D., Brown G. An examination of the effect of boundary layer ingestion on turboelectric distributed propulsion systems // Proc. of the 49th AIAA Aerospace Sciences Meeting including the New Horizons Forum and Aerospace Exposition. 2011. <https://doi.org/10.2514/6.2011-300>
4. Alrashed M., Nicoladis T., Pilidis P., Jafari S. Utilisation of turboelectric distribution propulsion in commercial aviation: A review on NASA's TeDP concept // Chinese Journal of Aeronautics. 2021. V. 34. N 11. P. 48–65. <https://doi.org/10.1016/j.cja.2021.03.014>
5. Chen Z., Zhang M., Chen Y., Sang W., Tan Z., Li D., Zhang B. Assessment on critical technologies for conceptual design of blended-wing-body civil aircraft // Chinese Journal of Aeronautics. 2019. V. 32. N 8. P. 1797–1827. <https://doi.org/10.1016/j.cja.2019.06.006>
6. Yaros S.F., Sextstone M.G., Huebner L.D., Lamar J.E., McKinley R.E., Jr., Torres A.O., Burley C.L., Scott R.C., Small W.J. Synergistic Airframe-Propulsion Interactions and Integrations. A White Paper Prepared by the 1996-1997 Langley Aeronautics Technical Committee: NASA/TM-1998-207644, 1998. 122 p.

References

1. Felder J.L., Tong M.T., Chu J. sensitivity of mission energy consumption to turboelectric distributed propulsion design assumptions on the N3-X hybrid wing body aircraft. *Proc. of the 48th AIAA/ASME/SAE/ASEE Joint Propulsion Conference and Exhibit*, 2012. <https://doi.org/10.2514/6.2012-3701>
2. Bradley M., Droney C. *Subsonic Ultra Green Aircraft Research. Phase I Final Report*. NASA/CR2011-216847. 2011, 193 p.
3. Felder J.L., Kim H.D., Brown G. An examination of the effect of boundary layer ingestion on turboelectric distributed propulsion systems. *Proc. of the 49th AIAA Aerospace Sciences Meeting including the New Horizons Forum and Aerospace Exposition*, 2011. <https://doi.org/10.2514/6.2011-300>
4. Alrashed M., Nicoladis T., Pilidis P., Jafari S. Utilisation of turboelectric distribution propulsion in commercial aviation: A review on NASA's TeDP concept. *Chinese Journal of Aeronautics*, 2021, vol. 34, no. 11, pp. 48–65. <https://doi.org/10.1016/j.cja.2021.03.014>
5. Chen Z., Zhang M., Chen Y., Sang W., Tan Z., Li D., Zhang B. Assessment on critical technologies for conceptual design of blended-wing-body civil aircraft. *Chinese Journal of Aeronautics*, 2019, vol. 32, no. 8, pp. 1797–1827. <https://doi.org/10.1016/j.cja.2019.06.006>
6. Yaros S.F., Sextstone M.G., Huebner L.D., Lamar J.E., McKinley R.E., Jr., Torres A.O., Burley C.L., Scott R.C., Small W.J. *Synergistic Airframe-Propulsion Interactions and Integrations. A White Paper Prepared by the 1996-1997 Langley Aeronautics Technical Committee*. NASA/TM-1998-207644, 1998, 122 p.

7. Burston M., Ranasinghe K., Gardi A., Parezanovic V., Ajaj R., Sabatini R. Design principles and digital control of advanced distributed propulsion systems // *Energy*, 2022. V. 241. P. 122788. <https://doi.org/10.1016/j.energy.2021.122788>
8. Richards E.J., Walker W.S., Greening J.R. Tests of a Griffith Aerofoil in the 13 ft x 9 ft Wind Tunnel. Part 1, Part 2, Part 3, Part 4: ARC R&M-2148 ARC-7464 ARC-7561 ARC-8054 ARC-8055. 1944.
9. Goldschmied F. Fuselage self-propulsion by static-pressure thrust: Wind-tunnel verification // *Proc. of the Aircraft Design, Systems and Operations Meeting*, 1987. <https://doi.org/10.2514/6.1987-2935>
10. Goldschmied F. Airfoil static-pressure thrust: flight test verification // *Proc. of the Aircraft Design, Systems and Operations Conference*, 1990. <https://doi.org/10.2514/6.1990-3286>
11. Goldschmied F. Thick-wing spanloader all-freighter: design concept for tomorrow's air cargo // *Proc. of the Aircraft Design, Systems and Operations Conference*, 1990. <https://doi.org/10.2514/6.1990-3198>
12. Cella U., Quagliarella D., Donelli R., Imperatore B. Design and test of the UW-5006 transonic natural-laminar-flow wing // *Journal of Aircraft*, 2010. V. 47. N 3. P. 783–795. <https://doi.org/10.2514/1.40932>
13. Булат П.В., Курнухин А.А., Продан Н.В. Численное моделирование характеристик высоконесущих профилей с энергетическими методами увеличения подъемной силы // *Научно-технический вестник информационных технологий, механики и оптики*, 2022. Т. 22. № 5. С. 1007–1015. <https://doi.org/10.17586/2226-1494-2022-22-5-1007-1015>
14. Булат П.В., Дудников С.Ю., Кузнецов П.Н. Основы аэродинамики беспилотных воздушных судов: учебное пособие. М.: Издательство «Спутник +», 2021. 273 с.
15. Tamaki T., Nagano S. Effects of inlet distortions on a multi-stage compressor // *Proc. of the 4th International Symposium on Air Breathing Engines*, 1979. <https://doi.org/10.2514/6.1979-7003>
16. Sandercock D.M., Sanger N.L. Some observations of the effects of radial distortions on performance of a transonic rotating blade row: NASA TN D-7824, 1974, 48 p.
17. Hancock J.P. Test of a high efficiency transverse fan // *Proc. of the 16th Joint Propulsion Conference*, 1980. <https://doi.org/10.2514/6.1980-1243>
18. Harloff G.J., Wilson D.R. Cross-flow propulsion fan experimental development and finite-element modeling // *Journal of Aircraft*, 1981. V. 18. N 4. P. 310–317. <https://doi.org/10.2514/3.57494>
19. Dygert R.K., Dang T.Q. Experimental investigation of embedded cross-flow fan for airfoil propulsion/circulation control // *Proc. of the 45th AIAA Aerospace Sciences Meeting and Exhibit*, 2007. <https://doi.org/10.2514/6.2007-368>
20. Gologan C., Mores S., Steiner H., Seitz A. Potential of the cross-flow fan for powered-lift regional aircraft applications // *Proc. of the 9th AIAA Aviation Technology, Integration, and Operations Conference (ATIO)*, 2009. <https://doi.org/10.2514/6.2009-7098>
21. Langel C.M., Chow R., Van Dam C.P., Maniaci D.C., Erhmann R.S., White E.B. A computational approach to simulating the effects of realistic surface roughness on boundary layer transition // *Proc. of the 52nd Aerospace Sciences Meeting*, 2014. <https://doi.org/10.2514/6.2014-0234>
22. Doll U., Migliorini M., Baikie J., Zachos K.P., Röhle I., Melnikov S., Steinbock J., Dues M., Kapulla R., MacManus D.G., Lawson N.J. Non-intrusive flow diagnostics for unsteady inlet flow distortion measurements in novel aircraft architectures // *Progress in Aerospace Sciences*, 2022. V. 130. P. 100810. <https://doi.org/10.1016/j.paerosci.2022.100810>
23. Dang T.Q., Bushnell P.P. Aerodynamics of cross-flow fans and their application to aircraft propulsion and flow control // *Progress in Aerospace Sciences*, 2009. V. 45. N 1-3. P. 1–29. <https://doi.org/10.1016/j.paerosci.2008.10.002>
24. Karpuk S.V., Kazarin P., Gudmundsson S., Golubev V.V. Preliminary feasibility study of a multi-purpose aircraft concept with a leading-edge embedded cross-flow fan // *Proc. of the 2018 AIAA Aerospace Sciences Meeting*, 2018. <https://doi.org/10.2514/6.2018-1744>
25. Kulkarni A.R., La Rocca G., Veldhuis L.L.M., Eitelberg G. Sub-scale flight test model design: Developments, challenges and opportunities // *Progress in Aerospace Sciences*, 2022. V. 130. P. 100798. <https://doi.org/10.1016/j.paerosci.2021.100798>
26. Perry A.T., Ansell P.J., Kerho M., Ananda G., D'Urso S. Design, analysis, and evaluation of a propulsive wing concept // *Proc. of the 34th AIAA Applied Aerodynamics Conference*, 2016. <https://doi.org/10.2514/6.2016-4178>
27. Prodan N.V., Kurnukhin A.A. Application of mathematical optimization methods for designing airfoil considering viscosity //
7. Burston M., Ranasinghe K., Gardi A., Parezanovic V., Ajaj R., Sabatini R. Design principles and digital control of advanced distributed propulsion systems. *Energy*, 2022, vol. 241, pp. 122788. <https://doi.org/10.1016/j.energy.2021.122788>
8. Richards E.J., Walker W.S., Greening J.R. *Tests of a Griffith Aerofoil in the 13 ft x 9 ft Wind Tunnel*. Part 1, Part 2, Part 3, Part 4. ARC R&M-2148 ARC-7464 ARC-7561 ARC-8054 ARC-8055. 1944.
9. Goldschmied F. Fuselage self-propulsion by static-pressure thrust: Wind-tunnel verification. *Proc. of the Aircraft Design, Systems and Operations Meeting*, 1987. <https://doi.org/10.2514/6.1987-2935>
10. Goldschmied F. Airfoil static-pressure thrust: flight test verification. *Proc. of the Aircraft Design, Systems and Operations Conference*, 1990. <https://doi.org/10.2514/6.1990-3286>
11. Goldschmied F. Thick-wing spanloader all-freighter: design concept for tomorrow's air cargo. *Proc. of the Aircraft Design, Systems and Operations Conference*, 1990. <https://doi.org/10.2514/6.1990-3198>
12. Cella U., Quagliarella D., Donelli R., Imperatore B. Design and test of the UW-5006 transonic natural-laminar-flow wing. *Journal of Aircraft*, 2010, vol. 47, no. 3, pp. 783–795. <https://doi.org/10.2514/1.40932>
13. Bulat P.V., Kurnukhin A.A., Prodan N.V. Numerical simulation of propulsive aerodynamic profiles. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 2022, vol. 22, no. 5, pp. 1007–1015. (in Russian). <https://doi.org/10.17586/2226-1494-2022-22-5-1007-1015>
14. Bulat P.V., Dudnikov S.Iu., Kuznetsov P.N. *Fundamental Aerodynamics of the Unmanned Aerial Vehicles*. Tutorial. Moscow, "Sputnik+" Publ., 2021, 273 p. (in Russian)
15. Tamaki T., Nagano S. Effects of inlet distortions on a multi-stage compressor. *Proc. of the 4th International Symposium on Air Breathing Engines*, 1979. <https://doi.org/10.2514/6.1979-7003>
16. Sandercock D.M., Sanger N.L. *Some observations of the effects of radial distortions on performance of a transonic rotating blade row*. NASA TN D-7824, 1974, 48 p.
17. Hancock J.P. Test of a high efficiency transverse fan. *Proc. of the 16th Joint Propulsion Conference*, 1980. <https://doi.org/10.2514/6.1980-1243>
18. Harloff G.J., Wilson D.R. Cross-flow propulsion fan experimental development and finite-element modeling. *Journal of Aircraft*, 1981, vol. 18, no. 4, pp. 310–317. <https://doi.org/10.2514/3.57494>
19. Dygert R.K., Dang T.Q. Experimental investigation of embedded cross-flow fan for airfoil propulsion/circulation control. *Proc. of the 45th AIAA Aerospace Sciences Meeting and Exhibit*, 2007. <https://doi.org/10.2514/6.2007-368>
20. Gologan C., Mores S., Steiner H., Seitz A. Potential of the cross-flow fan for powered-lift regional aircraft applications. *Proc. of the 9th AIAA Aviation Technology, Integration, and Operations Conference (ATIO)*, 2009. <https://doi.org/10.2514/6.2009-7098>
21. Langel C.M., Chow R., Van Dam C.P., Maniaci D.C., Erhmann R.S., White E.B. A computational approach to simulating the effects of realistic surface roughness on boundary layer transition. *Proc. of the 52nd Aerospace Sciences Meeting*, 2014. <https://doi.org/10.2514/6.2014-0234>
22. Doll U., Migliorini M., Baikie J., Zachos K.P., Röhle I., Melnikov S., Steinbock J., Dues M., Kapulla R., MacManus D.G., Lawson N.J. Non-intrusive flow diagnostics for unsteady inlet flow distortion measurements in novel aircraft architectures. *Progress in Aerospace Sciences*, 2022, vol. 130, pp. 100810. <https://doi.org/10.1016/j.paerosci.2022.100810>
23. Dang T.Q., Bushnell P.P. Aerodynamics of cross-flow fans and their application to aircraft propulsion and flow control. *Progress in Aerospace Sciences*, 2009, vol. 45, no. 1-3, pp. 1–29. <https://doi.org/10.1016/j.paerosci.2008.10.002>
24. Karpuk S.V., Kazarin P., Gudmundsson S., Golubev V.V. Preliminary feasibility study of a multi-purpose aircraft concept with a leading-edge embedded cross-flow fan. *Proc. of the 2018 AIAA Aerospace Sciences Meeting*, 2018. <https://doi.org/10.2514/6.2018-1744>
25. Kulkarni A.R., La Rocca G., Veldhuis L.L.M., Eitelberg G. Sub-scale flight test model design: Developments, challenges and opportunities. *Progress in Aerospace Sciences*, 2022, vol. 130, pp. 100798. <https://doi.org/10.1016/j.paerosci.2021.100798>
26. Perry A.T., Ansell P.J., Kerho M., Ananda G., D'Urso S. Design, analysis, and evaluation of a propulsive wing concept. *Proc. of the 34th AIAA Applied Aerodynamics Conference*, 2016. <https://doi.org/10.2514/6.2016-4178>
27. Prodan N.V., Kurnukhin A.A. Application of mathematical optimization methods for designing airfoil considering viscosity.

