VİTMO

НАУЧНО-ТЕХНИЧЕСКИЙ ВЕСТНИК ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ, МЕХАНИКИ И ОПТИКИ март–апрель 2025 Том 25 № 2 http://ntv.ifmo.ru/ SCIENTIFIC AND TECHNICAL JOURNAL OF INFORMATION TECHNOLOGIES, MECHANICS AND OPTICS March–April 2025 Vol. 25 No 2 http://ntv.ifmo.ru/en/ ISSN 2226-1494 (print) ISSN 2500-0373 (online)

ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ, МЕХАНИКИ И ОПТИКІ

ОПТИЧЕСКИЕ СИСТЕМЫ И ТЕХНОЛОГИИ OPTICAL ENGINEERING

doi: 10.17586/2226-1494-2025-25-2-179-189 УДК 535.211

Исследование фазовых превращений окалины углеродистой стали при лазерной очистке наносекундными импульсами в режиме испарения Данила Владимирович Журба¹⊠, Владимир Михайлович Журба², Вадим Павлович Вейко³, Дмитрий Васильевич Панькин⁴, Михаил Валерьевич Жуков⁵, Александр Эдуардович Пуйша⁶

1,2,6 ООО «НПП ВОЛО», Санкт-Петербург, 199034, Российская Федерация

1,3 Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация

⁴ Санкт-Петербургский государственный университет, Санкт-Петербург, 199034, Российская Федерация

⁵ Институт аналитического приборостроения РАН, Санкт-Петербург, 198095, Российская Федерация

¹ zhurba.danila306@ya.ru^{\[\]}, https://orcid.org/0009-0001-6814-1737

² vlzhurba@yandex.ru, https://orcid.org/0009-0006-9919-3106

³ vadim.veiko@mail.ru, https://orcid.org/0000-0001-6071-3449

⁴ dmitrii.pankin@spbu.ru, https://orcid.org/0000-0003-3114-3410

⁵ cloudjyk@yandex.ru, https://orcid.org/0000-0003-3361-6947

⁶ pujsha@mail.ru, https://orcid.org/0009-0000-6589-6723

Аннотация

Введение. Исследованы фазовые превращения, происходящие в окалине при воздействии на нее наносекундными лазерными импульсами. Определен исходный фазовый состав окалины и фазовый состав поверхностного слоя, подвергшегося лазерному воздействию. Обработка поверхности образцов осуществлялась в испарительном режиме лазерного воздействия и приводила к абляции окалины. Метод. Исследовались две группы образцов из листового горячекатаного проката углеродистой стали марки Ст3 (Е235-С, Fe 360-С). Первая группа состояла из образцов с исходной поверхностью окалины и образцов с механически шлифованной поверхностью. Исследовались фазовый и элементный составы, а также морфологические параметры исходной окалины. Во вторую группу были включены образцы с поверхностью окалины, обработанной наносекундными лазерными импульсами наносекундного иттербиевого волоконного лазера с максимальной средней мощностью 30 Вт. Для сканирования поверхности образцов пучком лазерного излучения применялась двухкоординатная сканирующая система на основе гальваносканеров. Фазовый состав окалины определялся методом спектроскопии комбинационного рассеяния света. Морфологические параметры поверхности и элементный состав образцов исследовались методами сканирующей электронной микроскопии, атомно-силовой микроскопии и энергодисперсионного анализа. Основные результаты. Исследования фазового состава исходной окалины показали, что она состоит в основном из магнетита, при этом вюстит в составе окалины обнаружен не был. Установлено, что при обработке окалины в испарительном режиме в зоне воздействия лазерного импульса формируется кратер, поверхность которого покрыта застывшим расплавом окалины. В расплаве происходит фазовое превращение с образованием вюстита. При застывании расплав растрескивается, что связано с произошедшим фазовым превращением. Показано, что в процессе лазерной очистки испарительному механизму удаления окалины сопутствует фазовое превращение смеси магнетита с металлическим железом в вюстит. Обсуждение. Полученные результаты могут быть положены в основу создания новой высокоэффективной технологии лазерной очистки поверхности стали от окалины.

Ключевые слова

лазерная очистка, фазовые превращения окалины, термомеханическое разрушение окалины, наносекундный волоконный лазер, структура прокатной окалины, рамановская спектроскопия окалины

Благодарности

Работа выполнена при научной и технической поддержке ООО «Научно-производственное предприятие Волоконно-Оптического и Лазерного Оборудования» (Российская Федерация).

Спектры комбинационного рассеяния света были получены в Ресурсном Центре «Оптические и лазерные методы исследования вещества» Научного парка ФГБОУ ВО «Санкт-Петербургский государственный университет». Работа выполнена при поддержке СПбГУ, шифр проекта АААА-А19-119082790069-6.

© Журба Д.В., Журба В.М., Вейко В.П., Панькин Д.В., Жуков М.В., Пуйша А.Э., 2025

Ссылка для цитирования: Журба Д.В., Журба В.М., Вейко В.П., Панькин Д.В., Жуков М.В., Пуйша А.Э. Исследование фазовых превращений окалины углеродистой стали при лазерной очистке наносекундными импульсами в режиме испарения // Научно-технический вестник информационных технологий, механики и оптики. 2025. Т. 25, № 2. С. 179–189. doi: 10.17586/2226-1494-2025-25-2-179-189.

Investigation of phase transformations of carbon steel scale during laser purification by nanosecond pulses in the evaporation mode

Danila V. Zhurba^{1⊠}, Vladimir M. Zhurba², Vadim P. Veiko³, Dmitrii V. Pankin⁴, Mikhail V. Zhukov⁵, Alexander E. Puisha⁶

1,2,6 NPP VOLO LLC, Saint Petersburg, 199034, Russian Federation

^{1,3} ITMO University, Saint Petersburg, 197101, Russian Federation

⁴ St. Petersburg State University (SPbSU), Saint Petersburg, 199034, Russian Federation

⁵ Institute for Analytical Instrumentation of the Russian Academy of Sciences (IAI RAS), Saint Petersburg, 198095, Russian Federation

¹ zhurba.danila306@ya.ru[⊠], https://orcid.org/0009-0001-6814-1737

² vlzhurba@yandex.ru, https://orcid.org/0009-0006-9919-3106

³ vadim.veiko@mail.ru, https://orcid.org/0000-0001-6071-3449

⁴ dmitrii.pankin@spbu.ru, https://orcid.org/0000-0003-3114-3410

⁵ cloudjyk@yandex.ru, https://orcid.org/0000-0003-3361-6947

⁶ pujsha@mail.ru, https://orcid.org/0009-0000-6589-6723

Abstract

The phase transformations occurring in scale when exposed to nanosecond laser pulses are investigated. The initial phase composition of the scale and the phase composition of the surface layer exposed to the laser have been determined. The surface treatment of the samples was carried out in the evaporative mode of laser exposure and led to scale ablation. Two groups of samples from hot-rolled carbon steel sheets of the St3 grade (E235, Fe 360) were studied. The first group consisted of samples with an initial scale surface and samples with a mechanically ground surface. Based on these samples, the phase and elemental composition as well as the morphological parameters of the initial scale were studied. The second group includes samples with a scale surface treated with nanosecond laser pulses. A pulsed nanosecond ytterbium fiber laser with a maximum average power of 30 watts was used. A two-coordinate scanning system based on electroplating scanners was used to scan the surface of the samples with a laser beam. The phase composition of the scale was determined by RAMAN spectroscopy. The morphological parameters of the surface and the elemental composition of the samples were determined by scanning electron microscopy (SEM), atomic force microscopy (AFM), and energy dispersion analysis (EDX). Studies of the phase composition of the initial scale showed that it consists mainly of magnetite, while wustite was not detected in the scale. It has been established that during the processing of scale in the evaporation mode a crater is formed in the area affected by the laser pulse the surface of which is covered with a solidified melt of scale. A phase transformation occurs in the melt with the formation of wustite. Upon solidification, the melt cracks, which is associated with the phase transformation that has occurred. Thus, it is shown that in the process of laser purification, the evaporative mechanism of scale removal is accompanied by a phase transformation of a mixture of magnetite and metallic iron into wustite. The results obtained can be used as a basis for the creation of a new highly efficient technology for laser surface cleaning of steels from scale.

Keywords

laser purification, phase transformations of scale, thermomechanical destruction of scale, nanosecond fiber laser, structure of mill scale, Raman spectroscopy of scale

Acknowledgements

The work was carried out with the scientific and technical support of NPP VOLO LLC (Russian Federation). The Raman spectra of light were obtained at the Resource Center "Optical and Laser Methods of Substance Research" of the Saint Petersburg State University Science Park. The work was carried out with the support of Saint Petersburg State University; the project code is AAAA-A19-119082790069-6.

For citation: Zhurba D.V., Zhurba V.M., Veiko V.P., Pankin D.V., Zhukov M.V., Puisha A.E. Investigation of phase transformations of carbon steel scale during laser purification by nanosecond pulses in the evaporation mode. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 2025, vol. 25, no. 2, pp. 179–189 (in Russian). doi: 10.17586/2226-1494-2025-25-2-179-189

Введение

Удаление оксидных слоев с поверхности горячекатаных углеродистых сталей (прокатной окалины) — актуальная задача для современного производства [1, 2]. Реализация технологий лазерной очистки поверхности металлопроката позволит решить множество проблем в данной сфере, связанных с использованием устаревающих механических и химических методов очистки. По сравнению с традиционными методами лазерная очистка металлопроката обладает следующими преимуществами: отсутствием вредного воздействия на окружающую среду и здоровье людей, отсутствием расходных материалов и инструментов, низкими эксплуатационными затратами, отсутствием или минимальным повреждением очищаемой поверхности металла [3–9].

В настоящее время наиболее распространены два основных подхода к лазерной очистке, основанные на

применении либо импульсных иттербиевых волоконных лазеров (чаще всего с наносекундной длительностью импульса), либо непрерывных иттербиевых волоконных лазеров. При обоих подходах, как правило, удаление загрязнений осуществляют за счет испарения при нагреве выше температуры кипения [3-12]. Однако испарение оксидных слоев и особенно окалины — весьма энергоемкий процесс. Испарительная лазерная очистка прокатной окалины с поверхности горячекатаной стали характеризуется низкой производительностью [10, 11]. Нагрев поверхности стали при ее очистке излучением непрерывного лазера в испарительном режиме воздействия приводит к окислению, оплавлению, закалке или отпуску [13]. В случаях, когда такое термическое воздействие на поверхность стали недопустимо, очистка непрерывным лазерным излучением не может быть применена.

Известно, что лазерную очистку возможно осуществлять с помощью различных физических механизмов, связанных с интенсивным, локальным и неравномерным нагревом удаляемого загрязненного слоя, отдельных частиц в этом слое, а также основного материала. В результате такого нагрева в зоне лазерной обработки в основном возникают термомеханические напряжения, термохимические преобразования, газификация основного материала, термодесорбция адсорбированного покрытием газа, и другие. При очистке также возможны дополнительные механизмы, которые могут вносить значительный вклад в эффективность и производительность процесса. Например, при поверхностном испарении материала и достаточных для этого плотностях мощности лазерного излучения разрушение материала происходит также за счет действия ударной волны, выноса расплава под действием давления отдачи пара или его вскипания [14, 15].

В работах [16–18] рассмотрена возможность термомеханического разрушения прокатной окалины. Для разрушения окалины использовался непрерывный волоконный лазер. Нагрев поверхности окалины осуществлялся до температур, не прерывающих температуру ее кипения. Разрушение окалины происходило в результате действия совокупности механизмов: изменений в химическом составе окалины и возникновения растягивающих напряжений при ее охлаждении. По результатам [16–19] разработан способ лазерной очистки металлопроката от окалины.

Цель настоящей работы — исследование фазовых превращений в окалине, а также определение роли термомеханических напряжений в процессе испарительной лазерной очистки металлопроката от нее. Актуальность работы связана с возрастающим интересом современного производства к технологиям лазерной антикоррозионной обработки и лазерной очистки поверхности проката. Для высокопроизводительной лазерной очистки поверхности стали от окалины необходимо искать эффективные механизмы ее разрушения и удаления. В этом контексте представляет интерес исследование механизмов разрушения окалины за счет термических напряжений (которые могут возникать при циклическом нагреве и охлаждении) и за счет фазовых превращений в окалине при ее нагреве. Новизна работы состоит в выявлении фазовых превращений в окалине во время лазерной очистки при испарении, и в обосновании и демонстрации возможности термомеханического разрушения прокатной окалины в режиме воздействия. Результаты работы могут быть положены в основу создания новой высокоэффективной технологии лазерной очистки поверхности сталей от окалины.

Образцы и методика эксперимента

Исследуемые образцы были вырезаны из листового горячекатаного проката стали марки Ст3¹ толщиной 5 мм. Ст3 – углеродистая сталь обыкновенного качества, аналогом по стандарту ИСО 630:1995² является сталь E235-C (Fe 360-C). Содержание углерода в стали 0,14–0,22 %. На поверхности стальных листов в процессе их производства был сформирован равномерный тонкий слой прокатной окалины темно-синего цвета.

Для обработки поверхности образцов металлопроката лазерным излучением использовалась установка лазерной гравировки с импульсным волоконным иттербиевым лазером RFL-P30Q (Raycus, Китай) максимальной средней мощности 30 Вт, с длиной волны 1060-1085 нм, диаметром выходного пучка 6-8 мм, параметром M² < 1,6. Сканирование поверхности образцов лазерным пучком осуществлялось с помощью двухкоординатной сканирующей головки производства Sinogalvo, с использованием F-theta объектива с фокусным расстоянием 300 мм. Для данного лазерного источника и объектива расчетный диаметр лазерного пучка в плоскости фокусировки составил около 100 мкм. Воздействие на поверхность окалины выполнено при частоте следования импульсов 60 кГц, энергии в импульсе 0,5 мДж (согласно характеристикам заявленным производителем лазерного источника). При данных значениях энергии в импульсе и частоты следования импульсов, длительность импульса, измеренная по полувысоте сигнала с помощью фотодиода ФД-256 (Россия), равен 190 ± 10 нс. При данных значениях энергии, длительности импульса и диаметре лазерного пятна средняя плотность мощности в пятне составляет примерно 3·107 Вт/см².

Образцы были разделены на две группы. Первая группа образцов не подвергалась лазерному воздействию. Образец 1.1 — исходное состояние поверхности окалины. Образец 1.2 — шлифовка окалины шлифовальной бумагой Р2000 зернистостью 5–7 мкм в двух перпендикулярных направлениях. Шлифовка произведена вручную до удаления окалины в центре шлифуемого участка, что было определено по появлению светлой области, соответствующей поверхности стали. Во вторую группу вошли образцы с обработкой поверхности сти импульсным лазерным излучением. Под действием

¹ ГОСТ 380-2005. Сталь углеродистая обыкновенного качества. Введен 01.01.2008. М.: Стандартинформ, 2007. 16 с.

² ИСО 630:1995. Конструкционные стали. Прокат толстолистовой, широкополосный, сортовые и фасонные профили (ISO 630:1995 «Structural steels — Plates, wide flats, bars, sections and profiles», NEQ).

лазерных импульсов происходило удаление окалины, в зоне действия лазерного импульса формировались кратеры. Воздействие лазерного излучения с перечисленными характеристиками на сильно поглощающие материалы, например окалину или сталь, обеспечило нагрев выше порога кипения [8]. Удаление окалины выполнено за счет нагрева ее в течение длительности лазерного импульса выше температуры кипения. На образцах 2.1 и 2.2 были сформированы кратеры, расстояние между краями кратеров превышало их диаметр. Образец 2.1 — обработка одиночным лазерным импульсом (в каждую область воздействия приходится один импульс), образец 2.2 — обработка 10 импульсами (приходится 10 импульсов). Образцы 2.3 и 2.4 — обработка при перекрытии лазерных пятен в строке сканирования на 75 % от их диаметра и перекрытии строк сканирования на 75 % от диаметра лазерного пятна. Образец 2.3 обрабатывался за один проход. Образец 2.4 — за четыре прохода, так как при этом окалина полностью удалялась. На рис. 1 схематично показаны трассы сканирования лазерным пучком.

Заметим, что образцы 2.1–2.4 выбраны из соответствующих подгрупп для демонстрации характерных результатов. Каждая подгруппа включала в себя не менее трех образцов, обработанных в одинаковых условиях. Каждый образец имел площадь поверхности, достаточную для проведения необходимого числа измерений и обеспечения статистической достоверности полученных результатов.

Внешний вид кратеров, полученных при воздействии лазерными импульсами, регистрировался с помощью сканирующего электронного микроскопа (СЭМ) Inspect (FEI, США). Измерения латеральных размеров кратеров в СЭМ проводились в вакууме при давлении $10^{-3}-10^{-4}$ Па и ускоряющем напряжении 20 кВ в режиме вторичных электронов, рабочее расстояние от колонны до образца — 7 мм, размер пятна пучка электронов составил — 3,5 усл. ед.

Профиль поверхности кратеров исследован с помощью сканирующего зондового микроскопа Ntagra Aura (NT-MDT, Россия). Топография поверхности кратеров изучена с помощью полуконтактного режима атомно-силовой микроскопии (ACM) для уменьшения действия зонда на структуры поверхности кратера. Изучение образцов происходило в воздушной среде



Рис. 1. Трасса сканирования лазерными импульсами поверхности для образцов: 2.1 и 2.2 (*a*); 2.3 и 2.4 (*b*) Стрелками обозначено направление сканирования, кругами — лазерные пятна на обрабатываемой поверхности образцов

Fig. 1. Surface scanning path with laser pulses: for Samples 2.1 and 2.2 (*a*); for Samples 2.3 and 2.4 (*b*)

при средней температуре 21–23 °С и относительной влажности в пределах 40–45 %. При измерениях использовалась пассивная и активная виброзащиты. Для юстировки зонда и выбора области для АСМ измерений применена интегрированная оптическая система Optem (Qioptiq, CША).

При исследовании исходного фазового состава (входящих в состав окалины оксидов железа) и преобразования фаз окалины под действием лазерного нагрева реализован метод спектроскопии комбинационного рассеяния света (КРС). Данный оптический метод успешно используется для исследования структуры, состава и пространственного распределения фаз, например оксидов и оксигидроксидов железа, образовавшихся как естественным образом [20-22], так и в результате направленного лазерного воздействия при различных температурных и влажностных режимах [23-25]. Сопряжение спектрометра КРС с оптическим микроскопом позволяет получать информацию о фазовом составе неметаллических (оксидных) образований (фазы вюстита, магнетита, гематита и другие) с пространственным разрешением в единицы микрометров [24].

Спектры были получены методом КРС в геометрии обратного рассеяния [26] на спектрометре LabRam HR800 с возбуждением при использовании встроенного He-Ne лазера (Horiba Jobin Yvon, Франция). Длина волны лазерного излучения составляла 632,8 нм. Его фокусировка осуществлялась с использованием 50× объектива с числовой апертурой 0,75. Средняя мощность лазерного излучения на фокусном расстоянии от объектива составляла 0,1 мВт (разрешение измерения — 0,01 мкВт) и измерена с использованием измерителя лазерной мощности марки LaserCheck¹ (Coherent, США). При съемке спектров применена дифракционная решетка 600 шт/мм. Диаметр ирисовой диафрагмы составлял 150 ± 0.5 мкм. Время накопления однократного спектра КРС в каждом случае подбиралось индивидуально и находилось в интервале 200-600 с. Выбор этого времени осуществлялся на основании совокупности следующих факторов: сечения рассеяния исследуемых веществ (оксидов и оксигидроксидов железа), их количества, чувствительности детектора, а также корректности работы алгоритма фильтрации спайков² [27]. В целом, подобранные времена накопления спектров КРС позволяли надежно регистрировать на спектрометре LabRam HR800 в интересующих областях (кратерах и возле них) даже слабые спектральные особенности, такие как, например, пик полосы поглощения магнетита расположенный около 310 см⁻¹, с соотношением сигнал/ шум более 3/1. Число повторов однократных спектров КРС было выбрано в количестве двух, что связано с работой алгоритма Spike Filter по удалению спайков

¹ LCh. LaserCheck Power Meter [Электронный ресурс]. Режим доступа: https://www.coherent.com/laser-power-energymeasurement/meters/lasercheck (дата обращения: 25.11.2024).

² Labspec. LabSpec 6 Spectroscopy Suite Software Meter [Электронный ресурс]. Режим доступа: https://www.horiba. com/int/scientific/products/detail/action/show/Product/labspec-6-spectroscopy-suite-software-1843/ (дата обращения: 25.11.2024).

(в англоязычных научных работах иногда вместо spike используется аналогичный по смыслу термин cosmic rays) [27].

Исследование фазовых превращений в окалине при лазерной очистке

Окалина на поверхности сталей — слой оксидов железа (черного, сизого или темно-красного цвета) формируется в результате термической обработки сталей в окислительной атмосфере, например, при прокатке (прокатная окалина). Тонкая пленка прокатной окалины получается однородной по своим свойствам и более пластичной, чем классическая печная окалина. Такая окалина в основном состоит из первичного магнетита и продукта распада вюстита — смеси магнетита и металлического железа [28-33]. Она не имеет выраженной слоистости, обладает высокой адгезией к поверхности стали и хорошо сопротивляется механическому воздействию [34]. При температуре выше 570 °С в окалине формируется вюстит. При температурах более 1250 °С вюстит становится пластичным и сопротивление сдвигу в слое вюстита становится близким к нулю. В связи с этим при высокой температуре характерна уменьшенная величина адгезии окалины к поверхности металла [35]. Различия коэффициентов теплового расширения слоев окалины и металла, а также снижение адгезии при формировании вюстита являются предпосылками для разрушения окалины при лазерном нагреве и последующем быстром охлаждении за счет теплопроводности. Основным преимуществом использования лазерного излучения в качестве источника тепла для нагрева и реализации механизмов, приводящих к удалению окалины, является локальность и высокая скорость нагрева. В этом случае удаление материала (лазерную очистку) можно проводить с вкладом энергии меньшим, чем при других способах нагрева (в печах, открытым пламенем, токами высокой частоты и т. п.). Выделение тепла в окалине, как и в других сильно поглощающих материалах, происходит в тонком поверхностном слое, толщину которого можно определить по показателю поглощения материала для длины волны используемого излучения. Показатель поглощения окалины при длине волны 1,06 мкм составляет 43 675 обратных сантиметров¹, а глубина проникновения излучения — 230 нм. Вглубь тепло распространяется за счет теплопроводности, толщину прогретого слоя *h* можно оценить согласно выражению $h = \sqrt{\alpha \tau}$, где α — температуропроводность материала (для магнетита 10-6 м²/с [36]), т — длительность воздействия (например, длительность лазерного импульса) [8]. Управляя параметрами лазерного воздействия: длительностью импульса, размером пятна, энергией в импульсе, частотой следования импульсов, а также скоростью сканирования лазерным пучком поверхности, можно управлять толщиной и температурой прогретого слоя материала.

При воздействии одиночного лазерного импульса на окалину (магнетит) в приведенном режиме работы лазерной установки (раздел «Образцы и методика эксперимента») глубина прогретого слоя достигает 0,5 мкм, при этом поверхность нагревается существенно выше температуры кипения.

На рис. 2 представлен внешний вид отдельного абляционного кратера, образованного в результате воздействия одиночного лазерного импульса (образец 2.1). Удаление окалины происходит в основном за счет испарения. Расплав окалины остается в области лазерного воздействия. Давление пара приводит к прогибу ванны расплава и выдавливанию части расплава на периферию кратера, где формируется валик. После окончания лазерного импульса расплав на поверхности абляционного кратера застывает.

Микрорельеф и микропрофиль поверхности кратера, образованного одиночным лазерным импульсом представлен на рис. 3. Максимальная глубина кратера над уровнем исходной поверхности образца и высота валика расплава составляла примерно 1,5 мкм. Высота валика расплава также близка к 1,5 мкм.

Для удаления окалины в центре кратера до поверхности стали потребовалось 10 импульсов, приходящихся в одну и ту же область. На рис. 4 представлен абляционный кратер в образце 2.2, сформированный при воздействии 10 лазерных импульсов. Появление поверхности стали на дне кратера идентифицируется как по внешнему виду (изменяется топология расплава), так и по рентгеновскому спектру (снижается массовая доля кислорода). Поверхность расплава из пористой и трещиноватой, характерной для расплава окалины, в области кратера, где была достигнута поверхность стали, становится гладкой, без трещин и пор. Массовая доля кислорода в окалине на краю кратера составляет более 41 %, в центре кратера — 12 %.

Элементный анализ проводился методом энергодисперсионной рентгеновской спектроскопии (EDX). Сбор данных EDX осуществляется в импульсной форме. Когда входящие импульсы напряжения или тока измеряются импульсным процессором в детекторе EDX, детектор фактически отключается на период времени, называемый мертвым временем. Другими словами,



Puc. 2. СЭМ-изображение кратера в окалине, образованного одиночным лазерным импульсом (образец 2.1)*Fig. 2.* SEM image of a crater in the scale formed by a single laser pulse (Sample 2.1)

¹ Optical constants of Fe3O4 (Iron(II,III) oxide, Magnetite) Querry 1985: n,k 0.21-55.6 µm [Электронный ресурс]. Режим доступа: https://refractiveindex.info/?shelf=main&book=Fe3O4 &page=Querry (дата обращения: 25.11.2024).



Рис. 3. Микрорельеф (a) и микропрофиль (b) поверхности кратера в окалине, образованного одиночным лазерным импульсом (образец 2.1), снятый вдоль отрезка на поверхности окалины, полученный с помощью атомно-силовой микроскопии

Fig. 3. Microrelief (*a*) and microprofile (*b*) of the crater surface in the scale formed by a single laser pulse (Sample 2.1) taken along a segment on the surface of the scale, obtained using atomic force microscopy

когда система EDX не подсчитывает входящие рентгеновские фотоны, а обрабатывает ранее собранный сигнал, она считается «мертвой». Мертвое время (Dead Time, DT) определяется как:

$$DT = (1 - Rout/Rin) \times 100 \%$$

где Rout и Rin — скорости счета на выходе и входе.

Сигнал EDX регистрировался с помощью СЭМ Inspect (FEI, США) при ускоряющем напряжении около 20 кВ, DT составляло 30–40 %, диаметр пятна электронного пучка составил — 4,5 усл. ед. Оптимальными в большинстве случаев являются значения DT от 15 до 40 % (около 1500 и 5000 имп./с) [37].

В таблице приведены волновые числа фононов, определенные на основании полученных спектров КРС для идентифицированных путем сравнения с данными работы [21] оксидов железа (гематита, магнетита, вюстита).



 Рис. 4. СЭМ-изображение кратера в окалине, образованного 10 лазерными импульсами (образец 2.2).
 В центре кратера видна оплавленная поверхность стали (указана стрелкой)

Fig. 4. SEM image of a crater in the scale formed by 10 laser pulses (Sample 2.2). A fused steel surface is visible in the center of the crater (indicated by an arrow)

Полученные спектры КРС образцов представлены на рис. 5. Гематит в образцах идентифицируется по характерным пикам полосы в спектре КРС примерно 220 см⁻¹ и 290 см⁻¹, магнетит — по характерному пику 670 см⁻¹, вюстит — по пику 660 см⁻¹.

Характерные для исходной поверхности окалины (образец 1.1) КРС (рис. 5, a) содержат полосы, которые можно отнести к магнетиту и гематиту. На рис. 6 показана область образца 1.2 (после шлифовки), в которой снимались КРС. Видно, что окалина имеет пористую структуру. Спектры от плоской поверхности (рис. 5, b) соответствуют магнетиту, а от пор (рис. 5, c) — магнетиту и гематиту и повторяют спектры исходной поверхности окалины.

По всей глубине окалины во всех образцах в основном содержится магнетит. Гематит присутствует только в виде следов на поверхностях, контактировавших с воздухом до шлифовки. Вюстита в исходной окалине обнаружено не было. Полученные результаты подтверждаются исследованиями процесса формирования прокатной окалины. Прокатная окалина состоит в основном из магнетита и металлического железа. Механическая смесь магнетита и металлического железа образуется в результате реакции распада вюстита во время остывания проката при температуре ниже 570 °C [28–35].

Особый интерес представляют химические преобразования в окалине, происходящие при лазерном нагреве. Полученные с поверхности кратеров спектры КРС показывают, что застывший расплав состоит из вюстита (рис. 5, d). Данный результат можно объяснить тем, что при нагреве окалины выше 570 °С из механической смеси магнетита и металлического железа может образовываться вюстит. В ходе исследований было обнаружено, что на поверхности кратера в образце 2.1 застывший расплав растрескался на фрагменты с продольными размерами около 2–5 мкм (рис. 7).

На рис. 8 приведены изображения поверхностей образцов 2.3 и 2.4 при обработке за один (рис. 8, *a*) и за

Таблица. Волновые числа пиков в спектрах комбинационного рассеяния оксидов железа в составе исследованной окалины и приведенные в работе [21]

<u> </u>	Волновое число, см ⁻¹				
Соединение	Результаты настоящей работы	Данные из работы [21]			
Гематит	220, 290, 410, 500, 610	225, 293, 412, 498, 612			
Магнетит	310, 540, 670	310, 540, 670			
Вюстит	660	652			





Рис. 5. Спектры комбинационного рассеяния: исходная поверхность окалины (образец 1.1) (*a*); шлифованная поверхность окалины, плоская область (образец 1.2) (*b*); пора на поверхности окалины (образец 1.2) (*c*); кратер от одиночного импульса (образец 2.1) (*d*)

Fig. 5. Raman spectra: the initial surface of the scale (Sample 1.1) (*a*); the polished surface of the scale, a flat area (Sample 1.2) (*b*); a pore on the surface of the scale (Sample 1.2) (*c*); a crater from a single pulse (Sample 2.1) (*d*)



Рис. 6. Микрофотография шлифованной поверхности окалины (образец 1.2), в перекрестии находится открытая пора

Fig. 6. Micrograph of the polished surface of the scale (Sample 1.2), with an open pore in the crosshair



Рис. 7. СЭМ-изображение поверхности кратера в окалине, образованного одиночным лазерным импульсом (образец 2.1)
 Fig. 7. SEM image of the crater surface in the scale formed by a single laser pulse (Sample 2.1)



Рис. 8. СЭМ-изображения поверхностей образцов, полученных в результате сканирования с 75 % перекрытием лазерных импульсов: образец 2.3 (1 проход) (*a*); образец 2.4 (4 прохода) (*b*)

Fig. 8. SEM images of the sample surface formed as a result of scanning with 75 % overlap of laser pulses: Sample 2.3 (1 pass) (*a*); Sample 2.4 (4 passes) (*b*)

четыре (рис. 8, *b*) прохода. За один проход происходит удаление слоя окалины на глубину меньше толщины всей пленки окалины. Поверхность образца 2.3 остается покрытой оплавленным остаточным слоем окалины. После четырех проходов поверхность образца 2.4 полностью очистилась от окалины, степень очистки соответствует Sa3 по стандарту очистки металлической поверхности ISO 8501-1¹. В отличие от расплава стали трещиноватость расплава окалины при застывании носит ярко выраженный характер. К растрескиванию, вероятно, приводят различие теплофизических свойств нагретого верхнего и более холодного нижнего слоев окалины и фазовые превращения в ее структуре. Содержание вюстита в зоне лазерного воздействия регистрируется вплоть до полного удаления окалины. Спектры КРС слоя на поверхности как отдельного кратера, так и поверхности, образованной в результате сканирования с перекрытием лазерных импульсов, например образца 2.3, соответствуют спектру вюстита.

Плотность твердого вюстита больше, чем плотность твердого магнетита примерно на 15 % и более чем на 30 % больше, чем плотность расплава магнетита [36]. К растрескиванию застывшего расплава окалины, вероятно, приводят растягивающие напряжения, возни-

кающие из-за увеличения ее плотности при фазовом превращении магнетита во вюстит.

Из полученных результатов, видно, что совокупность термоиндуцированных фазовых превращений и термомеханических напряжений вносит вклад в процесс испарительной лазерной очистки. Возникающая при образовании плазмы ударная волна в окалине и давление отдачи паров [15] могут отрывать фрагменты растрескавшейся окалины.

Заключение

В работе проведены исследования фазовых превращений в структуре прокатной окалины углеродистой стали, происходящих в процессе лазерной очистки в режиме испарения под действием наносекундных импульсов волоконного иттербиевого лазера. Обнаружено, что существенный вклад в процесс очистки вносят термоиндуцированные фазовые превращения и термомеханические напряжения.

Методом спектроскопии комбинационного рассеяния света исследован фазовый состав исходной прокатной окалины горячекатаного проката стали марки Ст3, а также фазовый состав окалины после лазерного воздействия. В результате воздействия лазерного импульса происходит испарение окалины и образование ее расплава. Показано, что основной неметаллической составляющей исходной прокатной окалины является магнетит. В застывшем после лазерного воздействия расплаве окалины происходит превращение смеси магнетита и металлического железа во вюстит.

Вскрытые термомеханические явления и фазовые превращения, сопутствующие испарению прокатной окалины, нуждаются в дальнейших исследованиях для разработки новых методов лазерной очистки с повышенным качеством и производительностью.

¹ ИСО 8501-1:2007. Подготовка стальной поверхности перед нанесением лакокрасочных материалов и относящихся к ним продуктов. Визуальная оценка чистоты поверхности. Часть 1. Степень окисления и степени подготовки непокрытой стальной поверхности и стальной поверхности после полного удаления прежних покрытий (ISO 8501-1:2007 Preparation of steel substrates before application of paints and related products — Visual assessment of surface cleanliness — Part 1: Rust grades and preparation grades of uncoated steel substrates and of steel substrates after overall removal of previous coatings).

Литература

- Wang X., Ai R., Yang Q., Wang S., Zhang Y., Meng Y., Ma X. Effect of oxide scale structure on shot-blasting of hot-rolled strip steel // PeerJ Materials Science. 2020. V. 2. P. e 9. https://doi.org/10.7717/ peerj-matsci.9
- Bin S., Cheng L., Du C.-Y., Zhang J.-K., He Y.-Q., Cao G.-M. Effect of oxide scale microstructure on atmospheric corrosion behavior of hot rolled steel strip // Coatings. 2021. V. 11. N 5. P. 517. https://doi. org/10.3390/coatings11050517
- Li X., Huang T., Chong A.W., Zhou R., Choo Y.S., Hong M. Laser cleaning of steel structure surface for paint removal and repaint adhesion // Guangdian Gongcheng/Opto-Electronic Engineering. 2017. V. 44. N 3. P. 340–344.
- Kayahan E., Candan L., Aras M., Gundogdu O. Surface cleaning of metals using low power fiber lasers // Acta Physica Polonica A. 2018. V. 134. N 1. P. 371–373. https://doi.org/10.12693/aphyspola.134.371
- Lu Y., Yang L., Wang Y., Chen H., Guo B., Tian Z. Paint removal on the 5A06 Aluminum alloy using a continuous wave fiber laser // Coatings. 2019. V. 9. N 8. P. 488. https://doi.org/10.3390/ coatings9080488
- Kravchenko Y., Klimentov S., Derzhavin S., Mamonov D., Karpov N., Mayorov A. Optimization of laser cleaning conditions using multimode short-pulse radiation // Optical and Quantum Electronics. 2020. V. 52. N 6. P. 280. https://doi.org/10.1007/s11082-020-02399-1
- Sun X., Yu Q., Bai X., Jin G., Cai J., Yuan B. Substrate cleaning threshold for various coated al alloys using a continuous-wave laser // Photonics. 2021. V. 8. N 9. P. 395. https://doi.org/10.3390/ photonics8090395
- Вейко В.П., Смирнов В.Н., Чирков А.М., Шахно Е.А. Лазерная очистка в машиностроении и приборостроении. СПб.: НИУ ИТМО, 2013. 103 с.
- Veiko V., Samohvalov A., Ageev E. Laser cleaning of engraved rolls coupled with spectroscopic control // Optics and Laser Technologies. 2013. V. 54. P. 170–175. https://doi.org/10.1016/j. optlastec.2013.05.015
- Chen G.X., Kwee T.J., Tan K.P., Choo Y.S., Hong M.H. High-power fibre laser cleaning for green shipbuilding // Journal of Laser Micro/ Nanoengineering. 2012. V. 7. N 3. P. 249–253. https://doi. org/10.2961/jlmn.2012.03.0003
- Deschênes J.M., Fraser A. Empirical study of laser cleaning of rust, paint, and mill scale from steel surfac // Minerals, Metals & Materials Series. 2020. P. 189–201. https://doi.org/10.1007/978-3-030-36556-1 17
- Вейко В.П., Шахно Е.А. Физические механизмы лазерной очистки поверхности // Известия Российской академии наук. Серия физическая. 2001. Т. 65. № 4. С. 584–587.
- Zhuang S., Kainuma S., Yang M., Haraguchi M., Asano T. Characterizing corrosion properties of carbon steel affected by highpower laser cleaning // Construction and Building Materials. 2021. V. 274. P. 122085. https://doi.org/10.1016/j.conbuildmat.2020.122085
- Вейко В.П., Мутин Т.Ю., Смирнов В.Н., Шахно Е.А., Батище С.А. Лазерная очистка поверхностей металлов: физические процессы и применение // Известия высших учебных заведений. Приборостроение. 2008. Т. 51. № 4. С. 30–36.
- Либенсон М.Н., Яковлев Е.Б., Шандыбина Г.Д. Взаимодействие лазерного излучения с веществом (силовая оптика). Часть 2. Лазерный нагрев и разрушение материалов. Учебное пособие. СПб.: НИУ ИТМО, 2014. 181 с.
- Журба Д.В., Журба В.М., Вейко В.П., Пуйша А.Э. Исследование процесса лазерной очистки металлопроката от окалины // Фотоника. 2024. Т. 18. № 6. С. 436–449. https://doi.org/10.22184/1993-7296.FRos.2024.18.6.436.449
- 17. Журба Д.В. Исследование физических механизмов лазерной очистки поверхности горячекатаного металлопроката от окалины при воздействии непрерывного лазерного излучения с длиной волны 1,06 мкм // XI Конгресс молодых учёных: Сборник научных трудов конгресса. Т. 2. СПб.: НИУ ИТМО, 2022. С. 231–233.
- Журба Д.В., Журба В.М., Пуйша А.Э. Лазерная очистка поверхностей конструкций трубопроводного транспорта углеводородов // Бурение и нефть. 2023. № S2. С. 61–67.
- Журба В.М., Журба Д.В., Пуйша А.Э. Способ лазерной очистки металлических поверхностей от окалины. Патент RU2812150C1. Бюл. 2024. № 3.
- 20. Hanesch M. Raman spectroscopy of iron oxides and (oxy) hydroxides at low laser power and possible applications in environmental