- Russian Aeronautics. 2021. V. 64. N 4. P. 670–677. <https://doi.org/10.3103/S1068799821040115>
28. Kramer B., Ansell Ph., D'Urso S., Ananda G., Perry A. Design, Analysis, and Evaluation of a Novel Propulsive Wing Concept: LEARN Phase I. Final Report. Contract Number NNX15AE39A. June 30th, 2016. 80 p.
 29. Булат П.В., Продан Н.В., Курнухин А.А. О влиянии ламинарно-турбулентного перехода при численном моделировании профиля крыла // Известия вузов. Авиационная техника. 2021. № 3. С. 89–98.
 30. Абзалилов Д.Ф., Марданов Р.Ф. Расчет и оптимизация аэродинамических характеристик крылового профиля с выдувом реактивной струи при наличии в потоке вихря // Известия вузов. Авиационная техника. 2016. № 3. С. 58–63.
 31. Drela M. XFOIL: An analysis and design system for low Reynolds number airfoils // Lecture Notes in Engineering. 1989. V. 54. P. 1–12. https://doi.org/10.1007/978-3-642-84010-4_1
 32. Langtry R.B., Menter F.R. Transition modeling for general CFD applications in aeronautics // Proc. of the 43rd AIAA Aerospace Sciences Meeting and Exhibit. 2005. <https://doi.org/10.2514/6.2005-522>
 33. Malan P., Suluksna K., Juntasaro E. Calibrating the γ -Re θ transition model for commercial CFD // Proc. of the 47th AIAA Aerospace Sciences Meeting including The New Horizons Forum and Aerospace Exposition. 2009. <https://doi.org/10.2514/6.2009-1142>
 34. Liu K., Wang Y., Song W.-P., Han Z.-H. A two-equation local-correlation-based laminar-turbulent transition modeling scheme for external aerodynamics // Aerospace Science and Technology. 2020. V. 106. P. 106128. <https://doi.org/10.1016/j.ast.2020.106128>
 35. Соломатин Р.С., Семенов И.В., Меньшов И.С. К расчету турбулентных течений на основе модели Спаларта-Аллараса с применением LU-SGS-GMRES алгоритма // Препринты ИПМ им. М.В.Келдыша. 2018. № 119. С. 1–30.
 36. Дудников С.Ю., Кузнецов П.Н., Мельникова А.И., Вокин Л.О. Моделирование течений при малых числах Рейнольдса применительно к проектированию несущих аэродинамических поверхностей беспилотных летательных аппаратов // Известия вузов. Авиационная техника. 2021. № 4. С. 39–48.
 37. Булат П.В., Продан Н.В., Дудников С.Ю., Курнухин А.А. Исследование характеристик аэродинамических профилей с отбором воздуха с верхней поверхности и заданным распределением давления // Известия вузов. Авиационная техника. 2022. № 3. Russian Aeronautics, 2021, vol. 64, no. 4, pp. 670–677. <https://doi.org/10.3103/S1068799821040115>
 28. Kramer B., Ansell Ph., D'Urso S., Ananda G., Perry A. Design, Analysis, and Evaluation of a Novel Propulsive Wing Concept. LEARN Phase I. Final Report. Contract Number NNX15AE39A. June 30th, 2016, 80 p.
 29. Bulat P.V., Prodan N.V., Kurnukhin A.A. On the influence of the laminar-turbulent transition in the numerical modeling of the wing airfoil. Russian Aeronautics, 2021, vol. 64, no. 3, pp. 455–465. <https://doi.org/10.3103/S1068799821030120>
 30. Abzalilov D.F., Mardanov R.F. Calculation and optimization of aerodynamic characteristics of airfoils with jet blowing in the presence of vortex in the flow. Russian Aeronautics, 2016, vol. 59, no. 3, pp. 358–363. <https://doi.org/10.3103/S1068799816030107>
 31. Drela M. XFOIL: An analysis and design system for low Reynolds number airfoils. Lecture Notes in Engineering, 1989, vol. 54, pp. 1–12. https://doi.org/10.1007/978-3-642-84010-4_1
 32. Langtry R.B., Menter F.R. Transition modeling for general CFD applications in aeronautics. Proc. of the 43rd AIAA Aerospace Sciences Meeting and Exhibit, 2005. <https://doi.org/10.2514/6.2005-522>
 33. Malan P., Suluksna K., Juntasaro E. Calibrating the γ -Re θ transition model for commercial CFD. Proc. of the 47th AIAA Aerospace Sciences Meeting including The New Horizons Forum and Aerospace Exposition, 2009. <https://doi.org/10.2514/6.2009-1142>
 34. Liu K., Wang Y., Song W.-P., Han Z.-H. A two-equation local-correlation-based laminar-turbulent transition modeling scheme for external aerodynamics. Aerospace Science and Technology, 2020, vol. 106, pp. 106128. <https://doi.org/10.1016/j.ast.2020.106128>
 35. Solomatin R.S., Semenov I.V., Men'shov I.S. Towards calculating turbulent flows with the Spalart-Allmaras model by using the LU-SGS-GMRES algorithm. Keldysh Institute Preprints, 2018, no. 119, pp. 1–30. (in Russian)
 36. Dudnikov S.Y., Kuznetsov P.N., Mel'nikova A.I., Vokin L.O. Simulation of flows at low Reynolds numbers as applied to the design of aerodynamic surfaces for unmanned aerial vehicles. Russian Aeronautics, 2021, vol. 64, no. 4, pp. 620–629. <https://doi.org/10.3103/S1068799821040061>
 37. Bulat P.V., Prodan N.V., Dudnikov S.YU., Kurnukhin A.A. Investigation of the characteristics of airfoils with air suction from the upper surface and a given pressure distribution. Izvestiya vysshikh uchebnykh zavedenii, 2022, no. 3. (in Russian)

Авторы

Булат Павел Викторович — доктор физико-математических наук, кандидат экономических наук, главный научный сотрудник, Севастопольский государственный университет, Севастополь, 299053, Российская Федерация; главный научный сотрудник, Балтийский государственный технический университет «ВОЕНМЕХ» им. Д.Ф. Устинова, Санкт-Петербург, 190005, Российская Федерация, [sc 55969578400](https://orcid.org/0000-0003-0099-9953), <https://orcid.org/0000-0003-0099-9953>, pavelbulat@mail.ru

Курнухин Антон Александрович — младший научный сотрудник, Севастопольский государственный университет, Севастополь, 299053, Российская Федерация; младший научный сотрудник, Балтийский государственный технический университет «ВОЕНМЕХ» им. Д.Ф. Устинова, Санкт-Петербург, 190005, Российская Федерация, [sc 57223276960](https://orcid.org/0000-0003-4851-6594), <https://orcid.org/0000-0003-4851-6594>, anton.kurnukhin@outlook.com

Продан Николай Васильевич — кандидат технических наук, старший научный сотрудник, Севастопольский государственный университет, Севастополь, 299053, Российская Федерация; старший научный сотрудник, Балтийский государственный технический университет «ВОЕНМЕХ» им. Д.Ф. Устинова, Санкт-Петербург, 190005, Российская Федерация, [sc 56177927100](https://orcid.org/0000-0002-6383-039X), <https://orcid.org/0000-0002-6383-039X>, kolinti@mail.ru

Authors

Pavel V. Bulat — D.Sc. (Physics & Mathematics), PhD (Economics), Chief Researcher, Sevastopol State University, Sevastopol, 299053, Russian Federation; Chief Researcher, Baltic State Technical University “VOENMEH” named after D.F. Ustinov, Saint Petersburg, 190005, Russian Federation, [sc 55969578400](https://orcid.org/0000-0003-0099-9953), <https://orcid.org/0000-0003-0099-9953>, pavelbulat@mail.ru

Anton A. Kurnukhin — Junior Researcher, Sevastopol State University, Sevastopol, 299053, Russian Federation; Junior Researcher, Baltic State Technical University “VOENMEH” named after D.F. Ustinov, Saint Petersburg, 190005, Russian Federation, [sc 57223276960](https://orcid.org/0000-0003-4851-6594), <https://orcid.org/0000-0003-4851-6594>, anton.kurnukhin@outlook.com

Nikolay V. Prodan — PhD, Senior Researcher, Sevastopol State University, Sevastopol, 299053, Russian Federation; Senior Researcher, Baltic State Technical University “VOENMEH” named after D.F. Ustinov, Saint Petersburg, 190005, Russian Federation, [sc 56177927100](https://orcid.org/0000-0002-6383-039X), <https://orcid.org/0000-0002-6383-039X>, kolinti@mail.ru

Статья поступила в редакцию 17.09.2022
Одобрена после рецензирования 05.12.2022
Принята к печати 18.01.2023

Received 17.09.2022
Approved after reviewing 05.12.2022
Accepted 18.01.2023



Работа доступна по лицензии
Creative Commons
«Attribution-NonCommercial»

doi: 10.17586/2226-1494-2023-23-1-192-201

УДК 62-83

Применение обратной связи переменной точности для повышения быстродействия контура тока в инверторах на базе GaN-транзисторов

Алексей Сергеевич Анучин¹, Мария Андреевна Гуляева²,
Максим Михайлович Лашкевич³, Александр Александрович Жарков⁴, Хао Чен⁵,
Антон Николаевич Дианов⁶

^{1,2,4} Национальный исследовательский университет «МЭИ», Москва, 111250, Российская Федерация

^{3,4} ООО «НПФ ВЕКТОР», Москва, 111396, Российская Федерация

⁵ China University of Mining & Technology, Сюйчжоу, 100083, Китай

⁶ Daeyoung R&D Center, Сеул, 03140, Южная Корея

¹ Anuchin.alecksey@gmail.com, <https://orcid.org/0000-0002-1019-756X>

² tilgivelse@yandex.ru, <https://orcid.org/0000-0002-2326-1153>

³ maxsoftpage@yandex.ru, <https://orcid.org/0000-0002-9699-7145>

⁴ zarckov@mail.ru, <https://orcid.org/0000-0001-5628-8101>

⁵ hchen@cumt.edu.cn, <https://orcid.org/0000-0001-9056-3760>

⁶ anton.dianov@ieee.org, <https://orcid.org/0000-0002-9778-6817>

Аннотация

Предмет исследования. С появлением силовых электронных устройств на полупроводниках с широкой запрещенной зоной, таких как SiC (карбид кремния) и GaN (нитрид галлия), частота широтно-импульсной модуляции увеличилась. В современных сервоприводах, использующих транзисторы на базе GaN, частота коммутации может достигать 100 кГц и более. В этом случае быстродействие электропривода ограничивается не частотой широтно-импульсной модуляции, а задержкой в измерении обратной связи по току. Однако при использовании в электроприводе датчиков тока с дельта-сигма модуляцией, задержку в обратной связи можно регулировать соответствующим изменением точности измерения. **Метод.** Рассмотрено применение менее точной, но более быстрой обратной связи в случае, когда ошибка между заданным и реальным значениями тока двигателя велика. В случае небольшой ошибки значения тока предложено использование более медленной, но точной обратной связи. Изменение скорости отработки задания требует одновременного изменения параметров регулятора тока. **Основные результаты.** Выполнено исследование представленного алгоритма на виртуальной модели сервопривода. Проведено сравнение адаптивного регулятора тока со стандартным регулятором при разных настройках. Показано, что за счет применения предложенного алгоритма быстродействие контура тока увеличивается вдвое без потери точности поддержания сигнала по сравнению со стандартным вариантом. Применение алгоритма позволило вдвое повысить частоту среза контура тока. Результаты получены для сигнала с линейной частотной модуляцией в качестве регулируемой величины. **Практическая значимость.** Предложенный способ измерения обратной связи в контуре тока позволяет увеличить быстродействие сервопривода в целом за счет увеличения быстродействия его внутреннего контура без потери точности регулирования.

Ключевые слова

контур тока, шунтовой датчик тока, GaN, сервопривод, дельта-сигма модуляция, адаптивный регулятор

Благодарности

Исследование выполнено за счет гранта Российского научного фонда (проект № 21-19-00696).

Ссылка для цитирования: Анучин А.С., Гуляева М.А., Лашкевич М.М., Жарков А.А., Чен Х., Дианов А.Н. Применение обратной связи переменной точности для повышения быстродействия контура тока в инверторах на базе GaN-транзисторов // Научно-технический вестник информационных технологий, механики и оптики. 2023. Т. 23, № 1. С. 192–201. doi: 10.17586/2226-1494-2023-23-1-192-201

Using variable-precision feedback to improve operational speed of the current loop in GaN-inverters

Alecksey S. Anuchin¹, Maria A. Gulyaeva²✉, Maxim M. Lashkevich³,
Alexandr A. Zharkov⁴, Hao Chen⁵, Anton N. Dianov⁶

^{1,2,4} National Research University “Moscow Power Engineering Institute” (MPEI), Moscow, 111250, Russian Federation

^{3,4} R&D Company “Vector”, LLC, Moscow, 111396, Russian Federation,

⁵ China University of Mining & Technology, Xuzhou, 100083, China

⁶ Daeyoung R&D Center, Seoul, 03140, South Korea

¹ Anuchin.alecksey@gmail.com, <https://orcid.org/0000-0002-1019-756X>

² tilgivelse@yandex.ru✉, <https://orcid.org/0000-0002-2326-1153>

³ maxsoftpage@yandex.ru, <https://orcid.org/0000-0002-9699-7145>

⁴ zarckov@mail.ru, <https://orcid.org/0000-0001-5628-8101>

⁵ hchen@cumt.edu.cn, <https://orcid.org/0000-0001-9056-3760>

⁶ anton.dianov@ieee.org, <https://orcid.org/0000-0002-9778-6817>

Abstract

With the advent of wide band-gap semiconductors like SiC and GaN, the frequency of pulse-width modulation has increased. In modern electric drives, the switching frequency can reach 100 kHz or more. In this case the performance of the drive is limited by the delay in the current feedback measurement. This delay can be changed by using delta-sigma modulators. This type of current sensors allows setting the measurement time. However, as the measurement time decreases, the accuracy of the feedback reduces. This paper proposes the algorithm in which the current controller uses variable-precision feedback. When the error between the reference and feedback is large, it uses faster but less accurate current feedback. When the error is small, it uses slower but accurate feedback. Changing the feedback measurement time requires changing the current controller gains. The algorithm was investigated on a virtual servo drive model. To evaluate the performance of the proposed regulator, the results were compared with standard regulators with different settings. It was proved that this approach allows increasing the speed of the current loop without loss in the transient performance. Besides, the algorithm increases the cut-off frequency in comparison with the standard slow and accurate controller.

Keywords

current loop, shunt current sensor, GaN, servo drive, delta-sigma modulation, adaptive controller

Acknowledgements

The research was carried out at the expense of a grant from the Russian Science Foundation (project No. 21-19-00696).

For citation: Anuchin A.S., Gulyaeva M.A., Lashkevich M.M., Zharkov A.A., Chen H., Dianov A.N. Using variable-precision feedback to improve operational speed of the current loop in GaN-inverters. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 2023, vol. 23, no. 1, pp. 192–201 (in Russian). doi: 10.17586/2226-1494-2023-23-1-192-201

Введение

В большинстве промышленных электроприводов используются системы управления с подчиненным регулированием. Контуры тока и скорости в таких системах являются внутренними, а контур положения, используемый в сервоприводах, — внешним. Быстродействие каждого внешнего контура ограничивается быстродействием внутреннего и обычно в два раза медленнее. Таким образом, быстродействие внутреннего контура определяет быстродействие всего электропривода, поэтому крайне важно сделать его максимально быстрым.

Быстродействие контура тока в электроприводе с цифровым управлением ограничено задержками: временем измерения тока и расчета контура управления, а также периодом широтно-импульсной модуляции (ШИМ). Распределение данных задержек менялось в течение долгого времени. С момента изобретения в 1980-х годах кремниевых IGBT-транзисторов на смену станочным приводам с тиристорными выпрямителями пришли прецизионные привода с частотами коммутаций до 10 кГц и более. В середине 90-х годов XX века системы управления, реализованные на цифровых сигнальных микроконтроллерах, при использовании

оптимизации программного кода смогли рассчитать контур тока за 20 мкс. В настоящее время библиотека функций «Fast Current Loop», разработанная компанией «Texas Instruments», позволяет сократить время расчета до 1 мкс. В связи с этим наибольшая задержка, варьирующаяся от 50 до 500 мкс, обусловлена ограниченной частотой коммутации IGBT-транзисторов.

С изобретением полупроводников с широкой запрещенной зоной, таких как SiC (карбид кремния) и GaN (нитрид галлия), частота ШИМ выросла. Наибольших частот достигают сервоприводы, разработанные с использованием GaN-транзисторов и напряжением звена постоянного тока 310 В, в которых частота ШИМ может достигать 200 кГц. Для увеличения быстродействия электропривода задаваемое значение тока следует обновлять дважды за период ШИМ: вверху и внизу несущего сигнала таймера перед применением уставок сравнения. Соответственно, время расчета контура тока и его измерение в обратной связи не должно превышать 2,5 мкс. Таким образом, важные факторы, ограничивающие быстродействие электропривода — время расчета системы управления и задержка на измерение тока.

Применение GaN-транзисторов способствует уменьшению объема и веса инвертора, размер которого определяется датчиками тока, которые обычно

представлены датчиками на основе эффекта Холла компенсационного типа. Датчики на эффекте Холла прямого измерения более компактны, однако, чем меньше их конструкция, тем сильнее электромагнитные помехи влияют на точность измерения. По этой причине датчики прямого измерения практически не используются в прецизионных приводах. В качестве альтернативного решения в настоящее время популярны шунтовые датчики, использующие дельта-сигма модуляцию ($\Delta\Sigma$ -модуляцию) [1–4]. Они измеряют ток непосредственно в фазе двигателя и обеспечивают высокую помехозащищенность. Проблема заключается в том, что точность таких датчиков зависит от времени измерения, необходимого для демодуляции битового потока, полученного от $\Delta\Sigma$ -модулятора. Так, сигнал с точностью 5 бит может быть получен за 1,2 мкс на максимальной частоте тактирования $\Delta\Sigma$ -модулятора равной 20 МГц, 8-битный — за 2,4 мкс и т. д., что сопоставимо с периодом ШИМ-инвертора на базе GaN-транзисторов [5, 6].

В нормальном режиме работы для точного поддержания тока требуется время измерения 4,8 мкс, если номинальный ток не превышает $\pm 50\%$ диапазона аналого-цифрового преобразователя (АЦП) [7]. Но если разница между заданным значением тока и обратной связью достаточно велика, то точность измерения тока не влияет на качество регулирования тока. Другими словами, точность может быть снижена ради получения более быстрого отклика в контуре тока.

В настоящей работе рассмотрен способ регулирования тока с переменным временем измерения обрат-

ной связи, которое зависит от рассогласования между заданным значением и обратной связью регулятора тока. При этом пропорциональный и интегральный коэффициенты регулятора меняются для поддержания максимальной полосы пропускания контура тока.

Контур тока и его настройка

Общая структура контура тока в виртуальной модели сервопривода показана на рис. 1. Разница между заданным значением тока и обратной связью поступает на пропорционально-интегральный (ПИ) регулятор, выход которого является заданием напряжения, которое подается на модуль ШИМ. На следующем периоде ШИМ-инвертор реализует задание напряжения (экстраполятор нулевого порядка на рис. 1), а значение тока считывается с помощью датчика и передается на регулятор тока.