References

- Wang X., Ai R., Yang Q., Wang S., Zhang Y., Meng Y., Ma X. Effect of oxide scale structure on shot-blasting of hot-rolled strip steel. *PeerJ Materials Science*, 2020, vol. 2, pp. e 9. https://doi.org/10.7717/peerjmatsci.9
- Bin S., Cheng L., Du C.-Y., Zhang J.-K., He Y.-Q., Cao G.-M. Effect of oxide scale microstructure on atmospheric corrosion behavior of hot rolled steel strip. *Coatings*, 2021, vol. 11, no. 5, pp. 517. https:// doi.org/10.3390/coatings11050517
- Li X., Huang T., Chong A.W., Zhou R., Choo Y.S., Hong M. Laser cleaning of steel structure surface for paint removal and repaint adhesion. *Guangdian Gongcheng/Opto-Electronic Engineering*, 2017, vol. 44, no. 3, pp. 340–344.
- Kayahan E., Candan L., Aras M., Gundogdu O. Surface cleaning of metals using low power fiber lasers. *Acta Physica Polonica A*, 2018, vol. 134, no. 1, pp. 371–373. https://doi.org/10.12693/aphyspola.134.371
- Lu Y., Yang L., Wang Y., Chen H., Guo B., Tian Z. Paint removal on the 5A06 Aluminum alloy using a continuous wave fiber laser. *Coatings*, 2019, vol. 9, no. 8, pp. 488. https://doi.org/10.3390/ coatings9080488
- Kravchenko Y., Klimentov S., Derzhavin S., Mamonov D., Karpov N., Mayorov A. Optimization of laser cleaning conditions using multimode short-pulse radiation. *Optical and Quantum Electronics*, 2020, vol. 52, no. 6, pp. 280. https://doi.org/10.1007/ s11082-020-02399-1
- Sun X., Yu Q., Bai X., Jin G., Cai J., Yuan B. Substrate cleaning threshold for various coated al alloys using a continuous-wave laser. *Photonics*, 2021, vol. 8, no. 9, pp. 395. https://doi.org/10.3390/ photonics8090395
- Veiko V.P., Smirnov V.N., Chirkov A.M., Shakhno E.A. Laser cleaning in mechanical and instrument engineering. St. Petersburg, NIU ITMO, 2013, 103 p. (in Russian)
- Veiko V., Samohvalov A., Ageev E. Laser cleaning of engraved rolls coupled with spectroscopic control. *Optics and Laser Technologies*, 2013, vol. 54, pp. 170–175. https://doi.org/10.1016/j. optlastec.2013.05.015
- Chen G.X., Kwee T.J., Tan K.P., Choo Y.S., Hong M.H. High-power fibre laser cleaning for green shipbuilding. *Journal of Laser Micro/ Nanoengineering*, 2012, vol. 7, no. 3, pp. 249–253. https://doi. org/10.2961/jlmn.2012.03.0003
- Deschênes J.M., Fraser A. Empirical study of laser cleaning of rust, paint, and mill scale from steel surface. *Minerals, Metals & Materials Series*, 2020, pp. 189–201. https://doi.org/10.1007/978-3-030-36556-1 17
- Veiko V.P., Shakhno E.A. Шахно E.A. Physical mechanisms of laser cleaning. *Izvestiya Rossiiskoi akademii nauk*. Seriya fizicheskaya, 2001, vol. 65, no. 4, pp. 584–587. (in Russian)
- Zhuang S., Kainuma S., Yang M., Haraguchi M., Asano T. Characterizing corrosion properties of carbon steel affected by highpower laser cleaning. *Construction and Building Materials*, 2021, vol. 274, pp. 122085. https://doi.org/10.1016/j. conbuildmat.2020.122085
- Veiko V.P., Mutin T.Yu., Smirnov V.N., Shakhno E.A., Batishche S.A. Laser clearing of surfaces of metals: physical processes and application. *Journal of Instrument Engineering*, 2008, vol. 51, no. 4, pp. 30–36. (in Russian)
- Libenson M.N., Iakovlev E.B., Shandybina G.D. Interaction Between Laser Radiation and Matter (Power Optics). Part 2: Laser Heating and Destruction of Materials. Study guide. St. Petersburg, NIU ITMO, 2014, 181 p. (in Russian)
- Zhurba D.V., Zhurba V.M., Veiko V.P., Puisha A.E. Investigation of the process of laser descaling of mill metal. *Photonics Russia*, 2024, vol. 18, no. 6, pp. 436–449. (in Russian). https://doi. org/10.22184/1993-7296.FRos.2024.18.6.436.449
- Zhurba D.V. Study of physical mechanisms of laser cleaning of hotrolled metal surface from scale under the influence of continuous laser radiation with a wavelength of 1.06 μm. *11th Congress of Young Scientists: Collection of Scientific Papers*. Vol. 2. St. Petersburg, NIU ITMO, 2022, pp. 231–233. (in Russian)
- Zhurba D.V., Zhurba V.M., Puisha A.E. Laser cleaning of surfaces of structures of pipeline transport of hydrocarbons. *Burenie i Neft*', 2023, no. S2, pp. 61–67. (in Russian)
- Zhurba D.V., Zhurba V.M., Pujsha A.E. Method for laser cleaning of metal surfaces from scale. *Patent RU2812150C1*. 2024. (in Russian)
- 20. Hanesch M. Raman spectroscopy of iron oxides and (oxy) hydroxides at low laser power and possible applications in environmental

magnetic studies // Geophysical Journal International. 2009. V. 177. N 3. P. 941–948. https://doi.org/10.1111/j.1365-246x.2009.04122.x

- De Faria D.L.A., Venâncio Silva S., de Oliveira M.T. Raman microspectroscopy of some iron oxides and oxyhydroxides // Journal of Raman Spectroscopy. 1997. V. 28. N 11. P. 873–878. https://doi.org/10.1002/(SICI)1097-4555(199711)28:11<873::AID-JRS177>3.0.CO;2-B
- 22. Prokuratov D.S., Davtian A.S., Vereshchagin O.S., Kurganov N.S., Samokhvalov A.A., Pankin D.V., Povolotckaia A.V., Shimko A.A., Mikhailova A.A., Somov P.A., Parfenov V.A. Laser cleaning of archaeologically corroded iron objects with inlays // Optical and Quantum Electronics. 2020. V. 52. N. 2. P. 113. https://doi. org/10.1007/s11082-020-2231-z
- Veiko V.P., Karlagina Y., Moskvin M., Mikhailovskii V., Odintsova G., Olshin P., Pankin D., Romanov V., Yatsuk R. Metal surface coloration by oxide periodic structures formed with nanosecond laser pulses // Optics and Lasers in Engineering. 2017. V. 96. P. 63–67. https://doi. org/10.1016/j.optlaseng.2017.04.014
- Sutton A.T., Kriewall C.S., Leu M.C., Newkirk J.W., Brown B. Characterization of laser spatter and condensate generated during the selective laser melting of 304L stainless steel powder // Additive Manufacturing. 2020. V. 31. P. 100904. https://doi.org/10.1016/j. addma.2019.100904
- Sánchez-Tovar R., Leiva-García R., García-Antón J. Characterization of thermal oxide films formed on a duplex stainless steel by means of confocal-Raman microscopy and electrochemical techniques // Thin Solid Films. 2015. V. 576. P. 1–10. https://doi.org/10.1016/j. tsf.2014.12.024
- 26. Encyclopedia of spectroscopy and spectrometry / ed. by Lindon J.C., Tranter G.E., Koppenaal D. Academic Press, 2016. 3584 p.
- Wiemann J., Fabbri M., Yang T.R., Stein K., Sander P.M., Norell M.A., Briggs D.E.G. Fossilization transforms vertebrate hard tissue proteins into N-heterocyclic polymers // Nature Communications. 2018. V. 9. P. 4741. https://doi.org/10.1038/s41467-018-07013-3
- Chen R., Yuen W. Oxidation of low-carbon, low-silicon mild steel at 450–900°C under conditions relevant to hot-strip processing // Oxidation of Metals. 2002. V. 57. N 1-2. P. 53–79. https://doi. org/10.1023/A:1013390628475
- Chen Rex, Yeun W.Y.D. Review of the High-Temperature Oxidation of Iron and Carbon Steels in Air or Oxygen // Oxidation of Metals. 2003. V. 59. N 5-6. P. 433-468. https://doi. org/10.1023/A:1023685905159
- Тепляков Ю.Н. Распад вюстита, входящего в состав окалины // Вестник Южно-Уральского государственного университета. Серия: Химия. 2009. № 23 (156). С. 36–42.
- Naipinij S., Sukieum S., Namprai R., Nilsonthi T. Formation of thermal oxide scale and its adhesion to hot-rolled low carbon steels with different final strip thicknesses // E3S Web of Conferences. 2022. V. 355. P. 02008. https://doi.org/10.1051/e3sconf/202235502008
- Cao G., Wu T., Xu R., Li, Z., Wang F., Liu Z. Effects of coiling temperature and cooling condition on transformation behavior of Tertiary oxide scale // Journal of Iron and Steel Research International. 2015. V. 22. N 10. P. 892–896. https://doi.org/10.1016/S1006-706X(15)30086-8
- 33. Shizukawa Y., Hayashi S., Yoneda S., Kondo Y., Tanei H., Ukai S. Mechanism of magnetite seam formation and its role for FeO scale transformation // Oxidation of Metals. 2016. V. 86. N 3-4. P. 315–326. https://doi.org/10.1007/s11085-016-9638-8
- Choi J.W., Choi J.W. Convective heat transfer coefficient for high pressure water jet // ISIJ International. 2002. V. 42. N 3. P. 283–289. https://doi.org/10.2355/isijinternational.42.283
- Грудев А.П., Зильберг Ю.В., Тилик В.Т. Трение и смазки при обработке металлов давлением: Справочник. М.: Металлургия, 1982. 310 с.
- 36. Devoino O.G., Gorbunov A.V., Lapkovsky A.S., Lutsko N.I., Shpackevitch D.A., Gorbunova V.A., Koval V.A. Data sets formation on the physical properties of oxide scale components for theoretical assessment of efficiency parameters of laser cleaning of carbon steels and related processes // Science & Technique. 2024. V. 23. N 3. P. 192–203. https://doi.org/10.21122/2227-1031-2024-23-3-192-203
- 37. Liao Y. Practical Electron Microscopy and Database. 2006. [Электронный ресурс]. URL: https://globalsino.com/EM/page4639. html (дата обращения: 02.12.2024).

magnetic studies. *Geophysical Journal International*, 2009, vol. 177, no. 3, pp. 941–948. https://doi.org/10.1111/j.1365-246x.2009.04122.x

- De Faria D.L.A., Venâncio Silva S., de Oliveira M.T. Raman microspectroscopy of some iron oxides and oxyhydroxides. *Journal* of Raman Spectroscopy, 1997, vol. 28, no. 11, pp. 873–878. https://doi.org/10.1002/(SICI)1097-4555(199711)28:11<873::AID-JRS177>3.0.CO;2-B
- 22. Prokuratov D.S., Davtian A.S., Vereshchagin O.S., Kurganov N.S., Samokhvalov A.A., Pankin D.V., Povolotckaia A.V., Shimko A.A., Mikhailova A.A., Somov P.A., Parfenov V.A. Laser cleaning of archaeologically corroded iron objects with inlays. *Optical and Quantum Electronics*, 2020, vol. 52, no. 2, pp. 113. https://doi. org/10.1007/s11082-020-2231-z
- Veiko V.P., Karlagina Y., Moskvin M., Mikhailovskii V., Odintsova G., Olshin P., Pankin D., Romanov V., Yatsuk R. Metal surface coloration by oxide periodic structures formed with nanosecond laser pulses. *Optics and Lasers in Engineering*, 2017, vol. 96, pp. 63–67. https:// doi.org/10.1016/j.optlaseng.2017.04.014
- Sutton A.T., Kriewall C.S., Leu M.C., Newkirk J.W., Brown B. Characterization of laser spatter and condensate generated during the selective laser melting of 304L stainless steel powder. *Additive Manufacturing*, 2020, vol. 31, pp. 100904. https://doi.org/10.1016/j. addma.2019.100904
- Sánchez-Tovar R., Leiva-García R., García-Antón J. Characterization of thermal oxide films formed on a duplex stainless steel by means of confocal-Raman microscopy and electrochemical techniques. *Thin Solid Films*, 2015, vol. 576, pp. 1–10. https://doi.org/10.1016/j. tsf.2014.12.024
- Encyclopedia of Spectroscopy and Spectrometry. Ed. by Lindon J.C., Tranter G.E., Koppenaal D. Academic Press, 2016, 3584 p.
- Wiemann J., Fabbri M., Yang TR., Stein K, Sander P.M., Norell M.A., Briggs D.E.G. Fossilization transforms vertebrate hard tissue proteins into N-heterocyclic polymers. *Nature Communications*, 2018, vol. 9, pp. 4741. https://doi.org/10.1038/s41467-018-07013-3
- Chen R., Yuen W. Oxidation of low-carbon, low-silicon mild steel at 450–900°C under conditions relevant to hot-strip processing. *Oxidation of Metals*, 2002, vol. 57, no. 1-2, pp. 53–79. https://doi. org/10.1023/A:1013390628475
- Chen Rex, Yeun W.Y.D. Review of the High-Temperature Oxidation of Iron and Carbon Steels in Air or Oxygen. Oxidation of Metals, 2003, vol. 59, no. 5-6, pp. 433-468. https://doi. org/10.1023/A:1023685905159
- 30. Teplyakov Yu.N. Disintegration of the wustite, intrance in composition iron oxidation. *Bulletin of the South Ural State University. Series Chemistry*, 2009, no. 23 (156), pp. 36–42. (in Russian)
- Naipinij S., Sukieum S., Namprai R., Nilsonthi T. Formation of thermal oxide scale and its adhesion to hot-rolled low carbon steels with different final strip thicknesses. *E3S Web of Conferences*, 2022, vol. 355, pp. 02008. https://doi.org/10.1051/e3sconf/202235502008
- Cao G., Wu T., Xu R., Li, Z., Wang F., Liu Z. Effects of coiling temperature and cooling condition on transformation behavior of Tertiary oxide scale. *Journal of Iron and Steel Research International*, 2015, vol. 22, no. 10, pp. 892–896. https://doi.org/10.1016/S1006-706X(15)30086-8
- 33. Shizukawa Y., Hayashi S., Yoneda S., Kondo Y., Tanei H., Ukai S. Mechanism of magnetite seam formation and its role for FeO scale transformation. *Oxidation of Metals*, 2016, vol. 86, no. 3-4, pp. 315– 326. https://doi.org/10.1007/s11085-016-9638-8
- Choi J.W., Choi J.W. Convective heat transfer coefficient for high pressure water jet. *ISIJ International*, 2002, vol. 42, no. 3, pp. 283– 289. https://doi.org/10.2355/isijinternational.42.283
- Grudev A.P. Friction and Lubrication in Metal Forming. Handbook. Moscow, Metallurgija, 1982, 310 p. (in Russian)
- 36. Devoino O.G., Gorbunov A.V., Lapkovsky A.S., Lutsko N.I., Shpackevitch D.A., Gorbunova V.A., Koval V.A. Data sets formation on the physical properties of oxide scale components for theoretical assessment of efficiency parameters of laser cleaning of carbon steels and related processes. *Science & Technique*, 2024, vol. 23, no. 3, pp. 192–203. https://doi.org/10.21122/2227-1031-2024-23-3-192-203
- Liao Y. Practical Electron Microscopy and Database. 2006. Available at: https://globalsino.com/EM/page4639.html (accessed: 02.12.2024).

Авторы

Журба Данила Владимирович — младший научный сотрудник, ООО «НПП ВОЛО», Санкт-Петербург, 199034, Российская Федерация; аспирант, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, https://orcid.org/0009-0001-6814-1737, zhurba. danila306@ya.ru

Журба Владимир Михайлович — генеральный директор, ООО «НПП ВОЛО», Санкт-Петербург, 199034, Российская Федерация, 5076349300, https://orcid.org/0009-0006-9919-3106, vlzhurba@ yandex.ru

Вейко Вадим Павлович — доктор технических наук, профессор, главный научный сотрудник, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, sc 7005095644, https://orcid.org/0000-0001-6071-3449, vadim.veiko@mail.ru

Панькин Дмитрий Васильевич — кандидат физико-математических наук, специалист, Санкт-Петербургский государственный университет, Санкт-Петербург, 199034, Российская Федерация, sc 57190487236, https://orcid.org/0000-0003-3114-3410, dmitrii.pankin@ spbu.ru

Жуков Михаил Валерьевич — кандидат технических наук, научный сотрудник, Институт аналитического приборостроения РАН, Санкт-Петербург, 198095, Российская Федерация, sc 55800160700, https:// orcid.org/0000-0003-3361-6947, cloudjyk@yandex.ru

Пуйша Александр Эдуардович — кандидат технических наук, начальник научно-технического отдела, ООО «НПП ВОЛО», Санкт-Петербург, 199034, Российская Федерация, https://orcid.org/0009-0000-6589-6723, pujsha@mail.ru

Статья поступила в редакцию 05.07.2024 Одобрена после рецензирования 21.02.2025 Принята к печати 30.03.2025

Authors

Danila V. Zhurba — Junior Researcher, NPP VOLO LLC, Saint Petersburg, 199034, Russian Federation; PhD Student, ITMO University, Saint Petersburg, 197101, Russian Federation, https://orcid.org/0009-0001-6814-1737, zhurba.danila306@ya.ru

Vladimir M. Zhurba — CEO, NPP VOLO LLC, Saint Petersburg, 199034, Russian Federation, sc 57076349300, https://orcid.org/0009-0006-9919-3106, vlzhurba@yandex.ru

Vadim P. Veiko — D.Sc., Professor, Chief Researcher, ITMO University, Saint Petersburg, 197101, Russian Federation, sc 7005095644, https://orcid.org/0000-0001-6071-3449, vadim.veiko@mail.ru

Dmitrii V. Pankin — PhD (Physics & Mathematics), Specialist, St. Petersburg State University (SPbSU), Saint Petersburg, 199034, Russian Federation, sc 57190487236, https://orcid.org/0000-0003-3114-3410, dmitrii.pankin@spbu.ru

Mikhail V. Zhukov — PhD, Scientific Researcher, Institute for Analytical Instrumentation of the Russian Academy of Sciences (IAI RAS), Saint Petersburg, 198095, Russian Federation, sc 55800160700, https://orcid. org/0000-0003-3361-6947, cloudjyk@yandex.ru

Alexander E. Puisha — PhD, Head of the Scientific and Technical Department, NPP VOLO LLC, Saint Petersburg, 199034, Russian Federation, https://orcid.org/0009-0000-6589-6723, pujsha@mail.ru

Received 05.07.2024 Approved after reviewing 21.02.2025 Accepted 30.03.2025



Работа доступна по лицензии Creative Commons «Attribution-NonCommercial» **I/İTMO**

НАУЧНО-ТЕХНИЧЕСКИЙ ВЕСТНИК ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ, МЕХАНИКИ И ОПТИКИ март–апрель 2025 Том 25 № 2 http://ntv.ifmo.ru/ SCIENTIFIC AND TECHNICAL JOURNAL OF INFORMATION TECHNOLOGIES, MECHANICS AND OPTICS March–April 2025 Vol. 25 No 2 http://ntv.ifmo.ru/en/ ISSN 2226-1494 (print) ISSN 2500-0373 (online)

ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ, МЕХАНИКИ И ОПТИКИ

университет итмо

doi: 10.17586/2226-1494-2025-25-2-190-198 УДК 538.975

Моделирование нагрева пленки селенида свинца непрерывным лазерным излучением с учетом процесса окисления

Максим Михайлович Сергеев¹, Александра Евгеньевна Пушкарева², Виктория Хассан³

1,2,3 Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация

¹ maxim.m.sergeev@gmail.com, https://orcid.org/0000-0003-2854-9954

² aepushkareva@itmo.ru^{\exists}, https://orcid.org/0000-0003-0082-984X

³ victoriahas566@gmail.com, https://orcid.org/0009-0005-4571-6461

Аннотация

Введение. Исследованы особенности нагрева пленки селенида свинца (PbSe) при воздействии непрерывным лазерным излучением с учетом зарождения и увеличения толщины оксидной фазы. Показано, что окисление пленки PbSe приводит к снижению скорости нагревания и падению максимальной температуры за счет уменьшения доли поглощенного лазерного излучения в оксидной фазе. Результаты моделирования, представленные в работе, стали обоснованием ранее полученных экспериментальных результатов. Объяснение механизма лазерного нагревания пленки PbSe впервые позволило определить наиболее эффективную длительность лазерного воздействия, обеспечивающую формирование структур с заданными оптическими характеристиками. Метод. Исследование выполнено методом аналитического моделирования. Для описания теплового источника использовалось частное решение уравнения теплопроводности. Оптические свойства пленки описывались на основе формул Френеля для отражения и пропускания света. На основании полученных ранее экспериментальных данных создана аналитическая модель, описывающая тепловой источник в пленке с учетом изменения ее оптических свойств за счет формирования слоя оксида PbSe и увеличения его толщины. Основные результаты. Показано, что при воздействии на пленку PbSe непрерывным лазерным излучением с длиной волны 405 нм, вследствие образования слоя оксида, происходит снижение коэффициента экстинкции пленки k_f со значения 0,488 до величины 1,62 \cdot 10⁻³. При этом показатель преломления пленки n_f также снижается со значения 3,532 до 1,925. Коэффициент поглощения пленки на длине волны лазерного излучения за время облучения уменьшается с 0,68 до 0,03. По мере увеличения толщины оксидной фазы от 0 до 600 нм происходит замедление роста температуры в зоне облучения и смещение максимального значения температуры с поверхности к границе пленка-подложка. При воздействии непрерывным лазерным излучением с плотностью мощности около 340 BT/см² в течение 9 с максимальная температура пленки не превышает 275 °С. Обсуждение. Полученные результаты могут быть применены при разработке фотодетекторов на основе пленок PbSe для работы в среднем инфракрасном диапазоне спектра. Лазерный отжиг пленки позволяет локально и контролируемо изменять оптические и электрические характеристики пленки PbSe в небольшом диапазоне значений, что влияет на фоточувствительность пленки, используемой в качестве детектора среднего и дальнего инфракрасного излучения.

Ключевые слова

непрерывное лазерное излучение, пленка селенида свинца (PbSe), модификация оптических свойств, окисление, моделирование теплового источника, аналитическое моделирование

Благодарности

Исследование финансировалось за счет гранта Российского научного фонда и гранта Санкт-Петербургского научного фонда (проект № 23-29-10081).

Ссылка для цитирования: Сергеев М.М., Пушкарева А.Е., Хассан В. Моделирование нагрева пленки селенида свинца непрерывным лазерным излучением с учетом процесса окисления // Научнотехнический вестник информационных технологий, механики и оптики. 2025. Т. 25, № 2. С. 190–198. doi: 10.17586/2226-1494-2025-25-2-190-198

^{© ©} Сергеев М.М., Пушкарева А.Е., Хассан В., 2025

Simulation of lead selenide film heating by CW laser radiation taking into considering oxidation process

Maksim M. Sergeev¹, Alexandra E. Pushkareva^{2⊠}, Viktoria Hassan³

^{1,2,3} ITMO University, Saint Petersburg, 197101, Russian Federation

¹ maxim.m.sergeev@gmail.com, https://orcid.org/0000-0003-2854-9954

² aepushkareva@itmo.ru^{\[\]}, https://orcid.org/0000-0003-0082-984X

³ victoriahas566@gmail.com, https://orcid.org/0009-0005-4571-6461

Abstract

The heating characteristics of a lead selenide (PbSe) film under continuous laser radiation were investigated, accounting for the nucleation and increase in the thickness of an oxide phase layer. It is shown that oxidation of the PbSe film reduces the heating rate and lowers the maximum temperature due to a decrease in the fraction of absorbed laser radiation within the oxide phase. The modeling results presented in this work substantiate earlier experimental findings. For the first time, the explanation of the laser heating mechanism of the PbSe film enabled determining the most effective laser exposure duration to form structures with specified optical characteristics. The study was conducted using analytical modeling. A particular solution of the heat conduction equation was employed to describe the heat source. The optical properties of the film were characterized using Fresnel equations for light reflection and transmission. Based on previously obtained experimental data, an analytical model was developed to describe the heat source in the film, considering changes in its optical properties due to the formation of a lead selenide oxide layer and its increasing thickness. The findings show that, when the PbSe film is exposed to continuous laser radiation with a wavelength of 405 nm, the extinction coefficient of the film, k_{f} decreases from 0.488 to 1.62 $\cdot 10^{-3}$ due to the formation of an oxide layer. In this case, the refractive index of the film, n_{f_2} similarly decreases from 3.532 to 1.925. The film absorption coefficient at the laser wavelength decreases from 0.68 to 0.03 during irradiation. As the thickness of the oxide phase increases from 0 to 600 nm, the temperature growth in the irradiated zone slows down, and the maximum temperature shifts from the surface toward the film-substrate interface. When exposed to continuous laser radiation with a power density of about 340 W/cm² for 9 s, the maximum film temperature does not exceed 275 °C. The obtained results can be applied in the development of mid-infrared spectrum photo detectors based on PbSe film. Laser annealing of the film allows local and controlled changes in the optical and electrical properties of the PbSe film within a narrow range of values, thereby influencing the photo sensitivity of the film used as a detector for mid- and far-infrared radiation.

Keywords

continuous wave laser, laser heating, lead selenide (PbSe) films, optical properties modification, oxidation, heat source simulation, analytical modeling

Acknowledgements

This research was funded by the Russian Science Foundation grant and a grant from the St. Petersburg Science Foundation (project no. 23-29-10081).

For citation: Sergeev M.M., Pushkareva A.E., Hassan V. Simulation of lead selenide film heating by CW laser radiation taking into considering oxidation process. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 2025, vol. 25, no. 2, pp. 190–198 (in Russian). doi: 10.17586/2226-1494-2025-25-2-190-198

Введение

Одним из ключевых свойств халькогенидных пленок является их высокая фоточувствительность в видимой и инфракрасной (ИК) областях спектра, благодаря чему они находят применение в качестве фотоэлементов в микроаналитических устройствах, датчиках и сенсорах различного типа [1]. В частности, пленки селенида свинца (PbSe) широко используются для преобразования ИК излучения в фототок [2]. На базе таких пленок уже давно промышленно изготавливают оптические сенсоры для газового анализа [3]. Недавно появились работы по применению пленок PbSe в качестве детекторов рентгеновского [4] и видимого излучения [5]. Расширение возможных применений обусловлено разнообразными методами нанесения пленки, ее обработки, и выбором подложки, на которую наносится PbSe.

Неотъемлемым этапом обработки пленок PbSe является тепловой отжиг в печи, при котором происходит окончательная коррекция оптических и электрических характеристик пленки, а также ее фиксация на подложке [6]. Тепловая обработка на воздухе приводит

к окислению и играет ключевую роль в достижении определенных параметров пленки как конечного фотоэлемента с необходимой фоточувствительностью [7]. Альтернативой тепловой обработке в печи стала лазерная модификация структуры пленки, которая имеет ряд преимуществ [8]. Модификация структуры пленки PbSe при лазерном воздействии как непрерывным излучением, так и наносекундными импульсами, отличается локальностью воздействия и экстремальными условиями нагревания: градиентом температур около 104 °С/мм, а также скоростью нагревания и охлаждения, варьируемой от 10 до 10² °C/с и выше. Такие условия теплового воздействия существенно отличаются от традиционной обработки в печи, где отсутствует градиент температур, а скорость нагревания редко превышает 0,6 К/с. Нагревание пленки PbSe при лазерном воздействии на воздухе также активирует процесс окисления, влияющий на ее структуру и, следовательно, на поглощательную способность и электрическую проводимость [9]. Именно повышение поглощения ИК излучения при сохранении или увеличении электрической проводимости способствует повышению фоточувствительности пленки. Сквозное окисление пленки PbSe приводит к

снижению поглощения в видимом диапазоне длин волн и к снижению электрической проводимости [8]. По этой причине сквозное окисление пленки считается браком, и такие образцы не используются в качестве фотоэлементов. При тепловой обработке крайне важным является достижение частичного окисления пленки, при котором еще сохраняется электрическая проводимость. Прогнозирование результатов лазерного отжига пленки PbSe осложняется тем, что при ее окислении изменяются не только теплофизические характеристики, но и поглощательная способность, что, в свою очередь, приводит к изменению доли поглощенного излучения и, следовательно, температуры в зоне облучения [10]. Моделирование теплового источника, возникающего в зоне лазерного облучения, становится одним из эффективных способов анализа механизмов лазерной модификации структуры пленок. Таким образом, исследование механизмов лазерного нагрева и окисления пленки PbSe является перспективным и крайне важным для разработки методов управляемой коррекции оптических и электрических характеристик пленок путем подбора режимов облучения.

В настоящей работе изучено влияние окисления пленки PbSe на условия лазерного воздействия непрерывным излучением по мере роста толщины оксидной фазы методом аналитического моделирования. Показано, что по мере увеличения толщины оксидной фазы происходит замедление роста температуры в зоне облучения и смещение ее максимума с поверхности к границе пленка-подложка. На основании результатов моделирования описан механизм окисления пленки с последующим ее просветлением, когда оксидная фаза достигает границы пленка-подложка.

Описание модели

Моделирование температуры в области лазерного воздействия. С применением математического моделирования в работе исследовался локальный нагрев халькогенидных пленок PbSe толщиной $h_f = 0,6$ мкм непрерывным лазерным излучением с длиной волны 405 нм, мощностью $P_0 = 60$ мВт, диаметром пятна 150 мкм и плоским профилем интенсивности [11]. Образцы пленок были нанесены на подложку в виде плоскопараллельных пластин покровного стекла марки К8 с размерами $20 \times 20 \times 0,2$ мм.

Для моделирования динамики температуры в области лазерного воздействия использовалась физико-математическая модель теплового источника, возникающего при непрерывном воздействии излучения, предложенная в работе [12]:

$$T(x, y, z) =$$

$$= \int_{0}^{\infty} \frac{A(z, \tau) P_0 I(z, \tau) \tau}{128 \kappa(z) R_{hs}} dT x(x, \tau) dT y(y, \tau) dT z(z, \tau) d\tau,$$
(1)

где

$$dTx(x, \tau) = \operatorname{erfc}\left(\frac{x-\omega}{R_{hs}\tau}\right) - \operatorname{erfc}\left(\frac{x+\omega}{R_{hs}\tau}\right);$$

$$dTy(y, \tau) = \operatorname{erfc}\left(\frac{y-\omega}{R_{hs}\tau}\right) - \operatorname{erfc}\left(\frac{y+\omega}{R_{hs}\tau}\right)$$
$$dTz(z, \tau) = \operatorname{erfc}\left(\frac{z-\omega}{R_{hs}\tau}\right) - \operatorname{erfc}\left(\frac{z+\omega}{R_{hs}\tau}\right);$$
$$R_{hs} = \sqrt[3]{\omega^2 h(\tau)};$$

T — температура; τ — время; $A(z, \tau)$ — поглощательная способность материала пленки; $I(z, \tau)$ — распределение интенсивности лазерного излучения в материале; R_{hs} — приведенный размер прогреваемой области; ω — радиус пятна; $h(\tau)$ — толщина прогреваемого слоя, равная значению *z*, при котором интенсивность излучения понижается в *e* раз; $\kappa(z)$ — значение теплопроводности материала¹.

Зависимость величин, входящих в выражение (1) от координаты z и времени обуславливалась появлением слоя оксида селенида свинца (PbSeO_x) и увеличением его толщины в процессе облучения. Поглощательная способность материала пленки определялась следующим образом:

$$A(z, \tau) = \begin{cases} A_{ox}, \ 0 \le z \le h_{ox}(\tau) \\ A_{ps}, \ h_{ox}(\tau) < z \le h_f, \\ 0, \ z > h_f \end{cases}$$

где A_{ox} и A_{ps} — поглощательные способности оксидной фазы (PbSeO_x) и PbSe; h_{ox} — толщина PbSeO_x.

Распределение интенсивности излучения в материале задавалось в относительных единицах в соответствии с законом Бугера–Лаберта–Берра:

$$I(z, \tau) = \begin{cases} \exp(-4\pi k_{ox} z/\lambda), \ 0 \le z \le h_{ox}(\tau) \\ \exp(-4\pi k_{ox} h_{ox}(\tau)/\lambda) \exp(-4\pi k_{ps}(z - h_{ox}(\tau))/\lambda) \\ h_{ox}(\tau) < z \le h_f \\ 0, \ z > h_f \end{cases},$$

где k_{ps} и k_{ox} — коэффициенты экстинкции исходной пленки PbSe и PbSeO_x. Коэффициент теплопроводности материала задавался отдельно для пленки и подложки:

$$\kappa(z) = \begin{cases} \kappa_{ps}, \ 0 \le z \le h_f \\ \kappa_{sub}, \ z > h_f \end{cases},$$

где к_{рs} и к_{sub} — коэффициенты теплопроводности PbSe и подложки.

- В основу модели легли следующие приближения:
- излучение лазера поглощалось материалом пленки и не поглощалось материалом подложки;
- окисление PbSe незначительно влияло на значение теплопроводности материала;
- теплоотвод от нагретой области происходил в окружающую среду (воздух) посредством теплоотдачи на границах (коэффициент теплоотдачи на границах с воздухом принят равным 25 Вт/(м²·K));
- распределение энергии в пучке равномерно по всей его площади;
- изменение общей толщины пленки в процессе лазерного воздействия отсутствовало.

¹ Физические величины: Справочник. Под ред. Григорьева И.С., Мейлихова Е.З. М., Энергоатомиздат, 1991. 1232 с.

Расчет оптических характеристик пленки. Для моделирования теплового источника потребовалось определить значение поглощательной способности пленки на длине волны лазерного излучения и ее изменение во время облучения. С этой целью пленка PbSe условно разделялась на два слоя, а для описания ее оптических характеристик использовались данные эксперимента, описанного в [11]. Верхний слой представлял собой PbSeO_x, для которого характерны собственные значения показателя преломления $n_{ox}(\lambda, \tau)$ и коэффициента экстинкции $k_{ox}(\lambda, \tau)$. Нижний слой оставался в виде исходной пленки PbSe с $n_{ps}(\lambda, \tau)$ и $k_{DS}(\lambda, \tau)$. По мере увеличения длительности облучения происходил рост PbSeO_x, что приводило к смещению границы между слоями вниз за счет увеличения толщины PbSeO_r. К окончанию облучения наблюдалось сквозное окисление пленки, что соответствовало исчезновению нижнего слоя из PbSe.

Моделирование оптических характеристик пленки до и после лазерного облучения с различным временем экспозиции осуществлялось с применением аналитических выражений на основе формул Френеля для отражения и пропускания света [13]. Изменение показателя преломления и экстинкции пленки связывалось с окислением и структурно-фазовыми превращениями в ней. До облучения пленка PbSe была не прозрачна в оптическом видимом диапазоне длин волн, поэтому ее пропускание отсутствовало, а показатель преломления вычислялся через выражение для отражения $R_{0m} = [(1 - n_{ps})^2 + k_{ps}^2]/[(1 + n_{ps})^2 + k_{ps}^2]$, преобразованное до следующего вида:

$$n_{ps} = \frac{1 + R_{0m}}{1 - R_{0m}} \pm \sqrt{\left(\frac{1 + R_{0m}}{1 - R_{0m}}\right)^2 - (k_{ps}^2 + 1)} .$$
(2)

Коэффициент экстинкции, k_{ps} , определялся через ослабление света в скин-слое толщиной δ_s по закону Бугера–Ламберта–Берра:

$$k_{ps} = \frac{\lambda}{4\pi\delta_s} \ln\left(\frac{1}{1 - R_{0m}}\right),\tag{3}$$

где δ_s — толщина скин-слоя на длине волны излучения λ . Величина δ_s определялась через удельную электропроводность пленки σ , равную 5,28·10⁵ Ом⁻¹м⁻¹ [2]:

$$\delta_s = \sqrt{\frac{\lambda}{\pi c \mu_0 \sigma}},\tag{4}$$

где величина $\mu_0 = 4\pi \cdot 10^{-7}$ H/A² — магнитная постоянная; $c = 3 \cdot 10^8$ м/с — скорость света в вакууме.

В итоге поглощательная способность пленки *A* на длине волны лазерного излучения рассчитывалась по формуле

$$A = (1 - R_m)[1 - \exp(-4\pi kh/\lambda)],$$
 (5)

где *k* — коэффициент экстинкции пленки; *h* — ее толщина.

Толщина PbSeO_x *h*_{ox} в зависимости от длительности облучения вычислялась через отражение пленки и ее

оптические константы, рассчитанные по выражениям (2) и (3), в предположении, что рост толщины $PbSeO_x$ происходил по экспоненциальному закону, при помощи следующего выражения:

$$R(h_{ox}) = \frac{(n_{pb} - 1)^2 + k_{pb}^2}{(n_{pb} + 1)^2 + k_{pb}^2} \exp\left(-2\frac{h_{ox}}{h_f}\right) + \frac{(n_{ox} - 1)^2 + k_{ox}^2}{(n_{ox} + 1)^2 + k_{ox}^2} \left[\operatorname{erfc}\left(-0.55\frac{h_{ox}}{h_f}\right) - 1 \right].$$
(6)

Результаты и обсуждение

Согласно результатам эксперимента, приведенного в [11], лазерное воздействие на пленку в первые моменты времени способствовало локальному нагреву с облучаемой поверхности и активации процесса окисления, что становилось причиной изменения поглощательной способности пленки (рис. 1). В отличие от пленки PbSe, ее оксидная фаза имела меньшую поглощательную способность и большую глубину проникновения излучения. Пленка PbSeO_x толщиной менее 1 мкм имела частичное пропускание в видимом спектральном диапазоне. В первые 3 с от начала облучения в центре зоны воздействия начиналось изменение цвета пленки, где возникали цвета побежалости, свидетельствующие о наличии тонкого оксидного слоя, прозрачного для видимого света с толщиной, меньше ее длины волны. Дальнейшее увеличение времени экспозиции приводило к разрастанию цветной области из центра на периферию зоны облучения, что свидетельствовало об увеличении толщины оксидного слоя. Спустя 9-12 с от начала облучения пленка начинала частично пропускать свет осветительной системы микроскопа, что указывало на сквозное окисление пленки.