Быстродействие контура тока ограничивается суммой задержек, которые не могут быть скомпенсированы регулятором тока. Например, если значение обратной связи точно известно в начале периода ШИМ и регулятор тока считается мгновенно, задание напряжения рассчитывается по формуле [8]:

$$v_{ref} = \frac{L}{T_{PWM}}(i_{ref} - i_{fbk}),$$

где L — индуктивность обмотки двигателя; T_{PWM} — период ШИМ (Pulse-Width Modulation, PWM); i_{ref} — задание на ток; i_{fbk} — обратная связь по току. Такой

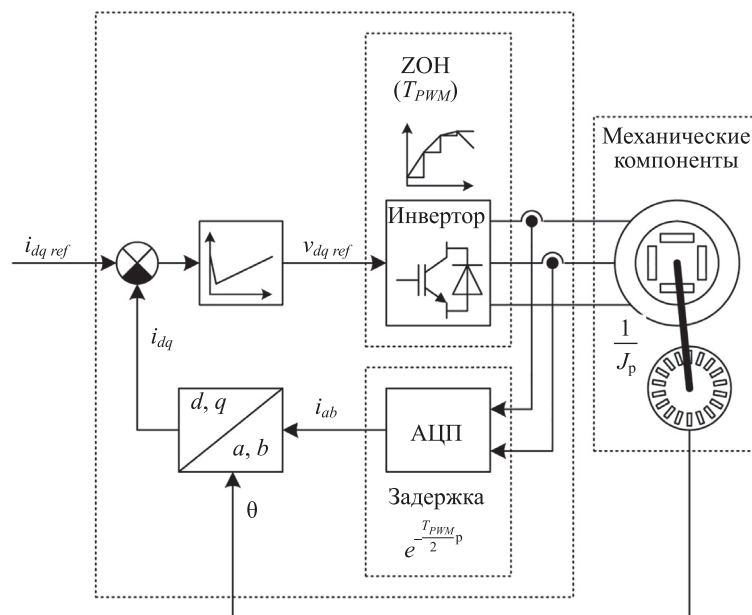


Рис. 1. Структура стандартного контура тока: ZOH — экстраполятор нулевого порядка; $i_{dq\ ref}$ — задания на ток по осям d и q ; $v_{dq\ ref}$ — задания на напряжение по осям d и q ; i_{dq} — токи по осям d и q ; i_{ab} — токи фаз A и B ; d, q — вращающиеся оси, связанные с ротором двигателя; a, b — фазы A и B ; θ — угол положения ротора; p — оператор дифференцирования; J — момент инерции двигателя

Fig. 1. Structure of the conventional current loop: ZOH — zero order hold; $i_{dq\ ref}$ — current references along d and q axes; $v_{dq\ ref}$ — voltage references along d and q axes; i_{dq} — currents along d and q axes; i_{ab} — currents of phases A and B ; d, q — rotating axes connected with the rotor; a, b — phases A and B ; θ — rotor position angle; p — differentiation operator; and J — inertia moment

подход предполагает, что задержки, обычно присущие системе, отсутствуют.

В этом случае заданный ток достигается через один период ШИМ, если напряжение не касается ограничений. Для упрощения представления регулятора компенсация обратной связи по ЭДС не учитывается. Сопротивление обмотки будет учтено в интегральной составляющей регулятора:

$$v_{ref}[k] = v_{ref}[k-1] + \frac{R}{T_{PWM}}(i_{ref} - i_{fbk})h,$$

где R — сопротивление обмотки; h — шаг интегрирования, равный периоду расчета регулятора тока; $[k]$ — обозначение значения напряжения, полученного на текущем шаге расчета; $[k-1]$ — обозначение значения напряжения, полученного на предыдущем шаге расчета.

Однако в реальных условиях система содержит задержки, перечисленные ранее. Для стандартных систем, у которых время измерения тока (k_P) и последующий расчет регулятора тока (k_I) синхронизированы с серединой периода ШИМ, пропорциональный и интегральный коэффициенты ПИ-регулятора снижаются в два раза.

$$\begin{cases} k_P = \frac{L}{2T_{PWM}} \\ k_I = \frac{R}{2T_{PWM}} \end{cases} \quad (1)$$

Такая настройка обеспечивает устойчивое поведение контура тока со временем переходного процесса около 5 периодов ШИМ и перерегулированием 5 %. Множитель «2» в знаменателе (1) учитывает то, что измерение тока не совпадает с концом периода ШИМ, а рассчитывается в середине периода, а также задержку

приложения напряжения, равную $0,5 T_{PWM}$ от начала периода ШИМ. Временная диаграмма представлена на рис. 2.

В контуре тока (рис. 2), показаны следующие задержки: измерение тока (1); время расчета регулятора тока (2); задержка между рассчитанным заданием напряжения и началом следующего периода ШИМ, где оно будет применено (3); задержка, которая составляет половину периода ШИМ, так как среднее значение задания напряжения приложено к обмоткам двигателя после половины периода ШИМ (4).

За последние десятилетия величины этих задержек изменялись. Например, долгое время расчет регуляторов тока занимал много времени из-за отсутствия цифровых сигнальных микроконтроллеров и слабой оптимизации кода. В настоящее время расчет контура тока с помощью библиотеки функций «Fast Current Loop» от компании «Texas Instruments» занимает 1 мкс и может быть дополнительно оптимизирован. Прорыв в развитии полупроводников с широкой запрещенной зоной, таких как GaN и SiC, содействовал получению частоты ШИМ 100 кГц и более [9–12], ранее недостижимой в электроприводе. Время измерения тока за последние годы уменьшилось с 500 до 150 нс, и новые принципы измерения продолжают появляться.

Шунтовое измерение тока с применением $\Delta\Sigma$ -модуляции становится все более популярным благодаря своей простоте, цене и точности. Это наиболее компактное решение по сравнению с датчиками на эффекте Холла компенсационного типа при аналогичной точности и помехозащищенности. Появляются силовые модули с уже встроенными шунтами [13]. Однако, чем большая точность требуется, тем больше времени необходимо для демодуляции битового потока, выходная частота которого в настоящее время ограничена на уровне 20 МГц. Частота $\Delta\Sigma$ -модулятора определяет

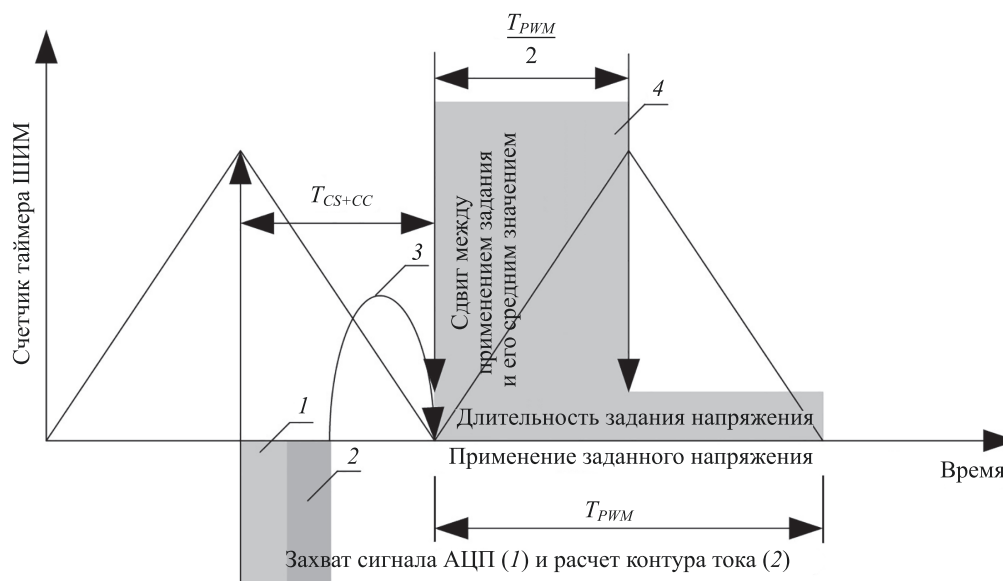


Рис. 2. Временная диаграмма контура управления током: T_{PWM} — период ШИМ; T_{CS+CC} (current sensing — измерение тока, current controller — регулятор тока) — время на измерение тока и расчет контура тока

Fig. 2. Time diagram of the current loop control: T_{PWM} — duration of PWM cycle, T_{CS+CC} — current sensing and current controller execution time

время, необходимое для измерения тока: например, для 10,5-битного разрешения АЦП оно равно 4,8 мкс, для 12-битного — 9,6 мкс, что слишком долго по сравнению с другими задержками в контуре тока.

Когда времена расчета системы управления и измерения тока близки или больше, чем период ШИМ, они должны быть учтены в некомпенсируемой постоянной времени регулятора. Учитывая это, формулу (1) перепишем в виде:

$$\begin{cases} k_P = \frac{L}{1(T_{PWM} + T_{CS+CC})}, \\ k_I = \frac{R}{2(T_{PWM} + T_{CS+CC})} \end{cases}$$

где T_{CS+CC} — время на измерение тока и расчет контура тока.

Шунтовое измерение тока

Токовый шунт подключается между выходом инвертора и фазой двигателя. Источник питания верхнего драйвера IGBT-транзистора может быть использован для питания интегральной схемы $\Delta\Sigma$ -модулятора (рис. 3). Такая схема обычно реализует $\Delta\Sigma$ -модуляцию второго порядка (рис. 4, а) [14], которая имеет лучшее

распределение единиц и нулей генерируемого битового потока по сравнению с $\Delta\Sigma$ -модуляцией первого порядка [15]. Выходной битовый поток (рис. 4, б) содержит информацию о токе, протекающем через фазу двигателя. Схема $\Delta\Sigma$ -модулятора синхронизируется тактовой частотой 20 МГц — f_{CLK} (clock frequency, CLK). Демодуляция битового потока осуществляется с помощью специального периферийного устройства современных микроконтроллеров. Битовый поток отфильтруем с помощью фильтра нижних частот. В общем случае Sinc-фильтр первого порядка является фильтром скользящего среднего и может быть представлен уравнениями фильтра с бесконечной или конечной импульсными характеристиками, при этом результат работы остается тем же:

$$H_{FIR}(z) = \frac{1}{N}(1 + z^{-1} + z^{-2} + \dots + z^{-(N-1)}),$$

где H_{FIR} — функция фильтра с конечной импульсной характеристикой (finite impulse response, FIR); z — оператор сдвига во времени; N — коэффициент передискретизации (Oversampling Ratio, OSR) или количество битов битового потока, обрабатываемых в фильтре.

Точность измерения тока зависит от количества битов в битовом потоке, используемом для одного измерения, и от порядка цифрового фильтра. Чем больше

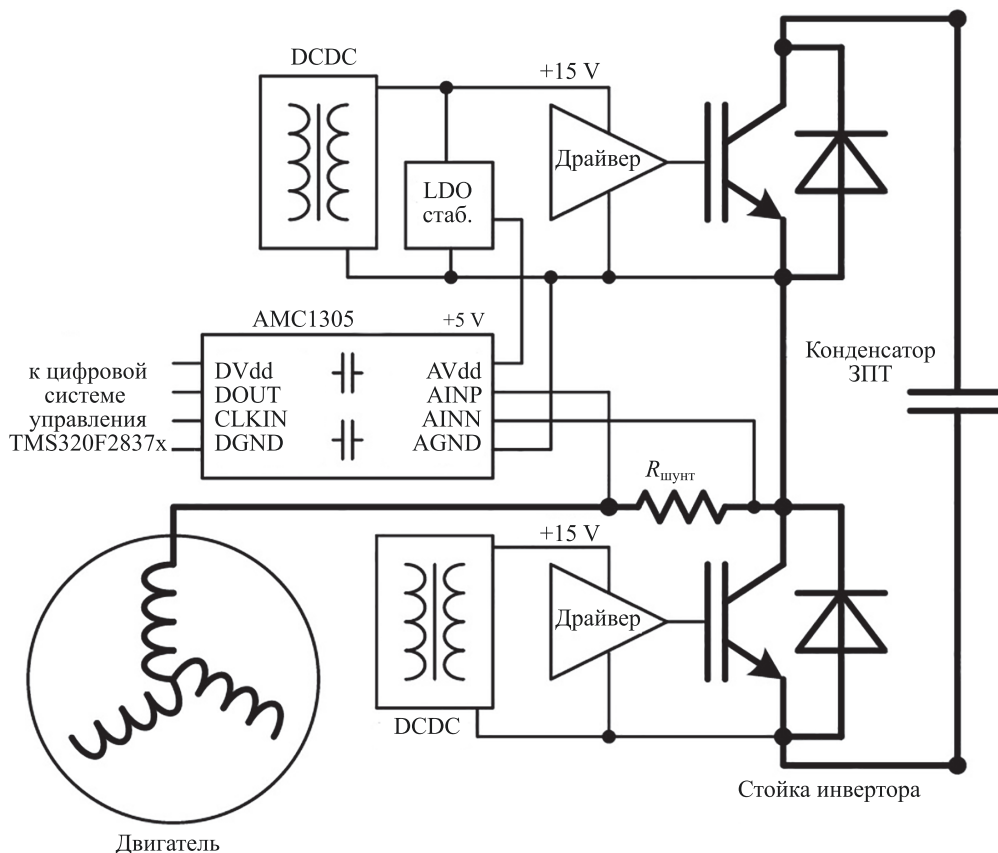


Рис. 3. Стандартная топология схемы измерения тока с помощью шунтового датчика для электроприводов:

DCDC — преобразователь уровней постоянного напряжения; LDO (low-dropout regulator — регулятор с низким перепадом) — линейный стабилизатор напряжения; ЗПТ — звено постоянного тока; $R_{шунт}$ — сопротивление шунта

Fig. 3. Basic shunt current sensing topology for electric drives: DCDC — DC-to-DC converter; LDO — low-dropout regulator; ЗПТ — DC link; $R_{шунт}$ — shunt resistance

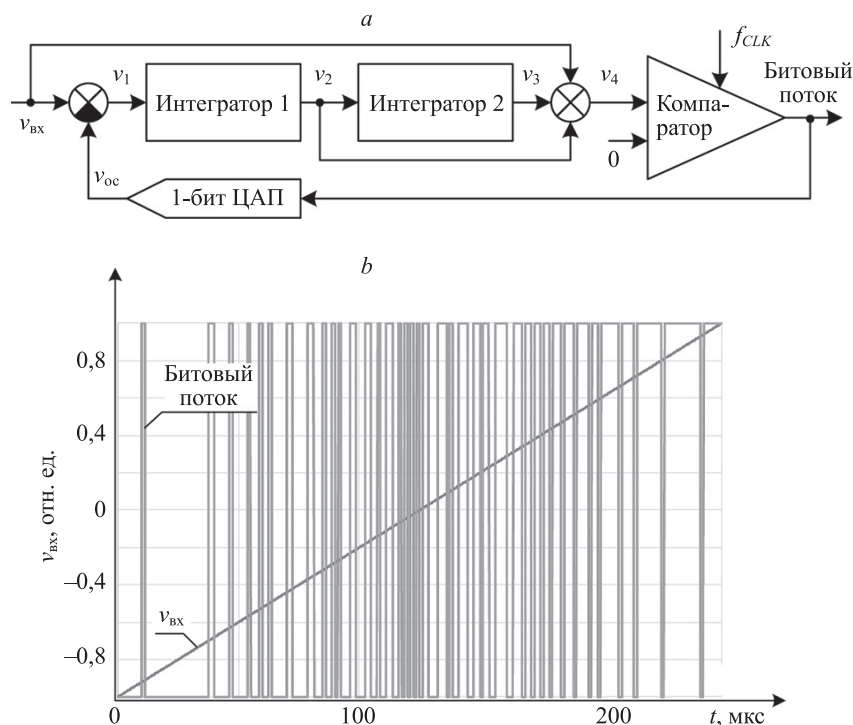


Рис. 4. Структура $\Delta\Sigma$ -модулятора второго порядка (а) и его работа (б): v_{BX} — входное напряжение, измеряемое шунтовым датчиком тока; v_{OC} — обратная связь по напряжению; v_1 — ошибка между входным напряжением и обратной связью; v_2, v_3 — напряжения после последовательных интеграторов; v_4 — сумма напряжений; ЦАП — цифро-аналоговый преобразователь

Fig. 4. Second-order delta-sigma modulator structure (a) and its operation (b): v_{BX} — input voltage measured by a shunt current sensor; v_{OC} — voltage feedback; v_1 — error between input voltage and feedback; v_2, v_3 — voltages after series integrators; v_4 — the sum of voltages; ЦАП — digital-to-analog converter

количество битов или коэффициент передискретизации, тем выше будет точность. Sinc3-фильтр имеет наиболее предпочтительный порядок фильтрации, обеспечивающий наибольшую точность в большом диапазоне коэффициентов передискретизации. При этом с увеличением точности, за счет увеличения количества обрабатываемых бит, увеличивается и продолжительность измерения. Следовательно, настройка цифрового фильтра должна быть компромиссной между точностью и скоростью измерения. Время измерения тока вычислим с помощью уравнения:

$$t = \frac{OSR}{f_{CLK}} F_{order},$$

где F_{order} — порядок Sinc-фильтра. Стандартные времена измерения — 1,2 мкс, 2,4 мкс, 4,8 мкс и 9,6 мкс. В два раза большая продолжительность времени обеспечивает примерно в 6 раз большую точность. Сравнение погрешностей измерения показано на рис. 5 [2].

Адаптивный регулятор тока с переменной точностью обратной связи

Чем быстрее измерение тока и меньше период ШИМ, тем выше может быть пропорциональный коэффициент регулятора тока и меньше время регулирова-

ния. С уменьшением времени точность измерения тока снижается, а шум в обратной связи влияет на работу контура тока, приводя к колебаниям вокруг заданного значения. Однако, если разница между заданным значением и обратной связью по току достаточно велика и недостижима в одном периоде расчета системы управления, то необходимости в высокой точности измерения для регулирования тока нет.

Другими словами, время измерения может быть выбрано путем анализа ошибки между заданием и обратной связью, а коэффициенты регулятора должны быть пересчитаны относительно времени измерения. Такой подход способен повысить быстродействие контура тока, сократив время регулирования. Блок-схема предлагаемого регулятора тока показана на рис. 6. Для простоты схема показана только для тока одной оси, при этом компенсация противо-ЭДС исключена.

Измеряемый ток преобразуется $\Delta\Sigma$ -модулятором в битовый поток, поступающий на три фильтра с тремя разными временами демодуляции. Самый быстрый из них получает данные за 2,4 мкс (рис. 6: Sinc3-фильтр, $OSR = 16$). В самом медленном фильтре $OSR = 64$, а время измерения тока 9,6 мкс. Разница между заданным значением тока и самым быстрым сигналом обратной связи сравнивается с двумя пороговыми значениями (уставками 1 и 2), определяющими, какие параметры обратной связи и коэффициенты регулятора тока должны быть использованы. Соответствующие параметры

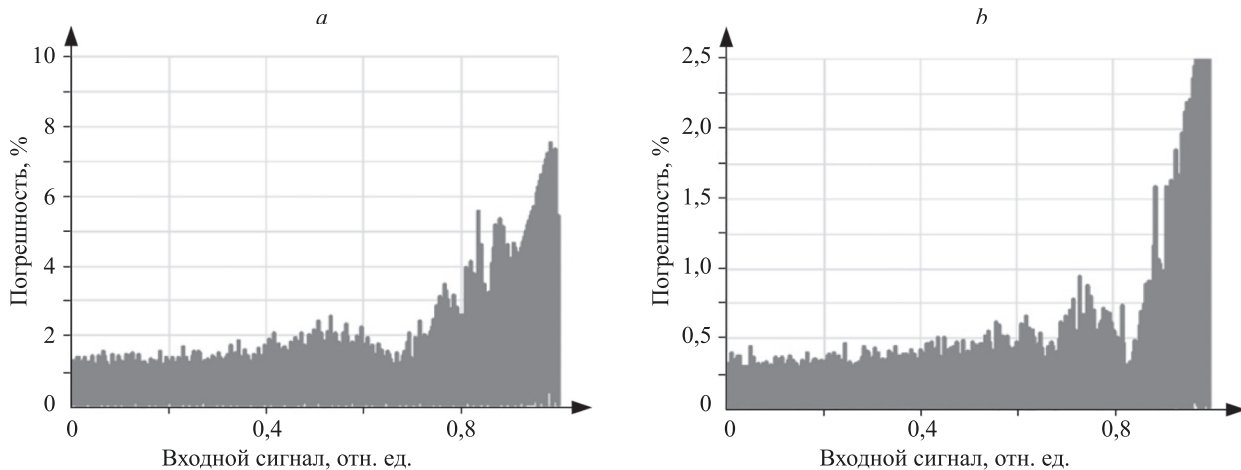


Рис. 5. Погрешность измерения тока при использовании Sinc3-фильтра и OSR = 8 (a), и OSR = 16 (b). Отрицательные значения не показаны, поскольку являются симметричными положительным

Fig. 5. Error of the measurement using Sinc3-filter with OSR = 8 (a) and OSR = 16 (b)

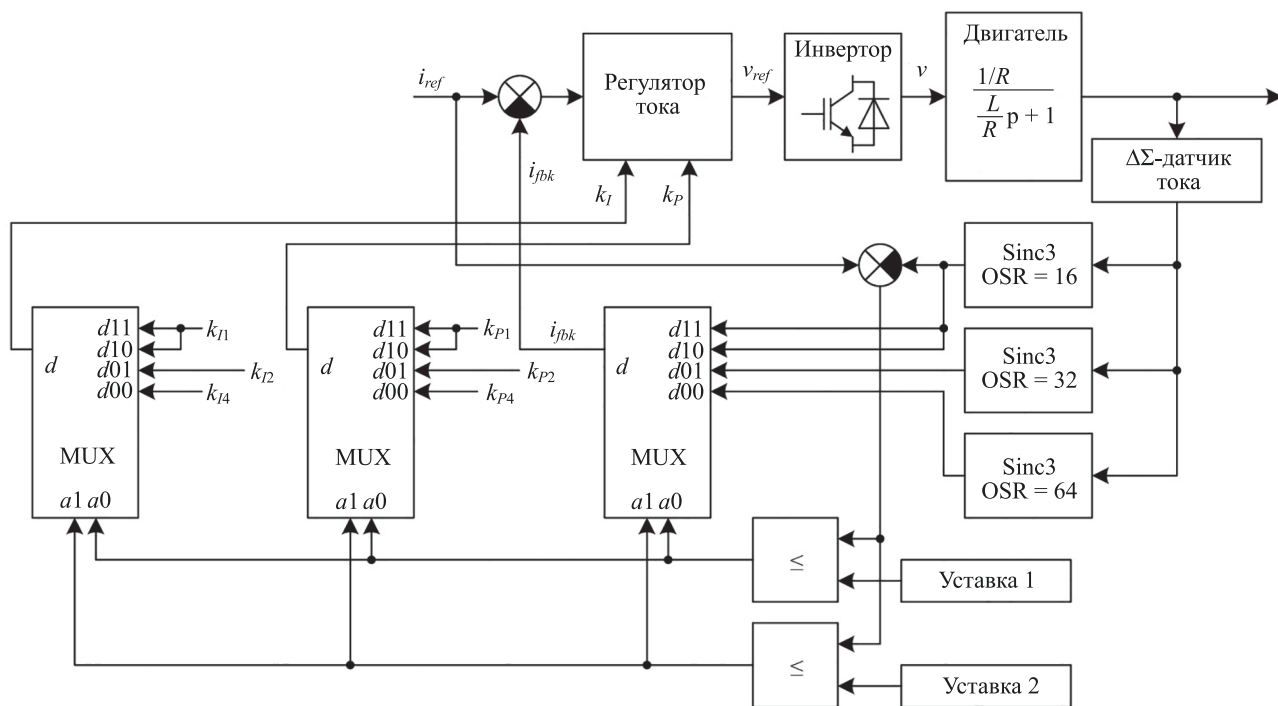


Рис. 6. Блок-схема адаптивного регулятора тока: $a0, a1$ — двоичные результаты сравнения значения обратной связи по току с уставками 1 и 2 соответственно; $d00, d01, d10, d11$ — наборы параметров (интегральный и пропорциональный коэффициенты, обратная связь), выбираемые в зависимости от результатов сравнения $a0$ и $a1$

Fig. 6. Adaptive current controller block diagram: $a0, a1$ — binary results of comparing the current feedback value with thresholds 1 and 2, respectively; $d00, d01, d10, d11$ — sets of parameters (integral and proportional coefficients, feedback) selected depending on the results of comparison of $a0$ and $a1$

регулятора и источник обратной связи тока выбираются мультиплексором и подаются на блок регулятора.

Результаты моделирования

Реакция на скачок. Параметры моделируемого привода показаны в таблице. Инвертор представлен экстраполятором нулевого порядка. Первый экспери-

мент выполнен при подаче задания скачком, которое изменялось от 0 до 10 А. При эксперименте выполнено исследование трех настроек. Первая — с точным измерением тока и «медленным» регулятором. Во второй настройке использовано самое быстрое измерение тока с низкой точностью и большими коэффициентами регулятора тока («быстрый» регулятор). Третья настройка является предложенным адаптивным регулятором тока

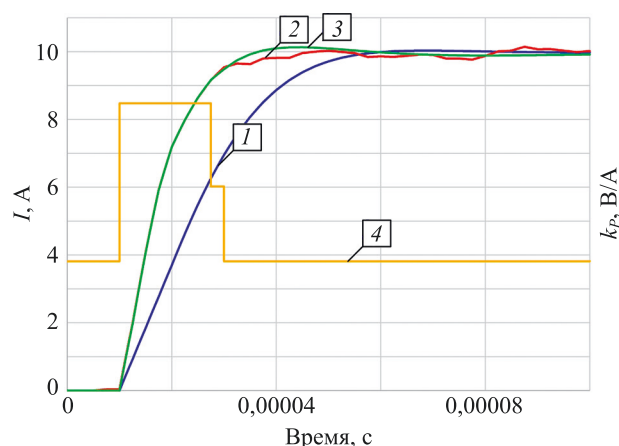


Рис. 7. Переходная характеристика регулятора тока при различных настройках: 1 — «медленный» регулятор тока; 2 — «быстрый» регулятор тока; 3 — адаптивный регулятор тока с переменными коэффициентами и точностью измерения; 4 — пропорциональный коэффициент адаптивного регулятора тока

Fig. 7. Step response of the current controller with various tunings: 1 — accurate current measurement and slow current controller tuning; 2 — fast current measurement with rapid current controller tuning; 3 — adaptive current controller with variable gains and measurement accuracy; 4 — proportional gain of the adaptive current controller

с переменными точностью обратной связи и параметрами регулятора. Результаты моделирования представлены на рис. 7.

В результате моделирования, видно, что самый затянутый процесс соответствует «медленному» регулятору с наибольшим временем измерения и наименьшими коэффициентами регулятора тока. Отметим, что данный регулятор продемонстрировал высокую точность регулирования. «Быстрый» регулятор показал низкое качество поддержания тока из-за значительных погрешностей в измерении. Регулятор пытается стабилизировать обратную связь, но из-за сильных помех ток колеблется вокруг заданного значения.

Адаптивный регулятор тока изменил свои коэффициенты усиления во время работы. При высокой ошибке между заданным значением и обратной связью использованы самые большие коэффициенты регу-

лятора k_{p1} и k_{I1} с наиболее быстрой обратной связью (OSR = 16).

В установившемся режиме применена самая медленная настройка (k_{p3} и k_{I3}), но наиболее точное измерение тока (OSR = 64), что помогло стабилизировать ток. Время регулирования с адаптивным регулятором тока аналогично времени регулирования с самой быстрой настройкой.

Реакция на сигнал с линейной частотной модуляцией. Второй эксперимент проведен для сравнения частот среза «медленного» регулятора и предложенного адаптивного регулятора (третья настройка). В качестве задания на регулятор тока установлен сигнал с линейной частотной модуляцией (ЛЧМ) с амплитудой 10 А и возрастающей частотой. Результаты моделирования представлены на рис. 8. Стандартный регулятор тока достиг частоты среза через 0,004 с при входной частоте

Таблица. Параметры моделируемого электропривода
Table. Parameters of the drive model

Параметр		Значение
Частота ШИМ, кГц		400
Напряжение звена постоянного тока, В		310
Индуктивность фазы, мкГн		100
Сопротивления фазы, Ом		3
Время расчета контура тока, мкс		1
Время измерения при OSR, мкс	16	2,4
	64	9,6
Пропорциональный коэффициент, В/А	«быстрого» регулятора при OSR = 16 k_{p1}	8,5
	регулятора при OSR = 32 k_{p2}	6,0
	«медленного» регулятора при OSR = 64 k_{p3}	3,8
Интегральный коэффициент, В/(А·с)	«быстрого» регулятора тока при OSR = 16 k_{I1}	$0,25 \cdot 10^6$
	регулятора тока при OSR = 32 k_{I2}	$0,18 \cdot 10^6$
	«медленного» регулятора тока при OSR = 64 k_{I3}	$0,11 \cdot 10^6$

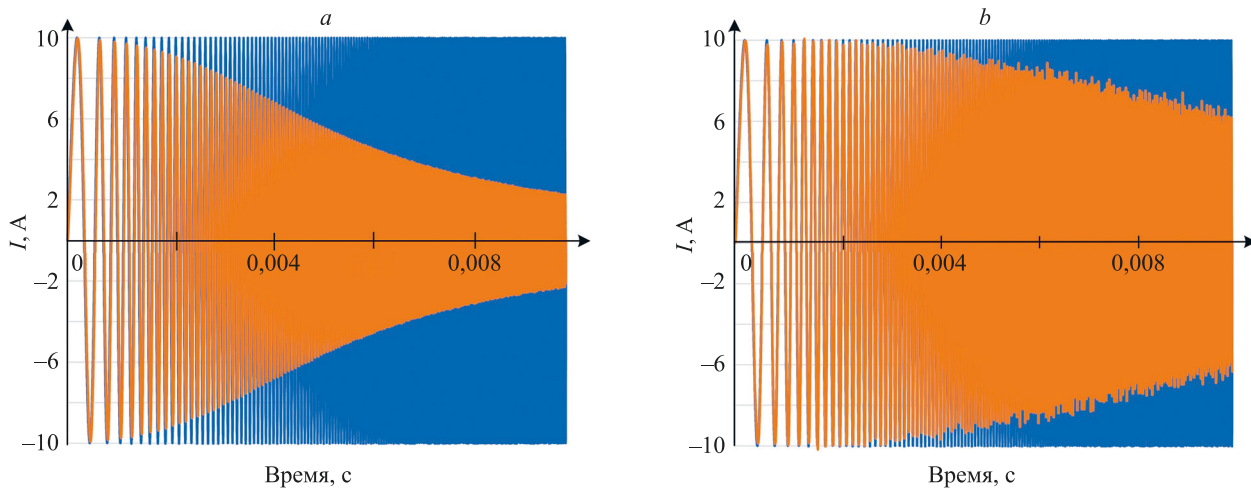


Рис. 8. Реакция медленного (а) и адаптивного (б) регуляторов на ЛЧМ-сигнал. Задание тока (кривая голубого цвета) и обратная связь (кривая оранжевого цвета)

Fig. 8. Chirp signal response of slow and accurate current controller (a) and adaptive current controller (b): blue — current reference; orange — actual current

заданного сигнала около 13 кГц. Для адаптивного регулятора тока такое же ослабление реализуемого тока относительно задания в $1/\sqrt{2}$ было достигнуто за 0,009 с при входной частоте 27 кГц. В данном эксперименте показатели предложенного решения оказались в два раза выше, чем у стандартного регулятора.

Заключение

В работе получено, что с увеличением частоты коммутации устройств с широкой запрещенной зоной влияние задержек в системе управления меняется. Преобладавшая ранее задержка, связанная с ограниченной частотой широтно-импульсной модуляции в силовых преобразователях, оказывается меньше, чем задержка в управлении и измерении тока. Сделан вывод, что традиционный подход к синтезу регуляторов

тока не может обеспечить оптимальную работу контура тока.

Предложен адаптивный регулятор тока, который изменяет свои коэффициенты в зависимости от ошибки между заданным и фактическим током, а также настраивает точность и время измерения обратной связи по току, получаемой с помощью шунтового датчика тока и $\Delta\Sigma$ -модуляции. Такой способ управления значительно повышает быстродействие контура тока. Для конкретных параметров привода в два раза улучшены как время регулирования, так и частота среза контура тока.

Описанные результаты получены на виртуальной модели электропривода. Дальнейшие исследования будут направлены на экспериментальную проверку предлагаемого регулятора с использованием GaN сервопривода с системой управления на программируемой логической интегральной схеме.