Спустя различные времена экспозиции регистрировалось изменение отражения пленки $R(\tau)$ на длине волны лазерного излучения (рис. 2). Отражение резко снижалось в первые $4,5 \pm 0,5$ с от начала облучения. При времени экспозиции более 6 с изменения отражения не наблюдалось и сохранялось на одинаковом уровне вплоть до 11 ± 1 с. Начиная с 9 с от начала облучения достигалось сквозное окисление, отражение при этом начинало снижаться.

Толщина скин-слоя в пленке PbSe на длине волны лазерного излучения, рассчитанная по выражению (4), составляла $\delta_s = 25,4$ нм. Отражение пленки PbSe равно $R_{0m} = 0,32$ (рис. 2), следовательно, коэффициент экстинкции, вычисленный по выражению (3), равнялся $k_{ps} = 0,488$. Показатель преломления пленки, определенный по выражению (2), достигал значения $n_{ps} = 3,532$, что согласуется с показателем преломления PbSe в оптическом видимом спектральном диапазоне, взятым из [14].

После начала облучения отражение снижалось за счет окисления пленки с поверхности. Часть излучения отражалась от оксидного слоя, другая часть проходила через этот слой, ослаблялась в нем, отражалась от границы между слоями и выходила из оксидного слоя



Рис. 1. Описание механизма окисления пленки PbSe. Изображения пленки PbSe, полученные с помощью оптической микроскопии в светлом поле: отраженного света после лазерной обработки в течение 3, 6, 9 и 12 с (*a*, *b*, *c*, *e*); проходящего света после лазерной обработки пленок в течение 9 и 12 с (*d*, *f*).

 R_{ps} — отражение от пленки PbSe перед облучением, R_{ox} — отражение от слоя PbSeO_x, R_{ps}^* — отражение от границы между PbSeO_x и PbSe

Fig. 1. Description of the oxidation mechanism of the PbSe film. Optical microscopy images of PbSe film obtained in the bright field of reflected light after 3, 6, 9 and 12 s of laser exposure (a, b, c, e), optical microscopy images of PbSe film obtained in the bright field of transmitted light after 9 µ 12 s of laser exposure (d, f).

 R_{ps} — reflection from the PbSe film before irradiation, R_{ox} — reflection from the PbSeO_x layer, R_{ps}^* — reflection from the boundary between PbSeO_x μ PbSe

обратно. После 9 с от начала облучения достигалось сквозное окисление. Пленка в этом случае приобретала частичное пропускание. После окисления удельная проводимость PbSe снижалась до $\sigma = 78 \text{ Om}^{-1}\text{m}^{-1}$ [15]. Глубина проникновения излучения, рассчитанная по (4), в этом случае увеличивалась до $\delta_{sox} = 2,1$ мкм. Это значение превышало более чем в три раза толщину пленки, что и свидетельствовало о возникновении пропускания у пленки при окончании облучения.



Рис. 2. Изменение отражения пленки PbSe на длине волны лазерного излучения [11]

Fig. 2. Changes in the reflection of the PbSe film at the laser radiation wavelength [11]

Ввиду малых размеров зоны лазерной модификации структуры пленки и высокого светорассеяния в этой области, измеренное пропускание PbSeO_x было на уровне приборной погрешности. По этой причине измерить спектральное пропускание пленки не представлялось возможным. Было принято решение коэффициент экстинкции вычислить через выражение (3) для случая $R_m = 0,1$ (рис. 2) и $\delta_{sox} = 2,1$ мкм. В итоге, на длине волны лазерного излучения коэффициент экстинкции пленки в режиме сквозного окисления был принят равным $k_{ox} = 1,62 \cdot 10^{-3}$. Показатель преломления пленки, вычисленный по (2), снижался при окислении до значения $n_{ox} = 1,925$.

Поглощательная способность пленки, вычисляемая по (5), за время лазерного облучения уменьшалась с 0,68 (соответствует A_{ps} PbSe) до 0,03 (соответствует A_{ox} PbSeO_x).

С помощью вычисленных оптических констант для PbSe и PbSeO_x, а также полученных в эксперименте значений отражения в зависимости от длительности облучения, было оценено изменение толщины слоя PbSeO_x при помощи выражения (6) (таблица).

Согласно вычислениям, толщина $PbSeO_x$ увеличивалась со временем облучения, тогда как толщина нижнего слоя PbSe, соответственно, уменьшалась (рис. 3). Изменение толщин происходило достаточно равномерно, и к 9 с толщина $PbSeO_x$ достигала толщины пленки, а слой PbSe полностью исчезал.

Для удобства дальнейших расчетов зависимость толщины PbSeO_x от времени была аппроксимирова-

Таблица. Изменение толщины PbSeO_x по мере изменения отражения пленки

Table.	Changes	in	PbSeO _x	thickness	corresp	ponding	to	variations
			in	film reflec	tion			

Время экспозиции, с	Отражение пленки, %	Толщина пленки, мкм
0	0,32	0,000
3	0,17	0,236
4	0,15	0,295
5	0,13	0,375
6	0,12	0,425
8	0,11	0,500
9	0,10	0,600

на полиномом третьей степени: $h_{ox}(\tau) = 101,44028\tau - 8,0038\tau^2 + 0,46\tau^3$. Появление оксидного слоя на поверхности пленки и увеличение его толщины в процессе обработки существенно изменили характер распределения температуры по глубине материала (рис. 4).

Основное тепловыделение в процессе воздействия лазерным излучением происходило на границе PbSeO_x вследствие более интенсивного поглощения излучения PbSe ($A_{ps} = 0,68$) по сравнению с PbSeO_x ($A_{ox} = 0,03$). Плавное возрастание температуры в слое PbSeO_x от поверхности вглубь обусловлено оттоком тепла от нагретой границы PbSeO_x-PbSe посредством теплопроводности. В слое PbSe под PbSeO_x наблюдался достаточно резкий спад температуры вследствие сильного ослабления интенсивности излучения за счет поглощения. Косвенно, на такой характер нагревания указывала граница модифицированной области в пленке (рис. 1). По снимкам оптической микроскопии видно, что с увеличением экспозиции облучения, размер модифицированной области не увеличивался, а оставался на одном и том же уровне. При этом модифицированную область окружало кольцо в виде пленки с измененным показателем преломления (более светлая область, чем исходная пленка), которая с увеличением длительности облучения становилась более выраженной. Подобная



Puc. 3. Расчетные зависимости толщин слоев PbSe и PbSeO_x от времени облучения *Fig. 3.* Calculated dependencies of PbSe and PbSeO_x layer thicknesses vs. irradiation time



Рис. 4. Зависимости температуры от глубины в пленке через время *t*, равное 1, 3, 5, 7 и 9 с. Пунктиром показана граница PbSeO_x в соответствующие моменты времени

Fig. 4. Temperature dependencies vs. depth within the film at time t equal to 1, 3, 5, 7, and 9 s (the dotted lines indicate the oxide layer boundary at the corresponding time points)



Рис. 5. Временные зависимости температуры на верхней границе (z = 0), на границе пленка-подложка (z = 600 нм), максимальной температуры в пленке (T_{max}) и температуры на границе между PbSeO_x и PbSe $(z = h_{ox})$ *Fig.* 5. Time dependencies of temperature at the upper boundary (z = 0), at the film-substrate interface (z = 600 nm), the maximum temperature in the film (T_{max}) , and the temperature at the oxide-lead selenide interface $(z = h_{ox})$

морфология модифицированной области указывала на снижение температуры в зоне облучения и ее выход в стационарный режим. В течение первых 6 с облучения в области границы между PbSeO_x и PbSe наблюдался максимум температуры, который со временем увеличивался и к 6-й секунде достигал примерно 253 °C (рис. 5).

Глубина расположения максимума температуры на 6-й секунде воздействия составляла примерно 420 нм, или приблизительно 2/3 от толщины исходной пленки (рис. 6). Далее, максимум температуры практически не смещался в глубину до конца облучения, несмотря на то что граница между PbSeO_x и PbSe продолжала смещение в сторону подложки. Таким образом, пленка PbSe с толщиной менее 200 нм начинала меньше поглощать падающее лазерное излучение, что и становилось причиной снижения температуры в пленке. Однако



Рис. 6. Временные зависимости толщины h_{ox} слоя PbSeO_x и глубины, на которой расположен максимум температуры $(h_{T_{\rm max}})$

Fig. 6. Time dependencies of the PbSeO_x layer thickness h_{ox} and the depth at which the maximum temperature $h_{T_{\text{max}}}$ is located

достигнутой температуры было достаточно для продолжения процесса окисления. Данный эффект объяснялся снижением интенсивности излучения, достигающего границы между PbSeO_x и PbSe, вследствие увеличения толщины оксидного слоя и интегрального поглощения света в нем, в результате чего тепловыделение на рассматриваемой границе становилось меньше, а ее температура снижалась (рис. 5). Косвенным подтверждением того, что температура на границе пленка–подложка не достигала 250 °C, является тот факт, что подложка покровного стекла не трескалась из-за высокого градиента напряжений, вызванных резким перепадом температур в зоне лазерного облучения [11]. Поскольку разрушений подложки не наблюдалось, то перепад температур в зоне облучения был ниже.

Начиная с 6-й секунды обработки, рост максимальной температуры материала резко замедлялся, и температура на границе между пленкой и подложкой начинала увеличиваться вследствие приближения к ней границы PbSeO_x–PbSe. Наконец, при сквозном окислении пленки PbSe режим нагревания выходил в стационарный режим, так как оптические константы пленки и, следовательно, ее поглощательная способность более не изменялись. По данным экспериментальных исследований [11], морфология пленки в зоне лазерного облучения после 9 с не изменялась, оставались без изменения и размеры модифицированных областей.

Заключение

В результате проведенного исследования морфологии модифицированных областей пленок селенида свинца (PbSe) и их отражения были определены показатель преломления и коэффициент экстинкции пленки и ее оксидной фазы (PbSeO_x). При лазерном воздействии непрерывным излучением показатель преломления пленки изменялся с $n_{ps} = 3,532$ до $n_{ox} = 1,925$, а коэффициент экстинкции уменьшался с $k_{ps} = 0,488$ до $k_{ox} = 1,62 \cdot 10^{-3}$. Причиной изменения оптических констант было окисление пленки. Поглощательная

способность пленки PbSe на длине волны лазерного излучения во время облучения снижалась с 0,68 до 0,03, что оказывало влияние на параметры теплового источника.

С учетом изменения оптических констант пленки в зависимости от времени экспозиции была восстановлена толщина PbSeO_x, значение которой увеличивалось, и к 9-й секунде от начала облучения достигала толщины пленки. Рост оксидной фазы в пленке описывался по экспоненциальному закону, что хорошо согласуется с экспериментальными данными. Основываясь на результатах роста оксидной фазы в пленке, было проведено моделирование теплового источника, оценены максимальная температура ее градиент и смещение максимальной температуры с облучаемой поверхности на границу PbSe-PbSeO_x. Из результатов моделирования видно, что PbSeO_x поглощала меньше лазерного излучения и нагревалась медленнее, чем материал пленки. Из-за этого максимум температуры всегда находился на границе PbSe-PbSeO_x, а скорость роста температуры замедлялась спустя 6 с от начала облучения. К этому моменту времени толщина PbSeO_x составляла около 0,420 мкм, что составляло 2/3 от общей толщины пленки. Далее, скорость повышения температуры на

Литература

- 1. Scheer R., Schock H.W. Chalcogenide photovoltaics: physics, technologies, and thin film devices. Wiley, 2011. 384 p.
- Gupta M.C., Harrison J.T., Islam M.T. Photoconductive PbSe thin films for infrared imaging // Materials Advances. 2021. V. 2. N 10. P. 3133–3160. https://doi.org/10.1039/D0MA00965B
- Tsiulyanu D., Marian S., Liess H-D., Eisele I. Chalcogenide based gas sensors // Journal of Optoelectronics and Advanced Materials. 2003. V. 5. N 5. P. 1349–1354.
- Khusayfan N.M., Khanfar H.K., Alharbi S.R.N. Fabrication and characterization of lead selenide thin film as X-ray sensors, photovoltaic devices and microwave resonators // Applied Physics A. 2023. V. 129. N 9. P. 639. https://doi.org/10.1007/s00339-023-06909-2
- Khusayfan N.M., Qasrawi A.F., Khanfar H.K., Alharbi S.R. Lead selenide thin films designed for laser sensing and visible light communications // Silicon. 2023. V. 15. N 16. P. 6971–6979. https:// doi.org/10.1007/s12633-023-02554-9
- Tretyakova N.A. IR sensitization of PbSnSe films by heat treatment in air // Inorganic Materials. 2017. V. 53. N 10. P. 1005–1008. https:// doi.org/10.1134/s002016851710017x
- Popov V.P., Tikhonov P.A., Tomaev V.V. Investigation into the mechanism of oxidation on the surface of lead selenide semiconductor structures // Glass Physics and Chemistry. 2003. V. 29. N 5. P. 494– 500. https://doi.org/10.1023/a:1026399332312
- Olkhova A.A., Patrikeeva A.A., Sergeev M.M. Electrical and optical properties of laser-induced structural modifications in PbSe films // Applied Sciences. 2022. V. 12. N 19. P. 10162. https://doi. org/10.3390/app121910162
- Olkhova A.A., Patrikeeva A.A., Dubkova M.A., Kuzmenko N.K., Nikonorov N.V., Sergeev M.M. Comparison of CW NUV and pulse NIR paser influence on PbSe films photosensitivity // Applied Sciences. 2023. V. 13. N 4. P. 2396. https://doi.org/10.3390/ app13042396
- Сергеев М.М., Пушкарева А.Е., Гресько В.Р. Моделирование теплового источника в пленке с наночастицами при воздействии ультракоротких лазерных импульсов // Известия высших учебных заведений. Приборостроение. 2024. Т. 67. № 7. С. 567–573. https://doi.org/10.17586/0021-3454-2024-67-7-567-573
- Olkhova A., Patrikeeva A., Butyaeva M., Pushkareva A., Hassan V., Omelchenko P., Shulga B., Kozodaev D., Korepanov O., Bogdanov K., Pestov G., Sergeev M. Dynamics of lead selenide films

границе PbSe–PbSeO_x снижалась, из-за чего замедлялся и рост оксидной фазы. Причиной таких особенностей нагрева является повышение влияния оттока тепла в подложку при приближении оксидной фазы к ней. Уменьшение толщины пленки PbSe приводило к снижению поглощения лазерного излучения, что и становилось причиной замедления скорости нагревания пленки. При достижении сквозного окисления температура в зоне облучения нормализировалась и выходила в стационарный режим.

Результаты моделирования позволили объяснить ряд важных процессов, происходящих в пленке PbSe при лазерном облучении, в частности влияние окисления на оптические свойства пленки и температуру в зоне облучения. Поскольку отжиг пленок при лазерном облучении планируется использовать как альтернативу тепловой обработки в печи, то длительность облучения и плотность мощности падающего излучения станут ключевыми параметрами в подборе оптимальных условий обработки. Именно от режима облучения будет зависеть толщина оксидной фазы пленки и, следовательно, такие важные параметры фотодетекторов на их основе, как поглощательная способность и электронная проводимость.

References

- 1. Scheer R., Schock H.W. Chalcogenide Photovoltaics: Physics, Technologies, and Thin Film Devices. Wiley, 2011, 384 p.
- Gupta M.C., Harrison J.T., Islam M.T. Photoconductive PbSe thin films for infrared imaging. *Materials Advances*, 2021, vol. 2, no. 10, pp. 3133–3160. https://doi.org/10.1039/D0MA00965B
- Tsiulyanu D., Marian S., Liess H-D., Eisele I. Chalcogenide based gas sensors. *Journal of Optoelectronics and Advanced Materials*, 2003, vol. 5, no. 5, pp. 1349–1354.
- Khusayfan N.M., Khanfar H.K., Alharbi S.R.N. Fabrication and characterization of lead selenide thin film as X-ray sensors, photovoltaic devices and microwave resonators. *Applied Physics A*, 2023, vol. 129, no. 9, pp. 639. https://doi.org/10.1007/s00339-023-06909-2
- Khusayfan N.M., Qasrawi A.F., Khanfar H.K., Alharbi S.R. Lead selenide thin films designed for laser sensing and visible light communications. *Silicon*, 2023, vol. 15, no. 16, pp. 6971–6979. https://doi.org/10.1007/s12633-023-02554-9
- Tretyakova N.A. IR sensitization of PbSnSe films by heat treatment in air. *Inorganic Materials*, 2017, vol. 53, no. 10, pp. 1005–1008. https://doi.org/10.1134/s002016851710017x
- Popov V.P., Tikhonov P.A., Tomaev V.V. Investigation into the mechanism of oxidation on the surface of lead selenide semiconductor structures. *Glass Physics and Chemistry*, 2003, vol. 29, no. 5, pp. 494–500. https://doi.org/10.1023/a:1026399332312
- Olkhova A.A., Patrikeeva A.A., Sergeev M.M. Electrical and optical properties of laser-induced structural modifications in PbSe films. *Applied Sciences*, 2022, vol. 12, no. 19, pp. 10162. https://doi. org/10.3390/app121910162
- Olkhova A.A., Patrikeeva A.A., Dubkova M.A., Kuzmenko N.K., Nikonorov N.V., Sergeev M.M. Comparison of CW NUV and pulse NIR paser influence on PbSe films photosensitivity. *Applied Sciences*, 2023, vol. 13, no. 4, pp. 2396. https://doi.org/10.3390/app13042396
- Sergeev M.M., Pushkareva A.E., Gresko V.R. Modeling of a heat source in a film with nanoparticles under the action of ultrashort laser pulses. *Journal of Instrument Engineering*, 2024, vol. 67, no. 7, pp. 567–573. (in Russian). https://doi.org/10.17586/0021-3454-2024-67-7-567-573
- Olkhova A., Patrikeeva A., Butyaeva M., Pushkareva A., Hassan V., Omelchenko P., Shulga B., Kozodaev D., Korepanov O., Bogdanov K., Pestov G., Sergeev M. Dynamics of lead selenide films

modification by laser annealing // Physica B: Condensed Matter. 2025. V. 698. P. 416771. https://doi.org/10.1016/j.physb.2024.416771

- Miyamoto I., Horn A., Gottmann J. Local melting of glass material and its application to direct fusion welding by Ps-laser pulses // Journal of Laser Micro Nanoengineering. 2007. V. 2. N 1. P. 7–14. https://doi.org/10.2961/jlmn.2007.01.0002
- 13. Борн М., Вольф Э. Основы оптики. М.: Наука, 1973. 720 с.
- Palik E.D. Handbook of Optical Constants of Solids. Academic Press, 1985.V. 1. 804 p.
- Jang M.H., Yoo S.S., Kramer M.T., Dhar N.K., Gupta M.C. Electrical transport properties of sensitized PbSe thin films for IR imaging sensors // Semiconductor Science and Technology. 2019. V. 34. N 6. P. 065009. https://doi.org/10.1088/1361-6641/ab19e7