Литература

1. Homann M., Schumacher W. High bandwidth phase voltage and phase current control loop of a permanent magnet synchronous motor based on delta sigma bitstreams // Proc. of the 2016 18th European Conference on Power Electronics and Applications (EPE'16 ECCE Europe), 2016. P. 1–10. <https://doi.org/10.1109/EPE.2016.7695308>
2. Hirota A., Saha S., Mun S.-P., Nakaoka M. An advanced simple configuration delta-sigma modulation three-phase inverter implementing space voltage vector approach // Proc. of the 2007 IEEE Power Electronics Specialists Conference. 2007. P. 453–457. <https://doi.org/10.1109/PESC.2007.4342030>
3. Jarzebowicz L. Impact of low switching-to-fundamental frequency ratio on predictive current control of PMSM: A simulation study // Proc. of the 2018 25th International Workshop on Electric Drives: Optimization in Control of Electric Drives (IWED), 2018. P. 1–5. <https://doi.org/10.1109/IWED.2018.8321400>
4. Li C., Vankayalapati B., Akin B. Latency compensation of SD-ADC for high performance motor control and diagnosis // Proc. of the IEEE 13th International Symposium on Diagnostics for Electrical Machines, Power Electronics and Drives (SDEMPED), 2021. P. 289–294. <https://doi.org/10.1109/SDEMPED51010.2021.9605522>
5. Anuchin D., Surnin D., Lashkevich M. Accuracy analysis of shunt current sensing by means of delta-sigma modulation in electric drives

References

1. Homann M., Schumacher W. High bandwidth phase voltage and phase current control loop of a permanent magnet synchronous motor based on delta sigma bitstreams. *Proc. of the 2016 18th European Conference on Power Electronics and Applications (EPE'16 ECCE Europe)*, 2016, pp. 1–10. <https://doi.org/10.1109/EPE.2016.7695308>
2. Hirota A., Saha S., Mun S.-P., Nakaoka M. An advanced simple configuration delta-sigma modulation three-phase inverter implementing space voltage vector approach. *Proc. of the 2007 IEEE Power Electronics Specialists Conference*, 2007, pp. 453–457. <https://doi.org/10.1109/PESC.2007.4342030>
3. Jarzebowicz L. Impact of low switching-to-fundamental frequency ratio on predictive current control of PMSM: A simulation study. *Proc. of the 2018 25th International Workshop on Electric Drives: Optimization in Control of Electric Drives (IWED)*, 2018, pp. 1–5. <https://doi.org/10.1109/IWED.2018.8321400>
4. Li C., Vankayalapati B., Akin B. Latency compensation of SD-ADC for high performance motor control and diagnosis. *Proc. of the IEEE 13th International Symposium on Diagnostics for Electrical Machines, Power Electronics and Drives (SDEMPED)*, 2021, pp. 289–294. <https://doi.org/10.1109/SDEMPED51010.2021.9605522>
5. Anuchin D., Surnin D., Lashkevich M. Accuracy analysis of shunt current sensing by means of delta-sigma modulation in electric drives.

- // Proc. of the 2018 17th International Ural Conference on AC Electric Drives (ACED). 2018. P. 1–5. <https://doi.org/10.1109/ACED.2018.8341706>
6. Zhuo Y.-C., Hsu C.-J., Qiu Z.-H., Lai Y.-S. Bandwidth boost method of current control for servo motor drives with current observer // Proc. of the 2021 IEEE International Future Energy Electronics Conference (IFEEC). 2021. P. 1–6. <https://doi.org/10.1109/IFEEC53238.2021.9661769>
7. Anuchin A., Lashkevich M., Shpak D., Aliamkin D., Zharkov A., Briz F. Current control of ac drives using shunt current sensors and delta-sigma modulation // Proc. of the IECON 2018 — 44th Annual Conference of the IEEE Industrial Electronics Society. 2018. P. 445–449. <https://doi.org/10.1109/IECON.2018.8591705>
8. Anuchin A., Kozachenko V. Current loop dead-beat control with the digital PI-controller // Proc. of the 2014 16th European Conference on Power Electronics and Applications. 2014. P. 1–8. <https://doi.org/10.1109/EPE.2014.6910795>
9. Ohta K., Kanazawal Y., Akatuka H., Hori S., Doki S., Tadano H., Shiozaki K. Variable switching frequency control for efficiency improvement of motor drive system by using GaN three phase inverter // Proc. of the 2020 IEEE International Conference on Industrial Technology (ICIT). 2020. P. 119–123. <https://doi.org/10.1109/ICIT45562.2020.9067266>
10. Lu J., Hou R., Di Maso P., Styles J. A GaN/Si hybrid T-Type Three-level configuration for electric vehicle traction inverter // Proc. of the 2018 IEEE 6th Workshop on Wide Bandgap Power Devices and Applications (WiPDA). 2018. P. 77–81. <https://doi.org/10.1109/WiPDA.2018.8569194>
11. Shang S., Yang M., Xu D. Design and research of servo drive system based on GaN power device // Proc. of the 2018 21st International Conference on Electrical Machines and Systems (ICEMS). 2018. P. 1313–1317. <https://doi.org/10.23919/ICEMS.2018.8549260>
12. Kurniawan A.W., Firmansyah E., Wijaya F.D. Experiment result of high frequency switching SiC mosfet gate driver // Proc. of the 2022 14th International Conference on Information Technology and Electrical Engineering (ICITEE). 2022. P. 54–57. <https://doi.org/10.1109/ICITEE56407.2022.9954115>
13. Spang M., Hofstoetter N. Evaluation of current measurement accuracy for a power module with integrated shunt resistors // Proc. of the PCIM Europe 2017; International Exhibition and Conference for Power Electronics, Intelligent Motion, Renewable Energy and Energy Management. 2017. P. 1–8.
14. AMC1305x Small High-Precision Reinforced Isolated Delta-Sigma Modulators Datasheet Texas Instruments. 2017.
15. Schreier R. Second and Higher-Order Delta-Sigma Modulators. MEAD, March 2008.
16. // Proc. of the 2018 17th International Ural Conference on AC Electric Drives (ACED). 2018. P. 1–5. <https://doi.org/10.1109/ACED.2018.8341706>
6. Zhuo Y.-C., Hsu C.-J., Qiu Z.-H., Lai Y.-S. Bandwidth boost method of current control for servo motor drives with current observer. *Proc. of the 2021 IEEE International Future Energy Electronics Conference (IFEEC)*, 2021, pp. 1–6. <https://doi.org/10.1109/IFEEC53238.2021.9661769>
7. Anuchin A., Lashkevich M., Shpak D., Aliamkin D., Zharkov A., Briz F. Current control of ac drives using shunt current sensors and delta-sigma modulation. *Proc. of the IECON 2018 — 44th Annual Conference of the IEEE Industrial Electronics Society*, 2018, pp. 445–449. <https://doi.org/10.1109/IECON.2018.8591705>
8. Anuchin A., Kozachenko V. Current loop dead-beat control with the digital PI-controller. *Proc. of the 2014 16th European Conference on Power Electronics and Applications*, 2014, pp. 1–8. <https://doi.org/10.1109/EPE.2014.6910795>
9. Ohta K., Kanazawal Y., Akatuka H., Hori S., Doki S., Tadano H., Shiozaki K. Variable switching frequency control for efficiency improvement of motor drive system by using GaN three phase inverter. *Proc. of the 2020 IEEE International Conference on Industrial Technology (ICIT)*, 2020, pp. 119–123. <https://doi.org/10.1109/ICIT45562.2020.9067266>
10. Lu J., Hou R., Di Maso P., Styles J. A GaN/Si hybrid T-Type Three-level configuration for electric vehicle traction inverter. *Proc. of the 2018 IEEE 6th Workshop on Wide Bandgap Power Devices and Applications (WiPDA)*, 2018, pp. 77–81. <https://doi.org/10.1109/WiPDA.2018.8569194>
11. Shang S., Yang M., Xu D. Design and research of servo drive system based on GaN power device. *Proc. of the 2018 21st International Conference on Electrical Machines and Systems (ICEMS)*, 2018, pp. 1313–1317. <https://doi.org/10.23919/ICEMS.2018.8549260>
12. Kurniawan A.W., Firmansyah E., Wijaya F.D. Experiment result of high frequency switching SiC mosfet gate driver. *Proc. of the 2022 14th International Conference on Information Technology and Electrical Engineering (ICITEE)*, 2022, pp. 54–57. <https://doi.org/10.1109/ICITEE56407.2022.9954115>
13. Spang M., Hofstoetter N. Evaluation of current measurement accuracy for a power module with integrated shunt resistors. *Proc. of the PCIM Europe 2017; International Exhibition and Conference for Power Electronics, Intelligent Motion, Renewable Energy and Energy Management*, 2017, pp. 1–8.
14. AMC1305x Small High-Precision Reinforced Isolated Delta-Sigma Modulators Datasheet Texas Instruments, 2017.
15. Schreier R. *Second and Higher-Order Delta-Sigma Modulators*. MEAD, March 2008.

Авторы

Анучин Алексей Сергеевич — доктор технических наук, профессор, заведующий кафедрой, Национальный исследовательский университет «МЭИ», Москва, 111250, Российская Федерация, [sc 56168843400](https://orcid.org/0000-0002-1019-756X), <https://orcid.org/0000-0002-1019-756X>, Anuchin.alecksey@gmail.com

Гуляева Мария Андреевна — ассистент, Национальный исследовательский университет «МЭИ», Москва, 111250, Российская Федерация, [sc 57195258176](https://orcid.org/0000-0002-2326-1153), <https://orcid.org/0000-0002-2326-1153>, tilgivelse@yandex.ru

Лашкевич Максим Михайлович — кандидат технических наук, ведущий инженер-программист, ООО «НПФ ВЕКТОР», Москва, 111396, Российская Федерация, [sc 56168455000](https://orcid.org/0000-0002-9699-7145), <https://orcid.org/0000-0002-9699-7145>, maxsoftpage@yandex.ru

Жарков Александр Александрович — кандидат технических наук, доцент, доцент, Национальный исследовательский университет «МЭИ», Москва, 111250, Российская Федерация; главный конструктор, ООО «НПФ ВЕКТОР», Москва, 111396, Российская Федерация, [sc 57112011700](https://orcid.org/0000-0001-5628-8101), <https://orcid.org/0000-0001-5628-8101>, zarckov@mail.ru

Чен Хао — кандидат технических наук, старший инженер-исследователь, China University of Mining & Technology, Сюйчжоу, 100083, Китай, <https://orcid.org/0000-0001-9056-3760>, hchen@cumt.edu.cn

Дианов Антон Николаевич — кандидат технических наук, старший инженер-исследователь, Daeyoung R&D Center, Сеул, 03140, Южная Корея, [sc 24483283500](https://orcid.org/0000-0002-9778-6817), <https://orcid.org/0000-0002-9778-6817>, anton.dianov@ieee.org

Authors

Alecksey S. Anuchin — D.Sc., Professor, Head of Department, National Research University “Moscow Power Engineering Institute” (MPEI), Moscow, 111250, Russian Federation, [sc 56168843400](https://orcid.org/0000-0002-1019-756X), <https://orcid.org/0000-0002-1019-756X>, Anuchin.alecksey@gmail.com

Maria A. Gulyaeva — Assistant, National Research University “Moscow Power Engineering Institute” (MPEI), Moscow, 111250, Russian Federation, [sc 57195258176](https://orcid.org/0000-0002-2326-1153), <https://orcid.org/0000-0002-2326-1153>, tilgivelse@yandex.ru

Maxim M. Lashkevich — PhD, Leading Software Engineer, R&D Company “Vector”, LLC, Moscow, 111396, Russian Federation, [sc 56168455000](https://orcid.org/0000-0002-9699-7145), <https://orcid.org/0000-0002-9699-7145>, maxsoftpage@yandex.ru

Alexandr A. Zharkov — PhD, Associate Professor, Associate Professor, National Research University “Moscow Power Engineering Institute” (MPEI), Moscow, 111250, Russian Federation; Chief Designer, R&D Company “Vector”, LLC, Moscow, 111396, Russian Federation, [sc 57112011700](https://orcid.org/0000-0001-5628-8101), <https://orcid.org/0000-0001-5628-8101>, zarckov@mail.ru

Hao Chen — PhD, Senior Research Engineer, China University of Mining & Technology, Xuzhou, 100083, China, <https://orcid.org/0000-0001-9056-3760>, hchen@cumt.edu.cn

Anton N. Dianov — PhD, Senior Research Engineer, Daeyoung R&D Center, Seoul, 03140, South Korea, [sc 24483283500](https://orcid.org/0000-0002-9778-6817), <https://orcid.org/0000-0002-9778-6817>, anton.dianov@ieee.org

doi: 10.17586/2226-1494-2023-23-1-202-209

УДК 536.4

Моделирование диффузионных процессов при электротермической обработке реакционных тиглей системы железо–олово

Владислав Евгеньевич Фомин¹✉, Анастасия Сергеевна Тукмакова²,
Геннадий Александрович Болкунов³, Анна Владимировна Новотельнова⁴,
Федор Юрьевич Бочканов⁵, Дмитрий Юрьевич Карпенков⁶

^{1,2,3,4} Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация

^{5,6} Национальный исследовательский технологический университет «МИСиС», Москва, 119049, Российская Федерация

⁶ Московский государственный университет имени М.В. Ломоносова, Москва, 119991, Российская Федерация

¹ vladdisslav.fomin@yandex.ru✉, <https://orcid.org/0000-0002-8970-0809>

² astukmakova@itmo.ru, <https://orcid.org/0000-0002-9123-8983>

³ genabolkun@gmail.com, <https://orcid.org/0000-0003-3297-3395>

⁴ novotelnova@yandex.ru, <https://orcid.org/0000-0003-0073-2415>

⁵ Bochkanov_fy@mail.ru, <https://orcid.org/0000-0001-9258-2551>

⁶ karpenkov_d_y@mail.ru, <https://orcid.org/0000-0001-8686-2303>

Аннотация

Предмет исследования. Методами численного моделирования исследована закономерность диффузионных процессов в реакционных тиглях системы железо–олово при их электротермической обработке. Изучено влияние плотности тока и температуры на процессы тепло- и массопереноса в реакционной зоне. **Метод.** Численное моделирование выполнено методом конечных элементов. Разработанная модель описывает механические, тепловые, электрические и химические процессы при электротермической обработке системы железо–олово в реакционном тигле с учетом распределения компонентов при различных условиях обработки реакционного тигля. **Основные результаты.** Выполнен сравнительный анализ расчетных данных по диффузии олова в железо в условиях длительного воздействия высоких температур без приложения электрического напряжения и при нагреве реакционной зоны за счет пропускания электрического тока высокой плотности. Получена картина распределения массовых долей компонентов в зависимости от вида воздействия. Определена глубина проникновения взаимодействующих компонентов и выполнена оценка интенсивности протекания массообменных процессов. Установлены закономерности тепло-массопереноса в системе железо–олово при изменении начальных параметров процесса. Верификация модели проведена путем сопоставления результатов моделирования с данными натурных экспериментов на контрольных образцах. **Практическая значимость.** Результаты исследований могут быть использованы при прогнозировании условий получения новых функциональных материалов.

Ключевые слова

компьютерное моделирование, теплопроводность, электропроводность, диффузия, тепловые и электрические поля

Ссылка для цитирования: Фомин В.Е., Тукмакова А.С., Болкунов Г.А., Новотельнова А.В., Бочканов Ф.Ю., Карпенков Д.Ю. Моделирование диффузионных процессов при электротермической обработке реакционных тиглей системы железо–олово // Научно-технический вестник информационных технологий, механики и оптики. 2023. Т. 23, № 1. С. 202–209. doi: 10.17586/2226-1494-2023-23-1-202-209

Simulation of diffusion processes during electrothermal treatment of reaction crucibles of the Fe-Sn system

Vladislav E. Fomin¹✉, Anastasiia S. Tukmakova², Gennady A. Bolkunov³,
Anna V. Novotelnova⁴, Fedor Yu. Bochkanov⁵, Dmitry Yu. Karpenkov⁶

^{1,2,3,4} ITMO University, Saint Petersburg, 197101, Russian Federation

^{5,6} National University of Science and Technology MISIS, Moscow, 119049, Russian Federation

⁶ Lomonosov Moscow State University, Moscow, 119991, Russian Federation

¹ vladdisslav.fomin@yandex.ru✉, <https://orcid.org/0000-0002-8970-0809>

² astukmakova@itmo.ru, <https://orcid.org/0000-0002-9123-8983>

³ genabolkun@gmail.com, <https://orcid.org/0000-0003-3297-3395>

⁴ novotelnova@yandex.ru, <https://orcid.org/0000-0003-0073-2415>

⁵ Bochkanov_fy@mail.ru, <https://orcid.org/0000-0001-9258-2551>

⁶ karpenkov_d_y@mail.ru, <https://orcid.org/0000-0001-8686-2303>

Abstract

The diffusion processes regularity in the reaction crucibles of the iron-tin system during their electrothermal treatment was studied by the numerical simulation methods. The effect of current density and temperature on the processes of heat and mass transfer in the reaction zone has been studied. Numerical simulation was performed by the finite element method. The developed model includes mechanical, thermal, electrical and chemical processes during the electrothermal treatment of the iron-tin system in the reaction crucible, taking into account the distribution of components under various processing conditions of the reaction crucible. A comparative analysis of the calculated data on the diffusion of tin into iron under conditions of long-term exposure to high temperatures without the application of an electric voltage and when the reaction zone is heated by passing a high-density electric current is performed. A picture of the distribution of mass fractions of components depending on the type of impact is obtained. The penetration depth of the interacting components was determined and the intensity of the mass transfer processes was assessed. The regularities of heat and mass transfer in the system of iron and tin with a change of the process initial parameters are established. The model was verified by comparing the simulation results with the data of full-scale experiments on control samples. The research results can be used to predict the conditions for obtaining new functional materials.

Keywords

computer simulation, thermal conductivity, electrical conductivity, diffusion, thermal and electric fields

For citation: Fomin V.E., Tukmakova A.S., Bolkunov G.A., Novotelnova A.V., Bochkanov F.Yu., Karpenkov D.Yu. Simulation of diffusion processes during electrothermal treatment of reaction crucibles of the Fe-Sn system. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 2023, vol. 23, no. 1, pp. 202–209 (in Russian). doi: 10.17586/2226-1494-2023-23-1-202-209

Введение

В настоящее время ведется поиск новых материалов для твердотельных преобразователей электрической энергии и технических решений, основанных на альтернативных эффектах (например, аномальные термомагнитные эффекты, нелинейные термоэлектрические явления) [1, 2].

Возникновение исследовательского интереса к этой системе обусловлено обнаружением у ряда соединений высоких значений аномальных эффектов Холла и Нернста, а также наличием фаз, перспективных для использования в качестве постоянных магнитов [3].

Система железо–олово имеет большое количество интерметаллических соединений перспективных для использования, как в качестве термоэлектрических материалов, так и для изготовления магнитотвердых материалов [4].

Поиск новых фаз и разработка технологий их получения может привести к получению высокоэффективных магнитных устройств и преобразователей энергии, использующих в своей работе поперечный термомагнитный эффект Нернста.

Для эффективного поиска метастабильных фаз со структурой и свойствами, отличающимися от полученных традиционными методами сплавления компонентов, применяются методы комбинаторного ана-

лиза систем. Подвод тепла к реакционной зоне может быть организован различными методами, такими как инициирование реакции горения при самораспространяющемся высокотемпературном синтезе или прямым нагревом реакционной зоны электрическим током [5, 6].

В поисковых исследованиях одним из методов получения новых фаз является техника реакционных тиглей [7]. Процесс осуществляется путем пропускания электрического тока через находящийся в замкнутой форме образец [8]. В ходе электротермической обработки температура реакционной зоны достигает значений, при которых содержимое тигля переходит в жидкое состояние. При этом материал, из которого изготовлен тигель, остается в твердом состоянии. На границе между расплавом и телом тигля происходит твердо-жидкофазная реакция. Полученный градиент концентрации ускоряет формирование соединений [9]. Данный процесс может быть модифицирован путем применения электрического тока высокой плотности. При этом фазообразование в диффузионной зоне тигля сопровождается такими эффектами как тепловое расширение, диффузионные процессы (концентрационная диффузия, бародиффузия, термодиффузия и электродиффузия ионов) [10, 11]. Условия, при которых проходит обработка, в этом случае отличны от условий формирования фаз путем сплавления реагентов в заданной пропорции.

Температура обработки — один из важнейших параметров, оказывающих влияние на формирование фаз. Значение температуры в реакционной зоне зависит от условий электрического нагружения. Требуемые значения разности потенциалов, а также значения получаемых температур и данные по массообмену могут быть получены путем математического моделирования.

Цель данной работы — изучение процессов массопереноса в объеме реакционной установки при обработках системы прямым нагревом и электрическим током высокой плотности.

Моделирование процесса

Геометрия модели. Установка для проведения синтеза (рис. 1) представляет собой вакуумную камеру со стальными стенками. Внутри камеры между стальными электродами 1, 6 (рис. 1) размещен закрепленный графитовыми вставками реакционный тигель 3 цилиндрической формы. Тигель заполнен компонентами исследуемой фазовой диаграммы (образец, выполненный из олова 4). Отвод тепла от камеры происходит с помощью водяного охлаждения.

Метод и материалы. Процесс моделирования произведен в реакционном тигле, выполненном из же-

леза. В тигель помещено олово в твердом состоянии. Электрическое нагружение осуществляется путем подачи электрического напряжения между верхним и нижним электродами. Под воздействием приложенного электрического напряжения через реакционный тигель протекает электрический ток. За счет выделения джоулева тепла происходит нагрев тигля и плавление олова. Этот процесс сопровождается интердиффузией железа и олова. Диффузионный процесс стимулируется тепловым воздействием, однако немалую роль в процессе играет электромиграция ионов [12].

Свойства материалов, используемых в процессе, приведены в табл. 1.

Для описания температурных зависимостей коэффициентов тепло- и электропроводности использованы данные [13–19]. Значения коэффициентов диффузии железа и олова, и температурные зависимости взяты из работ [20, 21].

Математическая модель процесса

Определим следующие переменные: поле температуры, плотность тока и концентрацию компонентов. Получим уравнения, использованные для описания физических процессов при реакционном синтезе.

Для описания электрических процессов применим закон Ома:

$$\mathbf{J} = \sigma_e \mathbf{E},$$

где σ_e — коэффициент электропроводности; \mathbf{E} — вектор напряженности электрического поля; \mathbf{J} — вектор плотности тока.

Закон сохранения электрического заряда:

$$\text{div} \mathbf{J} = 0.$$

Для описания тепломассообменных процессов используем:

— уравнение теплового баланса:

$$\rho C_p \frac{\partial T}{\partial t} + \text{div} \mathbf{q} = Q,$$

где ρ — плотность; C_p — удельная изобарная теплоемкость; t — время; Q — источник джоулева тепла; T — абсолютная температура; \mathbf{q} — вектор плотности теплового потока;

— закон Фурье:

$$\mathbf{q} = -\kappa \nabla T,$$

где κ — коэффициент теплопроводности;

— закон сохранения массы для описания диффузионных процессов, последнее слагаемое которого учитывает электромиграцию ионов:

$$\mathbf{J}_d = -D(T, p) \nabla c - z u F c,$$

где \mathbf{J}_d — диффузионный поток; $D(T, p)$ — коэффициент диффузии, как функция от температуры и давления p ; c — молярная концентрация компонента; z — зарядовое число; u — подвижность ионов; F — константа Фарадея равная 96485,33 Кл/моль.

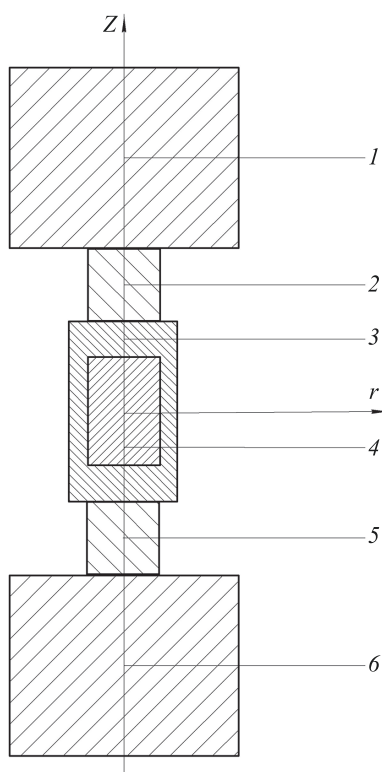


Рис. 1. Геометрическая модель установки: 1, 6 — верхний и нижний стальные электроды; 2 — графитовый пуансон; 3 — реакционный тигель; 4 — реакционная зона, заполненная оловом; 5 — пуансоны, выполненные из графита; Z — вертикальная ось; r — горизонтальная ось

Fig. 1. Geometrical model of the installation: 1, 6 — upper and lower steel electrodes; 2 — punch, made of graphite; 3 — reaction crucible; 4 — reaction zone filled with tin; 5 — punches made of graphite; Z — vertical axis; r — horizontal axis

Таблица 1. Свойства материалов, входящих в состав установки

Table 1. Properties of materials used in the installation

Свойство	Материал			
	Железо	Олово	Графит	Сталь
Плотность ρ , кг/м ³	7850 [11]	7310 [11]	1870 [12]	7740 [14]
Удельная изобарная теплоемкость C_p , Дж/(кг·К)	449 [13]	230 [13]	714 [15]	4620 [14]
Коэффициент теплового расширения $\alpha \cdot 10^{-6}$, 1/К	11,5 [11]	23,4 [11]	4 [16]	14 [14]
Модуль Юнга E_Y , ГПа	212 [11]	48,3 [11]	10 [12]	200 [14]

Для описания механических процессов с учетом теплового расширения применим закон Гука:

$$\sigma = E_Y \left(\frac{\Delta l}{l} + \alpha(T - T_0) \right),$$

где σ — нормальное напряжение в поперечном сечении; E_Y — модуль Юнга; $\frac{\Delta l}{l}$ — относительное удлинение; α — коэффициент теплового расширения; T_0 — начальная температура.

Граничные условия тепловых процессов

Примем, что начальная температура T_0 всех элементов установки равна температуре окружающей среды $T_{amb} = 293$ К. Во время проведения процесса синтеза система водяного охлаждения поддерживает температуру верхней поверхности верхнего электрода и нижней поверхности нижнего электрода, равную температуре окружающей среды.

На боковой поверхности деталей установки применено граничное условие лучистого теплообмена, описываемое законом Стефана–Больцмана:

$$-\mathbf{nq} = \varepsilon \sigma_{CB} (T_{amb}^4 - T_1^4),$$

где σ_{CB} — константа Стефана–Больцмана; ε — коэффициент черноты; T_{amb} — температура окружающей среды; T_1 — температура поверхности тела; \mathbf{n} — вектор нормали к поверхности.

Значение электрического потенциала ϕ_1 торца нижнего электрода принято равным нулю. Для создания электрического нагружения к верхней границе верхнего электрода необходимо приложить электрический потенциал ϕ_2 . Электрическое напряжение на установке, действующее в процессе синтеза, равно $U = \phi_2 - \phi_1$. Внешние боковые поверхности установки электроизолированы.

При задании граничных условий механических процессов принято, что нижняя поверхность установки механически жестко зафиксирована. К торцу верхнего электрода приложено давление 1 МПа. Все остальные поверхности могут смещаться свободно.

В качестве граничных условий при описании диффузионных процессов известно, что содержание железа перед началом процесса в тигле составило 100 %, а внутренний объем тигля полностью заполнен оловом, не содержащим примесей.

Основные результаты моделирования

Исследован процесс электропереноса в установке с геометрическими характеристиками (табл. 2).

Рассмотрен диффузионный процесс в двух режимах: под воздействием тепла без приложения электрического напряжения и в условиях нагрева реакционной зоны джоулевым теплом.

В первом случае при заданной температуре разность потенциалов между верхним и нижним электродами равнялась нулю. Температура всех деталей установки — 1073 К. Данное значение — температура проведения основных натурных экспериментов с целью получения перспективных материалов [6].

Во втором случае к верхнему электроду приложен электрический потенциал, при этом нижний электрод заземлен. Электрический ток, полученный под действием электрического напряжения U , выполнил нагрев всех деталей установки до определенной температуры. Для выявления влияния напряжения на распределение температуры и массообменные процессы использованы значения U равные 0,3; 0,7; 1; 1,4 В.

Под действием электрического тока, протекающего через детали установки, происходит разогрев реакционной зоны джоулевым теплом. При этом процессы диффузии значительно ускоряются за счет вклада от электромиграции.

Разогрев реакционной зоны происходит за первые минуты (рис. 2). Далее температура реакционной зоны стабилизируется и поддерживается практически постоянной. Это связано с формированием теплового равновесия между джоулевым нагревом и охлаждением путем теплопередачи и излучения.

Средние значения плотности тока в радиальном сечении и достигаемая температура при приложении различных напряжений приведены в табл. 3.

Таблица 2. Геометрические характеристики модели

Table 2. Geometry of the model

Элементы модели	d , мм	h , мм
Верхний и нижний стальные электроды	19,2	15,0
Графитовые пуансоны	6,0	6,0
Реакционный тигель	10,0	15,0
Внутренний объем тигля, заполненный оловом	6,0	9,0

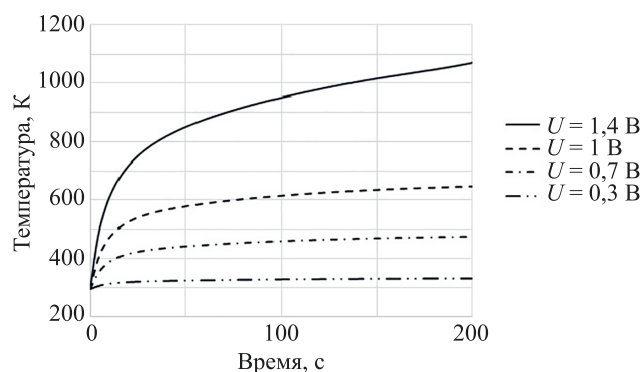


Рис. 2. Изменение средней температуры реакционной зоны от времени при различных значениях напряжения

Fig. 2. Change in the average temperature of the reaction zone versus time at various voltage values

Таблица 3. Значения плотности тока при заданных значениях разности потенциалов

Table 3. Values of current density for given values of potential difference

$U, \text{В}$	Плотность тока $J, \text{МА/м}^2$	$T, \text{К}$
0,3	0,6	330
0,7	1,2	469
1	1,5	632
1,4	2,7	1073

Выполним формирование диффузионного слоя при воздействии электрического тока при напряжении 1,4 В, так как в этом случае в реакционной зоне достигается необходимая температура $T = 1073 \text{ К}$.

Проведем анализ изменения концентрации олова в железном тигле при различных условиях синтеза. Рассчитаем изменения эффективной толщины диффузионного слоя в тигле d_{eff} (слоя с содержанием олова

40 %) от длительности процесса. Получим, что эффективность диффузии повышается при воздействии электрического напряжения (рис. 3). Рассчитанное по закону Фика значение d_{eff} увеличивается со временем процесса (рис. 3, а). Изменение массовой доли олова в приграничных слоях вблизи границы раздела железо–олово (Fe–Sn) показано на рис. 3, б. Значение $L = 0$ соответствует положению границы раздела.

Верификация модели

Верификация модели осуществлена путем сопоставления данных моделирования и натурных экспериментов, проведенных на контрольных образцах. В ходе экспериментов выполнен синтез образцов в течение фиксированных промежутков времени — 3, 24, 120 ч. После окончания каждого из экспериментов, полученный контрольный образец был разрезан и проведен анализ содержимого методами сканирующей электронной микроскопии и энергодисперсионной рентгеновской спектроскопии. Данные по глубине диффузии олова в железо сопоставлялись с концентрационными профилями, полученными при моделировании с аналогичными условиями.

Отметим, что при реакции в объеме реакционного тигля происходят несколько процессов: интердиффузия компонентов тигля на границе расплава и стенок тигля, а также растворение стенок тигля в расплаве.

Скорость растворения материала тигля определена экспериментально. После завершения процесса обработки изучены образцы с помощью микроскопа и проведены измерения расстояния сдвига границы. Процедуры измерения произведены несколько раз после обработок температурой и током. Таким образом, была измерена скорость растворения материала тигля в олове (рис. 4).