Авторы

Сергеев Максим Михайлович — кандидат технических наук, доцент, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, вс 55624732300, https://orcid.org/0000-0003-2854-9954, maxim.m.sergeev@gmail.com

Пушкарева Александра Евгеньевиа — кандидат технических наук, старший научный сотрудник, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, sc 12791000500, https://orcid.org/0000-0003-0082-984X, aepushkareva@itmo.ru

Хассан Виктория — инженер, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, https://orcid.org/0009-0005-4571-6461, victoriahas566@gmail.com

Статья поступила в редакцию 20.11.2024 Одобрена после рецензирования 18.02.2025 Принята к печати 21.03.2025 modification by laser annealing. *Physica B: Condensed Matter*, 2025, vol. 698, pp. 416771. https://doi.org/10.1016/j.physb.2024.416771

- Miyamoto I., Horn A., Gottmann J. Local melting of glass material and its application to direct fusion welding by Ps-laser pulses. *Journal* of *Laser Micro Nanoengineering*, 2007, vol. 2, no. 1, pp. 7–14. https:// doi.org/10.2961/jlmn.2007.01.0002
- 13. Born M., Wolf E. Principles of Optics. Pergamon Press, 1959, 852 p.
- 14. Palik E.D. Handbook of Optical Constants of Solids. Academic Press, 1985, vol. 1, 804 p.
- Jang M.H., Yoo S.S., Kramer M.T., Dhar N.K., Gupta M.C. Electrical transport properties of sensitized PbSe thin films for IR imaging sensors. *Semiconductor Science and Technology*, 2019, vol. 34, no. 6, pp. 065009. https://doi.org/10.1088/1361-6641/ab19e7

Authors

Maksim M. Sergeev — PhD, Associate Professor, ITMO University, Saint Petersburg, 197101, Russian Federation, Sc 55624732300, https://orcid.org/0000-0003-2854-9954, maxim.m.sergeev@gmail.com

Alexandra E. Pushkareva — PhD, Senior Researcher, ITMO University, Saint Petersburg, 197101, Russian Federation, se 12791000500, https://orcid.org/0000-0003-0082-984X, aepushkareva@itmo.ru

Viktoria Hassan — Engineer, ITMO University, Saint Petersburg, 197101, Russian Federation, https://orcid.org/0009-0005-4571-6461, victoriahas566@gmail.com

Received 20.11.2024 Approved after reviewing 18.02.2025 Accepted 21.03.2025



Работа доступна по лицензии Creative Commons «Attribution-NonCommercial» **I/İTMO**

НАУЧНО-ТЕХНИЧЕСКИЙ ВЕСТНИК ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ, МЕХАНИКИ И ОПТИКИ март–апрель 2025 Том 25 № 2 http://ntv.ifmo.ru/ SCIENTIFIC AND TECHNICAL JOURNAL OF INFORMATION TECHNOLOGIES, MECHANICS AND OPTICS March–April 2025 Vol. 25 No 2 http://ntv.ifmo.ru/en/ ISSN 2226-1494 (print) ISSN 2500-0373 (online)

ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ, МЕХАНИКИ И ОПТИКІ

doi: 10.17586/2226-1494-2025-25-2-199-211 VJIK 535.372: 539.196: 539.349

Синтез и исследование структуры и свойств фотокаталитических нанокомпозитов системы Cu/ZnO-ZnCr₂O₄

Сергей Константинович Евстропьев¹, Андрей Александрович Шелеманов², Николай Валентинович Никоноров³, Анна Владимировна Караваева⁴, Константин Владимирович Дукельский⁵, Григорий Сергеевич Полищук⁶, Марианна Александровна Гаврилова⁷, Ксения Александровна Портнова⁸, Игорь Викторович Багров⁹

1,5,6,9 Научно-производственное объединение Государственный оптический институт им. С.И. Вавилова, Санкт-Петербург, 192171, Российская Федерация

1.7 Санкт-Петербургский государственный технологический институт (технический университет),

Санкт-Петербург, 190013, Российская Федерация

1,2,3,8 Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация

⁴ Санкт-Петербургский государственный химико-фармацевтический университет, Санкт-Петербург, 197022, Российская Федерация

¹ evstropiev@bk.ru, https://orcid.org/0000-0002-0160-8443

² Shelemanov@mail.ru^{\box}, https://orcid.org/0000-0001-5854-9475

³ nikonorov@oi.ifmo.ru, https://orcid.org/0000-0002-1341-067X

⁴ anna.karavaeva@pharminnotech.com, https://orcid.org/0000-0001-8231-6364

⁵ kdukel@mail.ru, https://orcid.org/0000-0002-1627-7499

⁶ g.polishchuk@optrotech.ru, https://orcid.org/0009-0005-9358-6040

⁷ amonrud@yandex.ru, https://orcid.org/0000-0001-6163-9316

⁸ ksiu.san@gmail.com, https://orcid.org/0009-0004-3825-1291

9 i.bagrov2@gmail.com, https://orcid.org/0009-0005-2174-6412

Аннотация

Введение. В настоящее время разработка новых нанокомпозитных материалов с улучшенными фотокаталитическими и антибактериальными свойствами представляет собой актуальную задачу для экологически чистых технологий очистки воды и воздуха. В работе приведены результаты исследования порошковых нанокомпозитов ZnO-ZnCr₂O₄ и Cu/ZnO-ZnCr₂O₄, полученных полимерно-солевым методом. Метод. Для синтеза нанокомпозитов использовали растворы нитратов цинка и хрома с добавлением поливинилпирролидона в качестве растворимого органического полимера. Структура и морфология нанокомпозитов исследованы методами рентгенофазового анализа и электронной микроскопии, оптические и люминесцентные свойства с использованием спектроскопических методов. Основные результаты. В результате термообработки при 550 °С получены дисперсные порошки нанокомпозитов, состоящие из частиц размером несколько микрометров, включающих гексагональные нанокристаллы оксида цинка со средним размером около16 нм и кристаллы шпинели ZnCr₂O₄. В спектре люминесценции композита Cu/ZnO-ZnCr₂O₄ в видимой области наблюдаются полосы флуоресценции, характерные для кристаллов ZnCr₂O₄ и структурных дефектов кристаллов оксида цинка. Установлено, что интенсивность фотогенерации синглетного кислорода нанокомпозитом Cu/ZnO-ZnCr₂O₄ линейно зависит от плотности мощности возбуждающего излучения (длина волны 405 нм). Выявлена антибактериальная активность нанокомпозита Cu/ZnO-ZnCr2O4 в отношении бактерий Staphylococcus aureus АТСС 209Р. Обсуждение. Полученные нанокомпозитные порошки могут быть использованы в системах очистки и обеззараживания воды и воздуха.

Ключевые слова

фотокатализ, нанокомпозит, ZnO, бактерия

[©] Евстропьев С.К., Шелеманов А.А., Никоноров Н.В., Караваева А.В., Дукельский К.В., Полищук Г.С., Гаврилова М.А., Портнова К.А., Багров И.В., 2025

Ссылка для цитирования: Евстропьев С.К., Шелеманов А.А., Никоноров Н.В., Караваева А.В., Дукельский К.В., Полищук Г.С., Гаврилова М.А., Портнова К.А., Багров И.В. Синтез и исследование структуры и свойств фотокаталитических нанокомпозитов системы Cu/ZnO-ZnCr₂O₄ // Научно-технический вестник информационных технологий, механики и оптики. 2025. Т. 25, № 2. С. 199–211. doi: 10.17586/2226-1494-2025-25-2-199-211

Synthesis and study of the structure and properties of photocatalytic nanocomposites of the Cu/ZnO-ZnCr₂O₄ system

Sergey K. Evstropyev¹, Andrey A. Shelemanov^{2⊠}, Nikolay V. Nikonorov³, Anna V. Karavaeva⁴, Konstantin V. Dukelskii⁵, Grigorii S. Polischuk⁶, Marianna A. Gavrilova⁷, Ksenia A. Portnova⁸, Igor V. Bagrov⁹

^{1,5,6,9} JSC S.I. Vavilov State Optical Institute (SOI), Saint Petersburg, 192171, Russian Federation ^{1,7} Saint Petersburg State Technological Institute (Technical University), Saint Petersburg, 190013, Russian Federation

1,2,3,8 ITMO University, Saint Petersburg, 197101, Russian Federation

⁴ Saint Petersburg Chemical-Pharmaceutical University, Saint Petersburg, 197022, Russian Federation

¹ evstropiev@bk.ru, https://orcid.org/0000-0002-0160-8443

² Shelemanov@mail.ru^{\Box}, https://orcid.org/0000-0001-5854-9475

³ nikonorov@oi.ifmo.ru, https://orcid.org/0000-0002-1341-067X

⁴ anna.karavaeva@pharminnotech.com, https://orcid.org/0000-0001-8231-6364

⁵ kdukel@mail.ru, https://orcid.org/0000-0002-1627-7499

⁶ g.polishchuk@optrotech.ru, https://orcid.org/0009-0005-9358-6040

⁷ amonrud@yandex.ru, https://orcid.org/0000-0001-6163-9316

⁸ ksiu.san@gmail.com, https://orcid.org/0009-0004-3825-1291

⁹ i.bagrov2@gmail.com, https://orcid.org/0009-0005-2174-6412

Abstract

Currently, the development of new nanocomposite materials with improved photocatalytic and antibacterial properties is a topical task for environmentally friendly technologies for water and air purification. This paper presents the results of a study of ZnO-ZnCr₂O₄ and Cu/ZnO-ZnCr₂O₄ powder nanocomposites obtained by the polymer-salt method. For the synthesis of nanocomposites, zinc and chromium nitrate solutions with the addition of polyvinylpyrrolidone as a soluble organic polymer were used. The structure and morphology of the nanocomposites were studied by XRD analysis and electron microscopy, optical and luminescent properties - using spectroscopic methods. As a result of heat treatment at 550 °C, dispersed powders of nanocomposites were obtained, consisting of particles several micrometers in size, including hexagonal ZnO nanocrystals with an average size about16 nm and ZnCr₂O₄ spinel crystals. In the luminescence spectrum of the Cu/ZnO-ZnCr₂O₄ composite in the visible region, fluorescence bands are observed characteristic of ZnCr₂O₄ crystals and structural defects of ZnO crystals. It was found that the intensity of singlet oxygen photogeneration by the Cu/ZnO-ZnCr₂O₄ nanocomposite linearly depends on the power density of the exciting radiation (the wavelength is 405 nm). Antibacterial activity of the Cu/ZnO-ZnCr₂O₄ nanocomposite powders can be used in water and air purification and disinfection systems.

Keywords

photocatalysis, nanocomposites, ZnO, bacteria

For citation: Evstropyev S.K., Shelemanov A.A., Nikonorov N.V., Karavaeva A.V., Dukelskii K.V., Polischuk G.S., Gavrilova M.A., Portnova K.A., Bagrov I.V. Synthesis and study of the structure and properties of photocatalytic nanocomposites of the Cu/ZnO-ZnCr₂O₄ system. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 2025, vol. 25, no. 2, pp. 199–211 (in Russian). doi: 10.17586/2226-1494-2025-25-2-199-211

Введение

Все более широкое использование фотокаталитических процессов в различных современных практических приложениях (солнечная энергетика, сенсорика, медицина, экология) определяет актуальность разработки и исследования свойств новых материалов [1–4]. Оксидные полупроводниковые материалы являются одними из наиболее эффективных фотокатализаторов и отличаются высокой стабильностью, химической устойчивостью и технологической простотой их получения [5].

Фотокаталитические процессы происходят на поверхности материалов, и их морфология и величина удельной поверхности оказывают сильное влияние на скорость фотокаталитической реакции [6–9]. Процессы адсорбции органических соединений и микроорганизмов на поверхности материалов играют важную роль в фотокатализе и являются предметом интенсивных исследований [6, 10, 11].

Одними из наиболее эффективных оксидных фотокатализаторов являются материалы на основе оксида цинка (ZnO) [4, 6, 12, 13]. ZnO представляет собой широкозонный полупроводник с шириной запрещенной зоны 3,37 эВ при 300 К [12, 14], и для инициирования фотокаталитических процессов на поверхности ZnO обычно используется ультрафиолетовое (УФ) излучение [3, 15, 16].

Одними из эффективных фотокатализаторов являются композиционные материалы системы ZnO-ZnCr₂O₄ [17]. Нужно отметить, что кристаллы шпинели ZnCr₂O₄, так же, как и ZnO, проявляют высокие фотокаталитические свойства [17–20]. Для расширения спектрального диапазона возбуждающего излучения, необходимого для фотокаталитической реакции, а также для повышения фотокаталитических и бактерицидных свойств материалов на основе ZnO, в их состав вводятся соединения серебра [10, 21, 22] или меди (Cu) [23, 24], а также используются добавки других полупроводниковых оксидов [16, 24, 25]. Введение в состав фотокаталитического материала соединений меди приводит к изменению его электронной структуры и обеспечивает возможность применения для возбуждения фотокатализатора излучения видимого спектрального диапазона [26].

Актуальной задачей является исследование структуры и свойств фотокаталитических и бактерицидных композитов системы Cu/ZnO-ZnCr₂O₄, а также оптимизация методов их формирования. Несмотря на наличие работ, посвященных фотокаталитическим свойствам подобных материалов [17–20, 25], остаются недостаточно изученными взаимосвязи между морфологическими особенностями, фазовым составом и эффективностью фотокаталитических и бактерицидных процессов. Кроме того, актуальной остается задача расширения спектрального диапазона возбуждающего излучения до видимого диапазона.

Для получения фотокаталитических композитов широко применяются низкотемпературные жидкостные методы синтеза: полимерный золь-гель метод [6, 16, 22, 24], осаждение из растворов [12], распылительный пиролиз [17, 27] и другие. Выбор метода синтеза и условия его проведения могут оказывать сильное влияние на структуру и фотокаталитические свойства материалов [3, 28].

Полимерно-солевой метод синтеза наноматериалов является технологически простым, экономичным и широко используется для получения различных наноматериалов [22, 29, 30]. В настоящей работе данный метод выбран для синтеза порошковых нанокомпозитов, поскольку он обеспечивает высокую однородность распределения ионов металлов на молекулярном уровне за счет равномерного смешивания растворов солей металлов и растворимых органических полимеров. Такая особенность способствует получению материалов с контролируемым размером частиц, высокой удельной поверхностью и равномерным фазовым распределением, что является критически важным для оптимизации их фотокаталитических и бактерицидных свойств.

Цель работы — разработка методики синтеза Сисодержащего порошкового нанокомпозита системы $ZnO-ZnCr_2O_4$ и изучение его структуры, люминесцентных, адсорбционных и фотокаталитических свойств.

Материалы и методы

В качестве исходных компонентов нанокомпозита были использованы два водных раствора: нитратов цинка и хрома, органического полимера поливинилпирролидона (ПВП) в пропаноле-2. ПВП широко используется в жидкостных методах формирования различных наночастиц, в том числе ZnO [6, 16] и ZnCr₂O₄ [31]. После смешения двух растворов в заданных объемах полученные композиционные растворы подвергались сушке при 70 °C. Химические составы композиционных растворов, из которых были получены образец 1 и образец 2, приведены в табл. 1.

Полученные в результате сушки полимерно-солевые гели подвергали термообработке на воздухе в электрической лабораторной печи LH 30/12 (Nabertherm, Германия) при температуре 550 °C. В процессе термообработки происходило полное разложение солей металлов и органического полимера, удаление газообразных продуктов [30].

Изучение кристаллической структуры синтезированных нанокомпозитов ZnO-ZnCr₂O₄ и Cu/ZnO-ZnCr₂O₄ выполнено методом рентгенофазового анализа, используя дифрактометр Rigaku Ultima IV (Rigaku Согрогаtion, Япония). Средний размер *d* кристаллов ZnO был рассчитан по формуле Шеррера [32] на основании полученных дифрактограмм.

Морфология и химический состав порошковых образцов нанокомпозитов исследованы методами сканирующей электронной микроскопии и энергодисперсионного анализа на микроскопе TESCAN VEGA3 (TESCAN, Чехия), оснащенного приставкой рентгеноспектрального микроанализа Oxford INCA x-act (Oxford Instruments Analytical, Великобритания). При проведении исследований аналитического химического состава материала измерения выполнялись в трех различных точках каждого образца. В каждой точке проводилось не менее трех повторных измерений для повышения точности и надежности полученных данных. Результаты измерений усреднялись для каждой точки, после чего рассчитывалось среднее значение по всему образцу. Такой подход позволяет минимизировать влияние локальных неоднородностей материала и повысить достоверность результатов анализа.

Измерение спектров фотолюминесценции синтезированных композитов в спектральном диапазоне 230–650 нм выполнено на спектрофлюориметре Perkin Elmer LS50B (PerkinElmer, Inc., США). Спектры люминесценции в ближней ИК области спектра были измерены с использованием спектрометра SDH-IV (SOLAR Laser Systems, Беларусь) при возбуждении излучением

Таблица 1. Химический состав исходного раствора *Table 1*. Chemical composition of the initial solution

Ofmanay	Химический состав раствора, масс.%					
Ооразец	H ₂ O	Пропанол-2	ПВП	$Zn(NO_3)_2$	Cr(NO ₃) ₃	CuSO ₄
1	53,48	40,70	2,58	2,75	0,46	0,03
2	53,51	40,70	2,58	2,75	0,46	

Научно-технический вестник информационных технологий, механики и оптики, 2025, том 25, № 2 Scientific and Technical Journal of Information Technologies, Mechanics and Optics, 2025, vol. 25, no 2

светодиода HPR40E (Huey Jann Electronic, Китай) с максимальной длиной волны равной 405 нм.

Для исследования адсорбционных и фотокаталитических свойств полученных композитов использовались водные растворы органического диазокрасителя Chicago Sky Blue (CSB) (Sigma Aldrich, CША). Выбор этого красителя обусловлен его хорошо изученной химической структурой, стабильностью в водных растворах и выраженной окраской, что позволяет легко отслеживать изменения его концентрации с помощью спектрофотометрических методов. Кроме того, CSB ранее использовался в качестве модельного органического загрязнителя в ряде работ [6, 10, 16, 24], что обеспечивает возможность сопоставления полученных результатов с данными исследований и позволяет более объективно оценить эффективность фотокаталитической активности исследуемых композитов.

Молекулы CSB, находящиеся в водных растворах, интенсивно поглощают свет в красной области спектра с максимумом полосы поглощения при длине волны равной 611 нм. Измерения поглощения растворов на этой длине волны позволяли определять концентрацию красителя в растворах в процессе исследования адсорбционных и фотокаталитических свойств композитов. При проведении этих экспериментов навеска композита массой 0,02 г помещалась в 6 мл водного раствора красителя, находящегося в кювете из кварцевого стекла.

При исследованиях адсорбционных и фотокаталитических свойств композита проводились периодические измерения спектров поглощения растворов с целью определения изменения концентрации красителя CSB.

Адсорбционные свойства определялись как способность материала удерживать молекулы красителя на своей поверхности. Для их оценки проводились эксперименты в темноте, что исключало влияние фотокаталитических процессов. Навеска композита помещалась в раствор красителя, изменения концентрации красителя во времени фиксировались с помощью абсорбционной спектроскопии на спектрофотометре ПЭ-5400УФ (Экрос, Россия). Концентрация рассчитывалась на основе оптической плотности раствора при максимуме поглощения ($\lambda_{max} = 611$ нм). Снижение концентрации в этих условиях отражало исключительно вклад адсорбционного процесса.

Фотокаталитические свойства характеризовались способностью композита инициировать разложение красителя под воздействием УФ излучения. Для оценки этих свойств использовалась ртутная лампа ДРТ-230 (НИИИС имени А.Н. Лодыгина, Россия) $\lambda_{max} = 254$ нм. В ходе эксперимента водный раствор красителя с композитом подвергалась УФ облучению, изменения концентрации красителя также фиксировались на спектрофотометре ПЭ-5400УФ (Экрос, Россия). Проведено исследование фотокаталитических свойств композитов под воздействием излучения на границе мягкого УФ-А и видимого диапазонов. Источником был выбран светодиод HP-3003 (Litlight, Китай) с $\lambda = 390$ –400 нм.

Для оценки антибактериальной активности полученных композитов применен метод, подробно описанный в работе [8]. В качестве тест-организма использовалась бактерия *Staphylococcus aureus ATCC 209P*, выбранная в связи с ее клинической значимостью как одного из наиболее распространенных возбудителей инфекций, а также высокой устойчивостью к неблагоприятным условиям окружающей среды и антибиотикам [8].

Антибактериальная активность композитов оценивалась путем измерения диаметра зоны ингибирования роста бактерий — участка вокруг образца композита, свободного от бактериальной колонии. Для повышения точности измерений диаметр зоны ингибирования определялся в двух взаимно перпендикулярных направлениях, после чего рассчитывалось среднее значение для каждого образца.

В эксперименте использовались по три образца порошковых нанокомпозитов систем $ZnO-ZnCr_2O_4$ и Cu/ZnO-ZnCr_2O_4. Для каждого образца проводились три независимых измерения (итого 9 на каждый из порошковых нанокомпозитов). Затем результаты усреднялись, чтобы получить итоговую оценку антибактериальной активности.

Результаты и обсуждение

На рис. 1 приведены дифрактограммы синтезированных композитов, на которых хорошо видны интенсивные пики гексагональных кристаллов ZnO и кубических кристаллов шпинели ZnCr₂O₄. Положение пиков кристаллов ZnO на дифрактограмме соответствуют стандартным значениям 20 для кристаллов ZnO (Joint Committee on Powder Diffraction Standards (JCPDS) карточка № 01-074-9939).

На приведенной дифрактограмме (рис. 1) нанокомпозита Cu/Zn-ZnCr₂O₄ отсутствуют дифракционные пики, соответствующие соединениям меди. Это может быть связано с тем, что ионы Cu²⁺ (0,057 нм) имеют меньший ионный радиус по сравнению с ионами Zn²⁺ (0,072 нм) и Cr³⁺ (0,064 нм) [33], что позволяет им замещать ионы цинка в кристаллической решетке ZnO. Такое замещение приводит к формированию нанокомпозитов на основе ZnO без образования отдельной фазы Cu.

В табл. 2 приведены данные сопоставления положения некоторых пиков ZnO на дифрактограммах, приведенных на рис. 1, со стандартными значениями (карточка JCPDS № 01-074-9939), параметры элементарной ячейки кристаллов (a, c) и объем элементарной ячейки (V). Из табл. 2, видно, что введение Cu в состав композита приводит к небольшому сдвигу положения пиков ZnO в сторону больших значений 2 θ , относительно их положения на дифрактограмме нанокомпозита ZnO-ZnCr₂O₄. Это связано со сжатием элементарной ячейки кристаллов ZnO при замещении ионов Zn²⁺ меньшими по размеру ионами Cu²⁺.

В табл. 3 приведены данные сопоставления положения пиков ZnCr₂O₄ на дифрактограммах (рис. 1), со стандартными значениями (карточка JCPDS № 01-075-4058) и параметры элементарной ячейки кристаллов.

Из данных табл. 2 видно, что на рентгенограмме нанокомпозита $Cu/Zn-ZnCr_2O_4$ наблюдается сдвиг пиков кристаллов $ZnCr_2O_4$ в сторону больших значе-



Puc. 1. Дифрактограммы нанокомпозитов Cu/Zn-ZnCr₂O₄ (*a*) и Zn-ZnCr₂O₄ (*b*) *Fig. 1.* XRD patterns of Cu/Zn-ZnCr₂O₄ (*a*) and Zn-ZnCr₂O₄ (*b*) nanocomposites

Таблица 2. Сопоставление положения пиков ZnO на дифрактограммах (рис. 1) со стандартными значениями
(карточка JCPDS № 01-074-9939) и параметрами элементарной ячейки кристаллов
Table 2 Comparison of the positions of 7nO modes in the differentian potterns above in Fig. 1 with standard values

 Table 2. Comparison of the positions of ZnO peaks in the diffraction patterns shown in Fig. 1 with standard values (JCPDS card No. 01-074-9939) and parameters of the unit cell of crystals

	Положение пиков на дифрактограмме 20, град				
	карточка № 01-074-9939	Cu/Zn-ZnCr ₂ O ₄	Zn-ZnCr ₂ O ₄		
(100)	31,77	31,77	31,70		
(002)	34,43	34,45	34,40		
(101)	36,26	36,26	36,20		
(102)	47,55	47,55	47,45		
(110)	56,60	56,60	56,55		
(103)	62,87	62,85	62,80		
Параметры элементарной ячейки	a = 0,3249 HM; c = 0,5205 HM; V = 0,0476 HM ³	a = 0,3245 нм; c = 0,5200 нм; V = 0,0474 нм ³	a = 0,3251 HM; c = 0,5208 HM; V = 0,0477 HM ³		

Таблица 3. Сопоставление положения пиков ZnCr₂O₄, приведенных на рис. 1 дифрактограммах, со стандартными значениями (карточка JCPDS № 01-075-4058) и параметрами элементарной ячейки кристаллов

Table 3. Comparison of the positions of ZnCr₂O₄ peaks in the diffraction patterns shown in Fig. 1 with standard values (JCPDS card No. 01-075-4058) and parameters of the unit cell of crystals

	Положение пиков на дифрактограмме 20, град				
	карточка № 01-075-4058	Cu/Zn-ZnCr ₂ O ₄	Zn-ZnCr ₂ O ₄		
(220)	30,13	30,35	30,29		
(400)	43,13	43,30	43,36		
(422)	53,50	53,92	53,81		
Параметры элементарной ячейки	a = 0,8383 HM; V = 0,5892 HM ³	<i>a</i> = 0,8372 нм; <i>V</i> = 0,5868 нм ³	<i>a</i> = 0,8340 нм; <i>V</i> = 0,5802 нм ³		

ний 20, что свидетельствует о сжатии элементарной ячейки кристаллов. Видно также, что несколько меньшие, по сравнению со стандартными (карточка JCPDS № 01-075-4058) значениями, величины объема элементарной ячейки кристаллов ZnCr₂O₄ наблюдаются в обоих синтезированных композитах.

Расчеты, проведенные по формуле Шеррера, показали, что размер кристаллов ZnO в композитах составляет 13–17 нм. Такие небольшие размеры кристаллов определяют развитую поверхность материала и возможность эффективного контакта частиц материала с окружающей средой.

На рис. 2 приведены электронно-микроскопические снимки синтезированных композитов. Видно, что оба композита состоят из агрегатов микронного размера неправильной формы и частиц, имеющих субмикронные размеры. Высокая дисперсность материала на основании данных о размере кристаллов обеспечивает эффективный контакт частиц композита с окружающей средой, что является важным для фотокаталитической и антибактериальной активностей материала, повышая их на несколько порядков [6, 11, 34].

В табл. 4 приведен аналитический элементный состав ионов металлов в композитах системы ZnO-

 $ZnCr_2O_4$ по данным энергодисперсионного анализа. Полученные значения свидетельствуют о полном соответствии номинального (расчетного) и аналитического химического состава полученных материалов.

На рис. 3 показаны спектры фотолюминесценции композита ZnO-ZnCr2O4 при его возбуждении излучением УФ диапазона. В спектрах наблюдаются многочисленные полосы излучения, преимущественно расположенные в синей части видимой области спектра. На основании результатов работ [35–38], наблюдаемые в спектрах полосы люминесценции в основном связаны с люминесценцией структурных дефектов кристаллов ZnO (межрешеточные ионы цинка, кислородные вакансии). Например, заметная в спектрах полоса с максимумом при 490 нм определяется эмиссией кислородных вакансий V₀ [35, 36]. Относительно слабая по интенсивности полоса зеленой люминесценции (λ_{max} около 530 нм), наблюдавшаяся в [14, 37] в спектрах люминесценции нанокристаллов ZnO, связана с эмиссией вакансий цинка в структуре материалов [36, 37]. На основании экспериментальных данных, приведенных в [14], можно предположить, что эти дефектные центры расположены на поверхности нанокристаллов ZnO.



Puc. 2. Электронно-микроскопические снимки при различных увеличениях композитов ZnO-ZnCr₂O₄ (a, b) и Cu/ZnO-ZnCr₂O₄ (c, d)*Fig.* 2. SEM images at different magnifications of ZnO-ZnCr₂O₄ (a, b) and Cu/ZnO-ZnCr₂O₄ (c, d) composites

Научно-технический вестник информационных технологий, механики и оптики, 2025, том 25, № 2 Scientific and Technical Journal of Information Technologies, Mechanics and Optics, 2025, vol. 25, no 2

	Композиты					
элементный состав	ZnO-ZnCr ₂ O ₄		Cu/ZnO-ZnCr ₂ O ₄			
Элемент	Zn	Cr	Zn	Cr	Cu	
Содержание, %	86,62	13,15	86,34	12,47	1,19	
Среднеквадратичное отклонение, %	0,23	0,23	0,21	0,18	0,03	

Таблица 4. Аналитический элементный состав ионов металлов в композитах *Table 4*. Analytical elemental composition of metal ions in composites



Рис. 3. Спектры люминесценции композита системы ZnO-ZnCr₂O₄. Длина волны возбуждения люминесценции: 230 нм (кривая 1); 280 нм (кривая 2); 380 нм (кривая 3) *Fig. 3.* Luminescence spectra of the ZnO-ZnCr₂O₄ system composite. Luminescence excitation wavelength: 230 nm (curve 1); 280 nm (curve 2); 380 nm (curve 3)

Полосы люминесценции в синей части спектра с максимумами на 420 и 440 нм могут быть связаны по данным [38] с присутствием в структуре ZnO различных структурных дефектов межрешеточных ионов цинка Zn_i, вакансий цинка и кислорода. Кроме того, нанокристаллы ZnCr₂O₄ демонстрируют интенсивную синюю ($\lambda_{max} = 410$ нм) люминесценцию под действием коротковолнового УФ излучения ($\lambda = 260$ нм) [39].

Таким образом, данные спектров люминесценции свидетельствуют о присутствии в структуре синтезированного композита различных структурных дефектов. Следует отметить, что по данным работы [13] присутствие таких дефектов оказывает заметное влияние на фотокаталитическую активность материалов.

На рис. 4, *а* приведен спектр люминесценции композита ZnO-ZnCr₂O₄ под действием излучения с длиной волны $\lambda = 405$ нм. В спектре наблюдается полоса люминесценции, характерная для синглетного кислорода (электронный переход ${}^{1}\Delta_{g} {}^{-3}\Sigma_{g}$ [40]). Отметим, что генерация химически активного синглетного кислорода наблюдается под действием синего света ($\lambda = 405$ нм), энергия квантов которого (3,06 эВ) меньше значений ширины запрещенной зоны кристаллов ZnO и ZnCr₂O₄, входящих в структуру композита (3,37 эВ [12, 14] и 2,90–3,35 эВ [18, 41, 42] соответственно). Относительная интенсивность пика на рис. 4, *а* рассчитывалась на основе значения максимума интенсивности люминесценции на длине 1270 нм при возбуждении излучением с плотностью мощности 1700 мВт/см².

Увеличение плотности мощности возбуждающего излучения существенно повышает интенсивность люминесценции синглетного кислорода (рис. 4, *b*), что согласуется с результатами, полученными в работе [43]. Зависимость интенсивности люминесценции от плотности мощности излучения в диапазоне 1000–1700 мВт/см² носит линейный характер, что указывает на отсутствие насыщения процессов генерации



Puc. 4. Спектр люминесценции композита ZnO-ZnCr₂O₄ в ближней инфракрасной области спектра. Длина волны возбуждения 405 нм (*a*). Зависимость интенсивности люминесценции (λ = 1270 нм) синглетного кислорода от плотности мощности возбуждающего излучения (*b*)

Fig. 4. Luminescence spectrum of the ZnO-ZnCr₂O₄ composite in the near infrared region of the spectrum. Excitation wavelength is 405 nm (*a*). Dependence of the luminescence intensity ($\lambda = 1270$ nm) of singlet oxygen on the power density of the exciting radiation (*b*)

синглетного кислорода в данном интервале параметров облучения.

Интенсивность люминесценции синглетного кислорода на длине волны 1270 нм напрямую связана с его концентрацией [44]. Таким образом, увеличение интенсивности люминесценции свидетельствует о росте концентрации синглетного кислорода, а значит, о повышении скорости его генерации фотокатализатором. Полученные результаты демонстрируют, что синтезированный композит ZnO-ZnCr₂O₄ генерирует синглетный кислород под воздействием видимого излучения с длиной волны $\lambda = 405$ нм, а увеличение плотности мощности возбуждающего излучения видимого спектрального диапазона позволяет повышать скорость его генерации.

На рис. 5, *а* приведены кинетические зависимости адсорбции красителя CSB на поверхности композита ZnO-ZnCr₂O₄ (кривая *1*) и фотокаталитического раз-

ложения красителя под действием УФ излучения (кривая 2). Расчет относительной концентрации рассчитывался как отношение оптических плотностей растворов на длине волны 611 нм. Видно, что при адсорбции за 30 мин из раствора на композите осаждается около 20 % молекул красителя.

Для описания кинетики адсорбции органических красителей на поверхности оксидных материалов часто применяют кинетическую модель первого порядка, в которой скорость процесса адсорбции описывается уравнением [10, 24]:

$$\frac{dq_t}{dt} = k_f(q_e - q_t),\tag{1}$$

где q_t — масса красителя, адсорбированного 1 г сорбента к моменту времени t; q_e — равновесная адсорбционная емкость сорбента; k_f — константа ско-



Рис. 5. Кинетические зависимости адсорбции красителя CSB на поверхности композита ZnO-ZnCr₂O₄ (кривая 1) и фотокаталитического разложения красителя в присутствии композита под действием ультрафиолетового излучения (кривая 2) (*a*). Зависимости, рассчитанные на основании экспериментальных данных по адсорбции красителя на поверхности композита $-\ln(q_e - q_t) = f(t)$ (*b*) и по фотокаталитическому разложению красителя в растворе в присутствии композита $-\ln(C/C_0) = f(t)$ (*c*)

Fig. 5. Kinetic dependences of CSB dye adsorption on the surface of ZnO-ZnCr₂O₄ composite (curve *1*) and photocatalytic decomposition of the dye in the presence of the composite under the action of UV radiation (curve *2*) (*a*). Dependence $-\ln(q_e - q_l) = f(t)$ calculated on the basis of experimental data on dye adsorption on the surface of the composite (*b*). Dependence $-\ln(C/C_0) = f(t)$ calculated on the basis of experimental data on photocatalytic decomposition of the dye in solution in the presence of the composite (*c*) рости адсорбции; *t* — продолжительность процесса адсорбции.

Зависимость $-\ln(q_e - q_t) = f(t)$, рассчитанная на основании экспериментальных данных, приведена на рис. 5, *b*. Полученные в настоящей работе экспериментальные данные по кинетике адсорбции удовлетворительно описываются линейной зависимостью (1) с коэффициентом детерминации R² более 0,9. Можно отметить, что данный результат соответствует полученным ранее результатам [6, 10, 24]по исследованию кинетики адсорбции красителя CSB на поверхности наноматериалов на основе ZnO.

Процесс фотокаталитического разложения красителя протекает существенно быстрее процесса адсорбции (рис. 5, *a*). Это может быть связано с тем, что в процессе катализа разложению подвергаются не только молекулы CSB, адсорбированные на поверхности композита, но и находящиеся в растворе [10, 24].

Модель Лангмюра–Хиншельвуда [10, 24, 45] применяется для описания кинетики фотокаталитического разложения органических красителей в растворах. В рамках этой модели кинетика их разложения в разбавленных растворах (концентрация *C* <<10⁻³ M) соответствует уравнению псевдо-первого порядка:

$$-\ln(C/C_0) = kt, \tag{2}$$

где *C* и *C*₀ — текущая и начальная концентрации красителя в растворе; *t* — время; *k* — константа скорости фотокаталитического разложения красителя псевдопервого порядка. На рис. 5, *c* приведена зависимость $-\ln(C/C_0) = f(t)$, рассчитанная на основании экспериментальных данных (кривая *l*, рис. 5, *a*). Видно, что приведенная зависимость носит линейный характер и хорошо (\mathbb{R}^2 более 0,9) описывается уравнением (2).

На рис. 6 приведены кинетические зависимости фотолиза CSB в растворе, происходящего под воздействием излучения без катализатора, а также его фотокаталитического разложения в присутствии композитов ZnO-ZnCr₂O₄ (кривая 2) и Cu/ZnO-ZnCr₂O₄ (кривая 3) при облучении светодиодом HP-3003 (Litlight, Китай) с длиной волны в интервале 390–400 нм. Выбор спектрального состава излучения диода обусловлен тем, что он соответствует границе УФ-А и видимого диапазона, что позволяет оценить скорость фотокаталитической реакции под воздействием мягкого УФ и видимого света.

Из представленных данных (рис. 6) видно, что скорость фотолиза красителя более чем в два раза ниже, чем скорость его фотокаталитического разложения в присутствии нанокомпозита Cu/ZnO-ZnCr₂O₄. Это подтверждает, что каталитическая активность композитов способствует ускоренной деградации CSB под воздействием излучения мягкого УФ-А излучения и видимого света.

Исследования антибактериальной активности нанокомпозита показали, что он обладает высокими бактерицидными свойствами в отношении бактерий *Staphylococcus aureus ATCC 209P*. На рис. 7 приведена фотография чашки Петри, содержащей агар и бактерии,



Рис. 6. Кинетические зависимости фотолиза красителя в растворе под действием излучения (кривая 1) и фотокаталитического разложения красителя в растворе в присутствии композитов ZnO-ZnCr₂O₄ (кривая 2)

и Cu/ZnO-ZnCr₂O₄ (кривая 3). Длина волны излучения 395 нм

Fig. 6. Kinetic dependences of dye photolysis in solution under the action of radiation (curve *1*) and photocatalytic decomposition of the dye in solution in the presence of ZnO-ZnCr₂O₄ (curve *2*) and Cu/ZnO-ZnCr₂O₄ (curve *3*) composites. Radiation wavelength 395 nm



Рис. 7. Антибактериальная активность нанокомпозита ZnO-ZnCr₂O₄.