Скорость растворения материала тигля непостоянна и значительно увеличивается при воздействии электрического напряжения. Отмечена неравномерность ско-

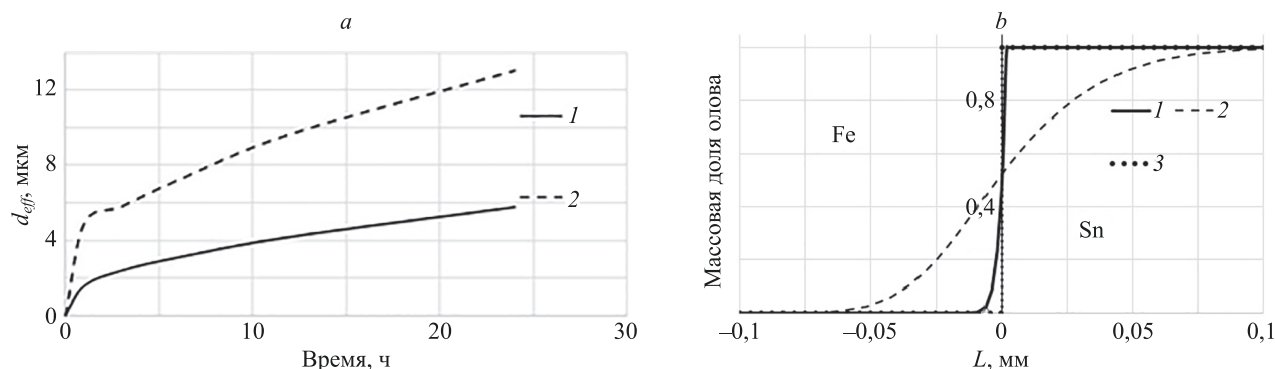


Рис. 3. Изменение эффективной толщины диффузионного слоя d_{eff} в тигле (а) и массовой доли олова в приграничных слоях вблизи нижней границы реакционной зоны у дна тигля от длительности процесса (б) при различных условиях: при высокой температуре без приложения электрического напряжения $U = 0 \text{ В}$ (кривая 1); при нагреве реакционной зоны электрическим током ($U = 1,4 \text{ В}$) (кривая 2); распределение компонентов в начале процесса (кривая 3). Температура $T = 1073 \text{ К}$

Fig. 3. Change in the effective thickness of the diffusion layer d_{eff} in the crucible (а) and the change in the tin mass fraction in the boundary layers near the lower boundary of the reaction zone at the bottom of the crucible vs the duration of the process (б) under different conditions: at high temperature without application of electric voltage $U = 0 \text{ V}$ (curve 1); when heating the reaction zone with electric current ($U = 1,4 \text{ V}$) (curve 2); distribution of components at the beginning of the process (curve 3). Temperature $T = 1073 \text{ К}$

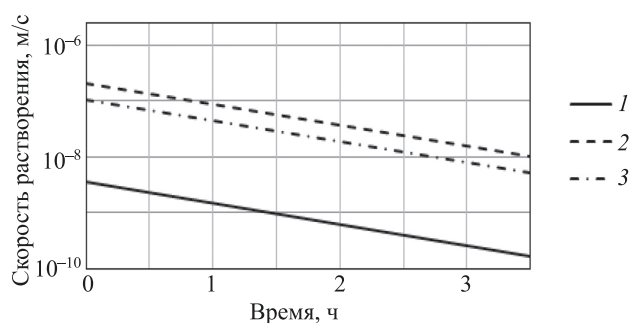


Рис. 4. Изменение скорости растворения материала тигля в ходе процесса синтеза при различных условиях проведения синтеза: при высокой температуре без приложения электрического напряжения ($U = 0$ В) (кривая 1); при нагреве реакционной зоны электрическим током ($U = 1,4$ В) у дна и боковой стенки тигля (кривые 2 и 3). Температура $T = 1073$ К

Fig. 4. Change in the dissolution rate of the crucible material during the synthesis process under various synthesis conditions: at high temperature without applying electrical voltage ($U = 0$ V) (curve 1); when heating the reaction zone with electric current ($U = 1.4$ V) at the bottom and side wall of the crucible (curves 2 and 3). Temperature $T = 1073$ K

рости растворения железа в объеме реакционной зоны. Так, скорость растворения дна тигля при напряжении 1,4 В примерно в два раза выше скорости растворения боковой стенки при том же напряжении. Это говорит о направленности массообменных процессов и показывает, что растворение и диффузия дна проходят более интенсивно, чем у стенок.

Сравнение результатов эксперимента и моделирования представлены в табл. 4.

Сопоставление результатов данных сканирующей электронной микроскопии и энергодисперсионной рентгеновской спектроскопии с данными моделирования позволило удостовериться в обоснованности полученных результатов. Расхождение в оценках глубины проникновения олова в железо натуральных и модельных экспериментов не превышало 4 %.

Обсуждение результатов

Условия проведения обработки и способ разогрева реакционной зоны влияют на скорость диффузионных процессов. При проведении процесса без воздействия

электрического напряжения происходит процесс диффузии железа в расплавленное олово и твердотельная диффузия олова в приграничные слои железа.

Толщина эффективного диффузионного слоя медленно увеличивается со временем, при этом скорость ее роста постепенно уменьшается. После 24 ч процесса значение эффективной толщины диффузионного слоя при синтезе в условиях воздействия высокой температуры без пропуска электрического тока достигает 6 мкм.

Толщина диффузионного слоя, получаемая при проведении экспериментов с нагревом реакционной зоны джоулевым теплом, увеличивается по сравнению со значением, полученным в ходе эксперимента при нагреве без приложения напряжения. Процесс в этом случае сопровождается увеличением толщины эффективного диффузионного слоя и повышением скорости растворения железа в олове. При воздействии электрическим током, величина d_{eff} повысилась до 13 мкм.

Наиболее интенсивно изменение массовой доли олова происходит вблизи границы реакционной зоны. Массовая доля олова внутри тигля постепенно убывает, в то же время происходит диффузия олова в приграничные слои железа и растворение приграничных слоев железа в олове. Изменение наклона кривой (рис. 2, б) в случае нагрева зоны электрическим током говорит об увеличении скорости массообменных процессов и ускорении диффузионных процессов, вызванных влиянием электропереноса/электродиффузии.

Данные, полученные методом математического моделирования, коррелируют с результатами натуральных экспериментов для образцов, синтезированных в аналогичных условиях. Уменьшение скорости роста толщины эффективного диффузионного слоя со временем связано с постепенным увеличением толщины слоя интерметаллида на границе расплава.

По мере увеличения толщины слоя, все процессы диффузии замедляются, за счет меньших коэффициентов диффузии в интерметаллиде и изменением характера диффузии с жидко-твердофазного к твердофазному. Этот процесс также сопровождается изменением концентраций, а значит, уменьшением разницы химических потенциалов на границе раздела [22]. Как итог, максимальная скорость растворения материала тигля достигается в самом начале процесса и затем снижается.

Таблица 4. Значения глубины проникновения олова и температуры реакционной зоны по результатам моделирования и эксперимента

Table 4. Values of the tin penetration depth and the temperature of the reaction zone according to the results of modeling and to the results of the experiment

Измеряемая величина	Результат эксперимента	Результат моделирования	Отклонение результатов, %
Глубина проникновения олова после трех часов выдержки при температуре 1073 К, мкм	7,9	7,6	3,8
Глубина проникновения после трех часов воздействия постоянного электрического тока напряжением 1,4 В, мкм	20,8	20,3	2,4
Температура поверхности реакционной зоны, К	1074	1073	0,09

Закключение

В результате проведенных исследований разработана математическая модель, описывающая электротермическую обработку реакционного тигля системы железо–олово. Верификация модели осуществлена путем сравнения полученных результатов по глубине проникновения олова в железо и по температуре с результатами экспериментов различной обработки. Показано, что компьютерное моделирование методом конечных элементов способно с достаточной точностью описывать процессы электродиффузии в системе реакционного тигля.

Методом математического моделирования исследовано влияние параметров обработки на поле температуры в реакционной зоне и проведен анализ диффузионных процессов железа и олова в реакционном тигле при

различных условиях проведения обработки. Рассчитана глубина проникновения олова в железный тигель под влиянием электрического тока и без него.

При синтезе в условиях электрического нагружения возрастают скорость диффузии и толщина эффективного слоя. Электрический ток оказывает решающее влияние на глубину проникновения атомов, а также значительно ускоряет все процессы диффузии и растворения внутри реакционного тигля.

Модель может быть использована при создании рекомендаций по выбору начальных параметров проведения синтеза интерметаллидов железа и олова.

Разработанная модель может быть распространена на анализ процессов электротермических обработок других систем, что поможет ускорить подбор начальных условий протекания процессов.

Литература

- Li X., Xu L., Ding L., Wang J., Shen M., Lu X., Zhu Z., Behnia K. Anomalous Nernst and Righi-Leduc effects in Mn_3Sn : Berry curvature and entropy flow // *Physical Review Letters*. 2017. V. 119. N 5. P. 056601. <https://doi.org/10.1103/PhysRevLett.119.056601>
- Буллат Л.П., Неведова И.А. О нелинейных термоэлектрических явлениях // Вестник Международной академии холода. 2012. № 4. С. 54–56.
- Sales B.C., Saporov B., McGuire M.A., Singh D.J., Parker D.S. Ferromagnetism of Fe_3Sn and alloys // *Scientific Reports*. 2014. V. 4. N 1. P. 7024. <https://doi.org/10.1038/srep07024>
- Predel B. Fe-Sn (Iron-Tin) // *Dy-Er-Fr-Mo*. 1995. P. 1–5. https://doi.org/10.1007/10474837_1342
- Levashov E.A., Mukasyan A.S., Rogachev A.S., Shtansky D.V. Self-propagating high-temperature synthesis of advanced materials and coatings // *International Materials Reviews*. 2017. V. 62. N 4. P. 203–239. <https://doi.org/10.1080/09506608.2016.1243291>
- Orrù R., Licheri R., Locci A.M., Cincotti A., Cao G. Consolidation/synthesis of materials by electric current activated/assisted sintering // *Materials Science and Engineering: R: Reports*. 2009. V. 63. N 4-6. P. 127–287. <https://doi.org/10.1016/j.mser.2008.09.003>
- Nikbakht R., Assadi H. Phase-field modelling of self-propagating high-temperature synthesis of NiAl // *Acta Materialia*. 2012. V. 60. N 10. P. 4041–4053. <https://doi.org/10.1016/j.actamat.2012.04.017>
- Lin S., Yeh C., Xie W., Liu Y., Yoshimura M. Ab initio-aided CALPHAD thermodynamic modeling of the Sn-Pb binary system under current stressing // *Scientific Reports*. 2013. V. 3. N 1. P. 2731. <https://doi.org/10.1038/srep02731>
- Fayyazi B., Skokov K.P., Faske T., Karpenkov D.Y., Donner W., Gutfleisch O. Bulk combinatorial analysis for searching new rare-earth free permanent magnets: Reactive crucible melting applied to the Fe-Sn binary system // *Acta Materialia*. 2017. V. 141. P. 434–443. <https://doi.org/10.1016/j.actamat.2017.09.036>
- Chen C.-M., Chen S.-W. Electromigration effect upon the Zn/Ni and Bi/Ni interfacial reactions // *Journal of Electronic Materials*. 2000. V. 29. N 10. P. 1222–1228. <https://doi.org/10.1007/s11664-000-0016-5>
- Pierce D.G., Brusius P.G. Electromigration: A review // *Microelectronics Reliability*. 1997. V. 37. N 7. P. 1053–1072. [https://doi.org/10.1016/s0026-2714\(96\)00268-5](https://doi.org/10.1016/s0026-2714(96)00268-5)
- Goll D., Loeffler R., Hohns D., Schneider G. Reaction sintering as a high-throughput approach for magnetic materials development // *Scripta Materialia*. 2018. V. 146. P. 355–361. <https://doi.org/10.1016/j.scriptamat.2017.05.004>
- Buch A. *Pure Metals Properties: A Scientific and Technical Handbook*. ASM International, 1999. 306 p.
- Iwashita N., Imagawa H., Nishiumi W. Variation of temperature dependence of electrical resistivity with crystal structure of artificial graphite products // *Carbon*. 2013. V. 61. P. 602–608. <https://doi.org/10.1016/j.carbon.2013.05.042>

References

- Li X., Xu L., Ding L., Wang J., Shen M., Lu X., Zhu Z., Behnia K. Anomalous Nernst and Righi-Leduc effects in Mn_3Sn : Berry curvature and entropy flow. *Physical Review Letters*, 2017, vol. 119, no. 5, pp. 056601. <https://doi.org/10.1103/PhysRevLett.119.056601>
- Bulat L.P., Nefedova I.A. Nonlinear thermoelectric phenomena. *Journal of International Academy of Refrigeration*, 2012, no. 4, pp. 54–56. (in Russian)
- Sales B.C., Saporov B., McGuire M.A., Singh D.J., Parker D.S. Ferromagnetism of Fe_3Sn and alloys. *Scientific Reports*, 2014, vol. 4, no. 1, pp. 7024. <https://doi.org/10.1038/srep07024>
- Predel B. Fe-Sn (Iron-Tin). *Dy-Er-Fr-Mo*, 1995, pp. 1–5. https://doi.org/10.1007/10474837_1342
- Levashov E.A., Mukasyan A.S., Rogachev A.S., Shtansky D.V. Self-propagating high-temperature synthesis of advanced materials and coatings. *International Materials Reviews*, 2017, vol. 62, no. 4, pp. 203–239. <https://doi.org/10.1080/09506608.2016.1243291>
- Orrù R., Licheri R., Locci A.M., Cincotti A., Cao G. Consolidation/synthesis of materials by electric current activated/assisted sintering. *Materials Science and Engineering: R: Reports*, 2009, vol. 63, no. 4-6, pp. 127–287. <https://doi.org/10.1016/j.mser.2008.09.003>
- Nikbakht R., Assadi H. Phase-field modelling of self-propagating high-temperature synthesis of NiAl. *Acta Materialia*, 2012, vol. 60, no. 10, pp. 4041–4053. <https://doi.org/10.1016/j.actamat.2012.04.017>
- Lin S., Yeh C., Xie W., Liu Y., Yoshimura M. Ab initio-aided CALPHAD thermodynamic modeling of the Sn-Pb binary system under current stressing. *Scientific Reports*, 2013, vol. 3, no. 1, pp. 2731. <https://doi.org/10.1038/srep02731>
- Fayyazi B., Skokov K.P., Faske T., Karpenkov D.Y., Donner W., Gutfleisch O. Bulk combinatorial analysis for searching new rare-earth free permanent magnets: Reactive crucible melting applied to the Fe-Sn binary system. *Acta Materialia*, 2017, vol. 141, pp. 434–443. <https://doi.org/10.1016/j.actamat.2017.09.036>
- Chen C.-M., Chen S.-W. Electromigration effect upon the Zn/Ni and Bi/Ni interfacial reactions. *Journal of Electronic Materials*, 2000, vol. 29, no. 10, pp. 1222–1228. <https://doi.org/10.1007/s11664-000-0016-5>
- Pierce D.G., Brusius P.G. Electromigration: A review. *Microelectronics Reliability*, 1997, vol. 37, no. 7, pp. 1053–1072. [https://doi.org/10.1016/s0026-2714\(96\)00268-5](https://doi.org/10.1016/s0026-2714(96)00268-5)
- Goll D., Loeffler R., Hohns D., Schneider G. Reaction sintering as a high-throughput approach for magnetic materials development. *Scripta Materialia*, 2018, vol. 146, pp. 355–361. <https://doi.org/10.1016/j.scriptamat.2017.05.004>
- Buch A. *Pure Metals Properties: A Scientific and Technical Handbook*. ASM International, 1999, 306 p.
- Iwashita N., Imagawa H., Nishiumi W. Variation of temperature dependence of electrical resistivity with crystal structure of artificial graphite products. *Carbon*, 2013, vol. 61, pp. 602–608. <https://doi.org/10.1016/j.carbon.2013.05.042>

15. Patel A.B., Bhatt N.K., Thakore B.Y., Vyas P.R., Jani A.R. The temperature-dependent electrical transport properties of liquid Sn using pseudopotential theory // *Molecular Physics*. 2014. V. 112. N 15. P. 2000–2004. <https://doi.org/10.1080/00268976.2013.877169>
16. Klemens P.G., Pedraza D.F. Thermal conductivity of graphite in the basal plane // *Carbon*. 1994. V. 32. N 4. P. 735–741. [https://doi.org/10.1016/0008-6223\(94\)90096-5](https://doi.org/10.1016/0008-6223(94)90096-5)
17. Eiling A., Schilling J.S. Pressure and temperature dependence of electrical resistivity of Pb and Sn from 1-300K and 0-10 GPa-use as continuous resistive pressure monitor accurate over wide temperature range; superconductivity under pressure in Pb, Sn and In // *Journal of Physics F: Metal Physics*. 1981. V. 11. N 3. P. 623–639. <https://doi.org/10.1088/0305-4608/11/3/010>
18. Chapman T.W. The heat capacity of liquid metals // *Materials Science and Engineering*. 1966. V. 1. N 1. P. 65–69. [https://doi.org/10.1016/0025-5416\(66\)90012-7](https://doi.org/10.1016/0025-5416(66)90012-7)
19. Taylor G.R., Isin A., Coleman R.V. Resistivity of iron as a function of temperature and magnetization // *Physical Review*. 1968. V. 165. N 2. P. 621–631. <https://doi.org/10.1103/physrev.165.621>
20. Torres D.N., Perez R.A., Dymont F. Diffusion of tin in α -iron // *Acta Materialia*. 2000. V. 48. N 11. P. 2925–2931. [https://doi.org/10.1016/S1359-6454\(00\)00074-4](https://doi.org/10.1016/S1359-6454(00)00074-4)
21. Neumann G., Tuijn C. Self-Diffusion and Impurity Diffusion in Pure Metals: *Handbook of Experimental Data*. Elsevier, 2011. 360 p.
22. Ishida T. The reaction of solid iron with molten tin // *Transactions of the Japan Institute of Metals*. 1973. V. 14. N 1. P. 37–44. <https://doi.org/10.2320/matertrans1960.14.37>
15. Patel A.B., Bhatt N.K., Thakore B.Y., Vyas P.R., Jani A.R. The temperature-dependent electrical transport properties of liquid Sn using pseudopotential theory. *Molecular Physics*, 2014, vol. 112, no. 15, pp. 2000–2004. <https://doi.org/10.1080/00268976.2013.877169>
16. Klemens P.G., Pedraza D.F. Thermal conductivity of graphite in the basal plane. *Carbon*, 1994, vol. 32, no. 4, pp. 735–741. [https://doi.org/10.1016/0008-6223\(94\)90096-5](https://doi.org/10.1016/0008-6223(94)90096-5)
17. Eiling A., Schilling J.S. Pressure and temperature dependence of electrical resistivity of Pb and Sn from 1-300K and 0-10 GPa-use as continuous resistive pressure monitor accurate over wide temperature range; superconductivity under pressure in Pb, Sn and In. *Journal of Physics F: Metal Physics*, 1981, vol. 11, no. 3, pp. 623–639. <https://doi.org/10.1088/0305-4608/11/3/010>
18. Chapman T.W. The heat capacity of liquid metals. *Materials Science and Engineering*, 1966, vol. 1, no. 1, pp. 65–69. [https://doi.org/10.1016/0025-5416\(66\)90012-7](https://doi.org/10.1016/0025-5416(66)90012-7)
19. Taylor G.R., Isin A., Coleman R.V. Resistivity of iron as a function of temperature and magnetization. *Physical Review*, 1968, vol. 165, no. 2, pp. 621–631. <https://doi.org/10.1103/physrev.165.621>
20. Torres D.N., Perez R.A., Dymont F. Diffusion of tin in α -iron. *Acta Materialia*, 2000, vol. 48, no. 11, pp. 2925–2931. [https://doi.org/10.1016/S1359-6454\(00\)00074-4](https://doi.org/10.1016/S1359-6454(00)00074-4)
21. Neumann G., Tuijn C. *Self-Diffusion and Impurity Diffusion in Pure Metals: Handbook of Experimental Data*. Elsevier, 2011, 360 p.
22. Ishida T. The reaction of solid iron with molten tin. *Transactions of the Japan Institute of Metals*, 1973, vol. 14, no. 1, pp. 37–44. <https://doi.org/10.2320/matertrans1960.14.37>

Авторы

Фомин Владислав Евгеньевич — аспирант, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, [sc 57202281009](https://orcid.org/0000-0002-8970-0809), <https://orcid.org/0000-0002-8970-0809>, vladdisslav.fomin@yandex.ru

Тукмакова Анастасия Сергеевна — кандидат технических наук, доцент, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, [sc 57103805600](https://orcid.org/0000-0002-9123-8983), <https://orcid.org/0000-0002-9123-8983>, astukmakova@itmo.ru

Болкунов Геннадий Александрович — аспирант, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, [sc 57202281009](https://orcid.org/0000-0003-3297-3395), <https://orcid.org/0000-0003-3297-3395>, genabolkun@gmail.com

Новотельнова Анна Владимировна — кандидат технических наук, доцент, доцент, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, [sc 56436689700](https://orcid.org/0000-0003-0073-2415), <https://orcid.org/0000-0003-0073-2415>, novotelnova@yandex.ru

Бочканов Федор Юрьевич — младший научный сотрудник, Национальный исследовательский технологический университет «МИСиС», Москва, 119049, Российская Федерация, [sc 36918372200](https://orcid.org/0000-0001-9258-2551), <https://orcid.org/0000-0001-9258-2551>, Bochkanov_fy@mail.ru

Карпенков Дмитрий Юрьевич — кандидат физико-математических наук, старший научный сотрудник, Национальный исследовательский технологический университет «МИСиС», Москва, 119049, Российская Федерация; старший научный сотрудник, Московский государственный университет имени М.В. Ломоносова, Москва, 119991, Российская Федерация, [sc 36918372200](https://orcid.org/0000-0001-8686-2303), <https://orcid.org/0000-0001-8686-2303>, karpenkov_d_y@mail.ru

Authors

Vladislav E. Fomin — PhD Student, ITMO University, Saint Petersburg, 197101, Russian Federation, [sc 57202281009](https://orcid.org/0000-0002-8970-0809), <https://orcid.org/0000-0002-8970-0809>, vladdisslav.fomin@yandex.ru

Anastasiia S. Tukmakova — PhD, Associate Professor, ITMO University, Saint Petersburg, 197101, Russian Federation, [sc 57103805600](https://orcid.org/0000-0002-9123-8983), <https://orcid.org/0000-0002-9123-8983>, astukmakova@itmo.ru

Gennady A. Bolkunov — PhD Student, ITMO University, Saint Petersburg, 197101, Russian Federation, [sc 57202281009](https://orcid.org/0000-0003-3297-3395), <https://orcid.org/0000-0003-3297-3395>, genabolkun@gmail.com

Anna V. Novotelnova — PhD, Associate Professor, Associate Professor, ITMO University, Saint Petersburg, 197101, Russian Federation, [sc 56436689700](https://orcid.org/0000-0003-0073-2415), <https://orcid.org/0000-0003-0073-2415>, novotelnova@yandex.ru

Fedor Yu. Bochkanov — Junior Researcher, National University of Science and Technology MISIS, Moscow, 119049, Russian Federation, [sc 36918372200](https://orcid.org/0000-0001-9258-2551), <https://orcid.org/0000-0001-9258-2551>, Bochkanov_fy@mail.ru

Dmitry Yu. Karpenkov — PhD (Physics & Mathematics), Senior Researcher, National University of Science and Technology MISIS, Moscow, 119049, Russian Federation; Senior Researcher, Lomonosov Moscow State University, Moscow, 119991, Russian Federation, [sc 36918372200](https://orcid.org/0000-0001-8686-2303), <https://orcid.org/0000-0001-8686-2303>, karpenkov_d_y@mail.ru

Статья поступила в редакцию 03.08.2022
Одобрена после рецензирования 07.12.2022
Принята к печати 16.01.2023

Received 03.08.2022
Approved after reviewing 07.12.2022
Accepted 16.01.2023



Работа доступна по лицензии
Creative Commons
«Attribution-NonCommercial»

Уважаемые подписчики научно-технической литературы!

Журнал выходит 6 раз в год.

На журнал «Научно-технический вестник информационных технологий, механики и оптики» можно оформить подписку в почтовых отделениях по объединенному каталогу «Пресса России», подписные индексы 47197 (полугодовая подписка),

а также online

по объединенному каталогу «Пресса России» и по каталогу «Пресса по подписке», подписные индексы Э47197 (полугодовая подписка) и по электронному каталогу Почты России, подписной индекс ПС543.

Корпоративная подписка и подписка физических лиц возможна по каталогу компаний «Урал-Пресс», подписные индексы 47197 (полугодовая подписка) и 70522 (годовая подписка).

Сведения о подписке можно уточнить в редакции журнала по адресу:

Санкт-Петербург, ул. Ломоносова., д.9, литера А, комн. 2136.

Тел.: +7(812) 480 02 75

ТРЕБОВАНИЯ К ОФОРМЛЕНИЮ СТАТЕЙ

Журнал публикует научные обзоры, полнотекстовые статьи и краткие сообщения.

Полнотекстовая статья должна иметь четкую структуру, включающую в себя аннотацию, ключевые слова, введение, несколько содержательных разделов и заключение.

В аннотации, рассчитанной на самый широкий круг читателей, необходимо в объеме 200–400 слов в форме краткого реферата (Abstract) изложить научное содержание статьи: предмет, цель работы, метод или методология проведения работы, краткое описание эксперимента, полученные результаты, рекомендации по их применению.

Ключевые слова должны отражать информацию, представленную в статье. Основное ключевое слово указывается первым.

Во введении необходимо представить содержательную постановку рассматриваемого вопроса, провести краткий анализ известных из научной литературы решений (со ссылками на источники), дать критику их недостатков, показать научную новизну и преимущество (особенности) предлагаемого подхода.

В основном тексте статьи должна быть представлена строгая постановка решаемой задачи, изложены и обстоятельно разъяснены (доказаны) полученные утверждения и выводы, приведены результаты экспериментальных исследований или математического моделирования, иллюстрирующие сделанные утверждения. Основной текст статьи должен быть разбит на содержательные разделы (2–3 раздела).

В заключении необходимо кратко сформулировать основные результаты, прокомментировать их и, если возможно, указать на направления дальнейших исследований и области применения.

Пристатейный список литературы (рекомендуется): для обзорной статьи — не менее 50, для полнотекстовой статьи — не менее 15, для краткого сообщения — не менее 8 литературных источников.

Ссылки на стандарты и иные нормативные документы, а также на неперсонифицированные интернет-ресурсы в список литературы не вносятся, а оформляются в виде сносок.

Объем обзорной статьи предварительно согласовывается с редакцией.

Объем полнотекстовой статьи, включая иллюстрации, таблицы и список литературы, не должен превышать 8 страниц машинописного текста (с рисунками и таблицами), шрифт 12 pt, один интервал.

Объем краткого сообщения — до 3 страниц, шрифт 12 pt, один интервал. Рубрикация текста не требуется. Статьи принимаются в электронном виде (ntvitmo@itmo.ru). В распечатанном виде с подписями авторов материалы представляются в редакцию по запросу редакции после прохождения процедуры рецензирования.

Комплект документов должен включать:

- текст статьи с заверстанными рисунками и таблицами;
- формы сведений об авторах (на каждого автора); допускается указание нескольких мест работы автора;
- файлы с рисунками к статье в оригинальном формате (предпочтительно JPEG) с максимальным разрешением; допускается представление цветных рисунков, если в черно-белом варианте теряется полезная информация;
- лицензионное соглашение;
- согласие на обработку персональных данных.

REQUIREMENTS FOR EXECUTION OF PAPERS

Scientific reviews, full-text and brief papers are published.

A full-text paper should have a well-defined structure, including an abstract, keywords, introduction, several substantive chapters and a conclusion.

An abstract is intended for general public. It is necessary to set forth the scientific content of the paper limited by 200–400 words in the form of a brief abstract: the subject, the work objective, the method or methodology of work, a brief description of the experiment, obtained results, recommendations for their application.

Keywords should be connected with the information presented in the paper. The main keyword is given first.

The introduction should contain a meaningful statement of the issue in question, a brief analysis of the solutions known from scientific literature (with references to sources), a criticism of their shortcomings, scientific novelty and advantages (features) of the proposed approach.

The main body of the paper should represent exact statement of the problem being solved, obtained assertions and conclusions are to be set forth and explained in detail (proved), the results of experimental studies or mathematical modeling should be given to illustrate the statements made. The main text of the paper needs to be split up into meaningful sections (2–3 sections).

In conclusion it is necessary to summarize the main results, comment on them and, if possible, to indicate the areas of future research and application.

A list of references is recommended: for a review paper — at least 50 items, for full-text paper — not less than 15, for a brief paper — at least 8 references. References to standards and other regulatory documents, as well as to Internet resources, are not included in a reference list and are documented as footnotes.

The length of a review paper is coordinated with the editorial board beforehand.

The length of a full-text paper, including illustrations, tables and references, should not exceed 8 pages of typewritten text (figures and tables included), font 12 pt, single-spaced.

The length of a brief paper is up to 3 pages, font 12 pt, single-spaced.

No rubrication is required.

Papers are accepted in electronic form (ntvitmo@itmo.ru).

Materials are submitted to the editors on request after the procedure of reviewing in a printed form signed by the authors.

A set of documents should include:

- The text of the paper with make-up figures and tables;
- The forms with information about the authors (for each author), several employers can be specified by the author;
- Files with pictures to the paper in their original format (preferably JPEG) with the maximum resolution; should useful information is lost in the black and white variant, color drawings are permitted;
- Publishers license agreement;
- Consent to the processing of personal data.

PROCEEDINGS OF THE XXXII SCHOOL ON HOLOGRAPHY**Part III**

Characterization of the holographic photopolymer Bayfol HX in the IR region	<i>Borisov V.N., Zverev A.D., Kamynin V.A., Kopyeva M.S., Okun R.A., Tsvetkov V.B.</i>	1
Study of blood vessels reaction to local heating by imaging photoplethysmography	<i>Belaventseva A.V., Podolyan N.P., Volynsky M.A., Zaytsev V.V., Sakovskaia A.V., Mamontov O.V., Romashko R.V., Kamshilin A.A.</i>	14
Transmission of 3D holographic information over a radio channel by a method close to SSB	<i>Pazoev A.L., Shoydin S.A.</i>	21

NANOSTRUCTURES. PROPERTIES AND APPLICATIONS

Anodization parameters influence on anodic aluminum oxide formed above the silver island film	<i>Nikitin I.Yu., Nabiullina R.D., Nashyokin A.V., Starovoytov A.A., Gladskikh I.A.</i>	28
---	---	----

AUTOMATIC CONTROL AND ROBOTICS

State estimation accuracy analysis of an induction electric drive by the algorithms of Luenberger and Kalman	<i>Bukreev V.G., Shandarova E.B., Perevoshchikov F.V.</i>	35
A method of optimizing the structure of hierarchical distributed control systems	<i>Onufrey A.Yu., Razumov A.V., Kakaev V.V.</i>	44
Method for identification of sinusoidal signal parameters with variable unknown amplitude	<i>Nguyen K.T., Vlasov S.M.</i>	54

NEW MATERIAL AND NANOTECHNOLOGIES

Improvement of the automatic temperature stabilisation process in the cryovacuum unit (in English)	<i>Golikov O.Yu., Yerezhap D., Sokolov D.Yu.</i>	62
Investigation on impact and wear behavior of Al6061 (SiC + Al ₂ O ₃) and Al7075 (SiC + Al ₂ O ₃) hybrid composites (in English)	<i>Ravichandaran R., Selvarasu S., Gopal S., Ramachandran R.</i>	68

COMPUTER SCIENCE

Computational methods to increase the speed of FPGA-based discrete wavelet transforms	<i>Sai S.V., Zinkevich A.V.</i>	79
Dialogue system based on spoken conversations with access to an unstructured knowledge base	<i>Masliukhin S.M.</i>	88

ARTIFICIAL INTELLIGENCE AND COGNITIVE INFORMATION TECHNOLOGIES

Multiobjective evolutionary discovery of equation-based analytical models for dynamical systems (in English)	<i>Maslyaev M.A., Hvatov A.A.</i>	97
Probabilistic criteria for time-series predictability estimation (in English)	<i>Kovantsev A.N.</i>	105
Software framework for hyperparameters optimization of models with additive regularization	<i>Khodorchenko M.A., Butakov N.A., Nasonov D.A., Firulik M.Yu.</i>	112
Value-based modeling of economic decision making in conditions of unsteady environment	<i>Guleva V.Yu., Kovatsev A.N., Surikov A.G., Chunaev P.V., Gornova G.V., Boukhanovskiy A.V.</i>	121
Methodology for organizing and conducting a study to assess consumer ability	<i>Aleynikov S.A., Gofman O.O., Basov O.O.</i>	136
Automated cluster analysis of communication strategies of educational telegram channels	<i>Nizomutdinov B.A., Uglova A.B., Bogdanovskaya I.M.</i>	142

MODELING AND SIMULATION

Computer modeling of non-Markovian processes based on the principle of balance of "complex probabilities"	<i>Gusenitsa Ya.N., Shiryamov O.A.</i>	150
The objectification method of the weight coefficients for decision-making in multicriteria problems	<i>Solovjev D.S.</i>	161
Assessment of the limits of applicability and methods of modulation of near-field magnetic coupling	<i>Grishentsev A.Yu., Goroshkov V.A., Chernov R.I.</i>	169
Numerical and analytical modeling of the propulsive wing and fuselage of an air taxi	<i>Bulat P.V., Kurnukhin A.A., Prodan N.V.</i>	178
Using variable-precision feedback to improve operational speed of the current loop in GaN-inverters	<i>Anuchin A.S., Gulyaeva M.A., Lashkevich M.M., Zharkov A.A., Chen H., Dianov A.N.</i>	192
Simulation of diffusion processes during electrothermal treatment of reaction crucibles of the Fe-Sn system	<i>Fomin V.E., Tukmakova A.S., Bolkunov G.A., Novotelnova A.V., Bochkhanov F.Yu., Karpenkov D.Yu.</i>	202