I — порошкообразные образцы нанокомпозита; 2 — агар
 с бактериями Staphylococcus aureus ATCC 209P; 3 — зона, свободная от бактерий

Fig. 7. Antibacterial activity of ZnO-ZnCr₂O₄ nanocomposite. *I* — powdered nanocomposite samples; *2* — agar with bacteria *Staphylococcus aureus ATCC 209P*; *3* — bacteria-free zone

в которой размещены три порошкообразных образца нанокомпозита. Величина зоны свободной от бактерий, окружающей каждый образец, превосходит 12 мм. Приведенный результат (рис. 7) свидетельствует об антибактериальной активности композита.

Заключение

Синтезированные полимерно-солевым методом композиты ZnO-ZnCr₂O₄ и Cu/ZnO-ZnCr₂O₄ состоят из дисперсных частиц микронного размера. Эти частицы сформированы из гексагональных нанокристаллов ZnO со средним размером около 16 нм и нанокристаллов шпинели ZnCr₂O₄. Кристаллическая решетка ZnCr₂O₄ в композите Cu/ZnO-ZnCr₂O₄ деформирована из-за внедрения в нее ионов меди. Данные исследования фотолюминесценции в видимой части спектра свидетельствуют о присутствии в структуре нанокристаллов ZnO и ZnCr₂O₄ различных дефектов (кислородные вакансии, межрешеточные ионы цинка и др.).

Полученный нанокомпозит демонстрирует способность к генерации синглетного кислорода под действием излучения видимого диапазона (длина волны равна 405 нм), при этом энергия фотона меньше значений ширины запрещенной зоны кристаллов ZnO и ZnCr₂O₄, входящих в структуру композита.

Скорость фотокаталитического разложения диазокрасителя Chicago Sky Blue (CSB) в присутствии композита ZnO-ZnCr₂O₄ в водном растворе под воздействием ультрафиолетового излучения выше скорости адсорбции CSB на этом композите. Что свидетельствует о наличии фотокаталитических свойств нанокомпозита ZnO-ZnCr₂O₄.

Литература

- 1. Хомутинникова Л.Л., Мешковский И.К., Евстропьев С.К., Литвинов М.Ю., Быков Е.П., Плясцов С.А. Методика оптического детектирования метана волоконно-оптическом сенсором при применении фотокаталитического нанокомпозита ZnO-SnO₂-Fe₂O₃ // Оптика и спектроскопия. 2023. Т. 131. № 3. С. 427–432. https://doi.org/10.21883/os.2023.03.55395.4525-23
- Gaya U.I., Abdullah A.H. Heterogeneous photocatalytic degradation of organic contaminants over titanium dioxide: a review of fundamentals, progress and problems // Journal of Photochemistry and Photobiology C: Photochemistry Reviews. 2008. V. 9. N 1. P. 1–12. https://doi.org/10.1016/j.jphotochemrev.2007.12.003
- Uribe-López M.C., Hidalgo-López M.C., López-González R., Frías-Márquez D.M., Núñez-Nogueira G., Hernández-Castillo D., Alvarez-Lemus M.A. Photocatalytic activity of ZnO nanoparticles and the role of the synthesis method on their physical and chemical properties // Journal of Photochemistry and Photobiology A: Chemistry. 2021. V. 404. P. 112866. https://doi.org/10.1016/j. jphotochem.2020.112866
- Ong C.B., Ng L.Y., Mohammad A.W. A review of ZnO nanoparticles as solar photocatalysts: Synthesis, mechanisms and applications // Renewable and Sustainable Energy Reviews. 2018. V. 81. Part 1. P. 536–551. https://doi.org/10.1016/j.rser.2017.08.020
- Gusain R., Gupta K., Joshi P., Khatri O.P. Adsorptive removal and photocatalytic degradation of organic pollutants using metal oxides and their composites: A comprehensive review // Advances in Colloid and Interface Science. 2019. V. 272. P. 102009. https://doi. org/10.1016/j.cis.2019.102009
- Gavrilova M., Gavrilova D., Evstropiev S., Shelemanov A., Bagrov I. Porous ceramic ZnO nanopowders: features of photoluminescence, adsorption and photocatalytic properties // Ceramics. 2023. V. 6. N 3. P. 1667–1681. https://doi.org/10.3390/ceramics6030103
- Li R., Zhang L., Wang P. Rational design of nanomaterials for water treatment // Nanoscale. 2015. V. 7. N 41. P. 17167–17194. https://doi. org/10.1039/C5NR04870B
- Raghupathi K.R., Koodali R.T., Manna A.C. Size-dependent bacterial growth inhibition and mechanism of antibacterial activity of zinc oxide nanoparticles // Langmuir. 2011. V. 27. N 7. P. 4020–4028. https://doi.org/10.1021/la104825u
- Qi K., Cheng B., Yu J., Ho W. Review on the improvement of the photocatalytic and antibacterial activities of ZnO // Journal of Alloys and Compounds. 2017. V. 727. P. 792–820. https://doi.org/10.1016/j. jallcom.2017.08.142
- Saratovskii A.S., Bulyga D.V., Evstrop'ev S.K., Antropova T.V. Adsorption and photocatalytic activity of the porous glass–ZnO–Ag

Нанокомпозиты ZnO-ZnCr₂O₄ и Cu/ZnO-ZnCr₂O₄ демонстрируют фотокаталитическую активность при разложении органического диазокрасителя CSB в водных растворах. Скорость фотокаталитического разложения этого красителя под воздействием излучения с диапазоном длин волн 390–400 нм описывается кинетическим уравнением псевдо-первого порядка (k = 0,009 мин⁻¹ и k = 0,012 мин⁻¹) и в два раза превышает скорость фотолиза CSB в отсутствии навески композита в аналогичных условиях (k = 0,005 мин⁻¹).

Изучение бактерицидных свойств нанокомпозита ZnO-ZnCr₂O₄ методом дисковой диффузии показало, что он обладает антибактериальной активностью в отношении бактерий *Staphylococcus aureus ATCC 209P*. Размер ингибированной зоны составил более 12 мм.

На основании полученных результатов можно сделать вывод о перспективности применения разработанных нанокомпозитов в фотокаталитических системах очистки воды и воздуха.

References

- Khomutinnikova L.L., Meshkovskii I.K., Evstropiev S.K., Litvinov M.Y., Bykov E.P., Plyastsov S.A. Method of methane detection by a fiber-optic sensor using a photocatalytic nanocomposite ZnO-SnO₂-Fe₂O₃. *Optics and Spectroscopy*, 2023, vol. 131, no. 3, pp. 399–404. (in Russian). https://doi.org/10.61011/ EOS.2023.03.56193.4525-23
- Gaya U.I., Abdullah A.H. Heterogeneous photocatalytic degradation of organic contaminants over titanium dioxide: a review of fundamentals, progress and problems. *Journal of Photochemistry and Photobiology C: Photochemistry Reviews*, 2008, vol. 9, no. 1, pp. 1–12. https://doi.org/10.1016/j.jphotochemrev.2007.12.003
- Uribe-López M.C., Hidalgo-López M.C., López-González R., Frías-Márquez D.M., Núñez-Nogueira G., Hernández-Castillo D., Alvarez-Lemus M.A. Photocatalytic activity of ZnO nanoparticles and the role of the synthesis method on their physical and chemical properties. *Journal of Photochemistry and Photobiology A: Chemistry*, 2021, vol. 404, pp. 112866. https://doi.org/10.1016/j. jphotochem.2020.112866
- Ong C.B., Ng L.Y., Mohammad A.W. A review of ZnO nanoparticles as solar photocatalysts: Synthesis, mechanisms and applications. *Renewable and Sustainable Energy Reviews*, 2018, vol. 81, part 1, pp. 536–551. https://doi.org/10.1016/j.rser.2017.08.020
- Gusain R., Gupta K., Joshi P., Khatri O.P. Adsorptive removal and photocatalytic degradation of organic pollutants using metal oxides and their composites: A comprehensive review. *Advances in Colloid and Interface Science*, 2019, vol. 272, pp. 102009. https://doi. org/10.1016/j.cis.2019.102009
- Gavrilova M., Gavrilova D., Evstropiev S., Shelemanov A., Bagrov I. Porous ceramic ZnO nanopowders: features of photoluminescence, adsorption and photocatalytic properties. *Ceramics*, 2023, vol. 6, no. 3, pp. 1667–1681. https://doi.org/10.3390/ceramics6030103
- Li R., Zhang L., Wang P. Rational design of nanomaterials for water treatment. *Nanoscale*, 2015, vol. 7, no. 41, pp. 17167–17194. https:// doi.org/10.1039/C5NR04870B
- Raghupathi K.R., Koodali R.T., Manna A.C. Size-dependent bacterial growth inhibition and mechanism of antibacterial activity of zinc oxide nanoparticles. *Langmuir*, 2011, vol. 27, no. 7, pp. 4020–4028. https://doi.org/10.1021/la104825u
- Qi K., Cheng B., Yu J., Ho W. Review on the improvement of the photocatalytic and antibacterial activities of ZnO. *Journal of Alloys* and Compounds, 2017, vol. 727, pp. 792–820. https://doi. org/10.1016/j.jallcom.2017.08.142
- Saratovskii A.S., Bulyga D.V., Evstrop'ev S.K., Antropova T.V. Adsorption and photocatalytic activity of the porous glass–ZnO–Ag

composite and ZnO–Ag nanopowder // Glass Physics and Chemistry. 2022. V. 48. N 1. P. 10–17. https://doi.org/10.1134/S1087659622010126

- Wang T., Tian B., Han B., Ma D., Sun M., Hanif A., Xia D., Shang J. Recent advances on porous materials for synergetic adsorption and photocatalysis // Energy & Environmental Materials. 2022. V. 5. N 3. P. 711–730. https://doi.org/10.1002/eem2.12229
- Rao L.S., Rao T.V., Naheed Sd., Rao P.V. Structural and optical properties of zinc magnesium oxide nanoparticles synthesized by chemical co-precipitation // Materials Chemistry and Physics. 2018.
 V. 203. P. 133-140. https://doi.org/10.1016/j. matchemphys.2017.09.048
- Bhatia S., Verma N. Photocatalytic activity of ZnO nanoparticles with optimization of defects // Materials Research Bulletin. 2017. V. 95. P. 468–476. https://doi.org/10.1016/j.materresbull.2017.08.019
- Guo L., Yang S., Yang C., Yu P., Wang J., Ge,p W., Wong G.K.L. Highly monodisperse polymer-capped ZnO nanoparticles: Preparation and optical properties // Applied Physics Letters. 2000. V. 76. N 20. P. 2901–2903. https://doi.org/10.1063/1.126511
- Chen X., Wu Z., Liu D., Gao Z. Preparation of ZnO photocatalyst for the efficient and rapid photocatalytic degradation of azo dyes // Nanoscale Research Letters. 2017. V. 12. N 1. P. 143. https://doi. org/10.1186/s11671-017-1904-4
- Evstropiev S.K., Lesnykh L.V., Karavaeva A.V., Nikonorov N.V., Oreshkina K.V., Mironov L.Y., Maslennikov S.Y., Kolobkova E.V., Vasilyev V.N., Bagrov I.V. Intensification of photodecomposition of organics contaminations by nanostructured ZnO-SnO₂ coatings prepared by polymer-salt method // Chemical Engineering and Processing – Process Intensification. 2019. V. 142. P. 107587. https:// doi.org/10.1016/j.cep.2019.107587
- Mimouni R., Askri B., Larbi T., Amlouk M., Meftah A. Photocatalytic degradation and photo-generated hydrophilicity of Methylene Blue over ZnO/ZnCr₂O₄ nanocomposite under stimulated UV light irradiation // Inorganic Chemistry Communications. 2020. V. 115. P. 107889. https://doi.org/10.1016/j.inoche.2020.107889
- Mousavi Z., Soofivand F., Esmaeili-Zare M., Salavati-Niasari M., Bagheri S. ZnCr₂O₄ nanoparticles: facile synthesis, characterization and photocatalytic properties // Scientific Reports. 2016. V. 6. P. 20071. https://doi.org/10.1038/srep20071
- Benrighi Y., Nasrallah N., Chaabane T., Sivasankar V., Darchen A., Baaloudj O. Photocatalytic performances of ZnCr₂O₄ nanoparticles for cephalosporins removal: Structural, optical and electrochemical properties // Optical Materials. 2021. V. 115. P. 111035. https://doi. org/10.1016/j.optmat.2021.111035
- 20. Peng C., Gao L. Optical and photocatalytic properties of spinel ZnCr₂O₄ nanoparticles synthesized by a hydrothermal route // Journal of the American Ceramic Society. 2008. V. 91. N 7. P. 2388–2390. https://doi.org/10.1111/j.1551-2916.2008.02417.x
- 21. Das S., Misra A.J., Rahman A.P.H., Das B., Jayabalan R., Tamhankar A.J., Mishra A., Lundborg C.S., Tripathy S.K. Ag@ SnO₂@ZnO core-shell nanocomposites assisted solar-photocatalysis downregulates multidrug resistance in Bacillus sp.: a catalytic approach to impede antibiotic resistance // Applied Catalysis B: Environmental. 2019. V. 259. P. 118065. https://doi.org/10.1016/j. apcatb.2019.118065
- Lu Y.H., Xu M., Xu L.X. Zhang C.L., Zhang Q.P., Xu X.N., Xu S., Ostrikov K. Enhanced ultraviolet photocatalytic activity of Ag/ZnO nanoparticles synthesized by modified polymer-network gel method // Journal of Nanoparticle Research. 2015. V. 17. N 9. P. 350. https:// doi.org/10.1007/s11051-015-3150-y
- Lavín A., Sivasamy R., Mosquera E., Morel M.J. High proportion ZnO/CuO nanocomposites: Synthesis, structural and optical properties, and their photocatalytic behavior // Surfaces and Interfaces. 2019. V. 17. P. 100367. https://doi.org/10.1016/j. surfin.2019.100367
- Shelemanov A., Tincu A., Evstropiev S., Nikonorov N., Vasilyev V. Cu-doped porous ZnO-ZnAl₂O₄ nanocomposites synthesized by polymer-salt method for photocatalytic water purification // Journal of Composites Science. 2023. V. 7. N 7. P. 263. https://doi. org/10.3390/jcs7070263
- 25. Wang C., Wang X., Xu B.Q., Zhao J.C., Mai B.X., Peng P., Sheng G.Y., Fu H.M. Enhanced photocatalytic performance of nanosized coupled ZnO/SnO₂ photocatalysts for methyl orange degradation // Journal of Photochemistry and Photobiology A: Chemistry. 2004. V. 168. N 1-2. P. 47–52. https://doi.org/10.1016/j. jphotochem.2004.05.014

composite and ZnO–Ag nanopowder. *Glass Physics and Chemistry*, 2022, vol. 48, no. 1, pp. 10–17. https://doi.org/10.1134/S1087659622010126

- Wang T., Tian B., Han B., Ma D., Sun M., Hanif A., Xia D., Shang J. Recent advances on porous materials for synergetic adsorption and photocatalysis. *Energy & Environmental Materials*, 2022, vol. 5, no. 3, pp. 711–730. https://doi.org/10.1002/eem2.12229
- Rao L.S., Rao T.V., Naheed Sd., Rao P.V. Structural and optical properties of zinc magnesium oxide nanoparticles synthesized by chemical co-precipitation. *Materials Chemistry and Physics*, 2018, vol. 203, pp. 133–140. https://doi.org/10.1016/j. matchemphys.2017.09.048
- Bhatia S., Verma N. Photocatalytic activity of ZnO nanoparticles with optimization of defects. *Materials Research Bulletin*, 2017, vol. 95. pp. 468–476. https://doi.org/10.1016/j.materresbull.2017.08.019
- Guo L., Yang S., Yang C., Yu P., Wang J., Ge,p W., Wong G.K.L. Highly monodisperse polymer-capped ZnO nanoparticles: Preparation and optical properties. *Applied Physics Letters*, 2000, vol. 76, no. 20, pp. 2901–2903. https://doi.org/10.1063/1.126511
- Chen X., Wu Z., Liu D., Gao Z. Preparation of ZnO photocatalyst for the efficient and rapid photocatalytic degradation of azo dyes. *Nanoscale Research Letters*, 2017, vol. 12, no. 1, pp. 43. https://doi. org/10.1186/s11671-017-1904-4
- Evstropiev S.K., Lesnykh L.V., Karavaeva A.V., Nikonorov N.V., Oreshkina K.V., Mironov L.Y., Maslennikov S.Y., Kolobkova E.V., Vasilyev V.N., Bagrov I.V. Intensification of photodecomposition of organics contaminations by nanostructured ZnO-SnO₂ coatings prepared by polymer-salt method. *Chemical Engineering and Processing — Process Intensification*, 2019, vol. 142, pp. 107587. https://doi.org/10.1016/j.cep.2019.107587
- Mimouni R., Askri B., Larbi T., Amlouk M., Meftah A. Photocatalytic degradation and photo-generated hydrophilicity of Methylene Blue over ZnO/ZnCr₂O₄ nanocomposite under stimulated UV light irradiation. *Inorganic Chemistry Communications*, 2020, vol. 115, pp. 107889. https://doi.org/10.1016/j.inoche.2020.107889
- Mousavi Z., Soofivand F., Esmaeili-Zare M., Salavati-Niasari M., Bagheri S. ZnCr₂O₄ nanoparticles: facile synthesis, characterization and photocatalytic properties. *Scientific Reports*, 2016, vol. 6, pp. 20071. https://doi.org/10.1038/srep20071
- Benrighi Y., Nasrallah N., Chaabane T., Sivasankar V., Darchen A., Baaloudj O. Photocatalytic performances of ZnCr₂O₄ nanoparticles for cephalosporins removal: Structural, optical and electrochemical properties. *Optical Materials*, 2021, vol. 115, pp. 111035. https://doi. org/10.1016/j.optmat.2021.111035
- 20. Peng C., Gao L. Optical and photocatalytic properties of spinel ZnCr₂O₄ nanoparticles synthesized by a hydrothermal route. *Journal* of the American Ceramic Society, 2008, vol. 91, no. 7, pp. 2388–2390. https://doi.org/10.1111/j.1551-2916.2008.02417.x
- 21. Das S., Misra A.J., Rahman A.P.H., Das B., Jayabalan R., Tamhankar A.J., Mishra A., Lundborg C.S., Tripathy S.K. Ag@ SnO2@ZnO core-shell nanocomposites assisted solar-photocatalysis downregulates multidrug resistance in Bacillus sp.: a catalytic approach to impede antibiotic resistance. *Applied Catalysis B: Environmental*, 2019, vol. 259, pp. 118065. https://doi.org/10.1016/j. apcatb.2019.118065
- 22. Lu Y.H., Xu M., Xu L.X. Zhang C.L., Zhang Q.P., Xu X.N., Xu S., Ostrikov K. Enhanced ultraviolet photocatalytic activity of Ag/ZnO nanoparticles synthesized by modified polymer-network gel method. *Journal of Nanoparticle Research*, 2015, vol. 17, no. 9, pp. 350. https://doi.org/10.1007/s11051-015-3150-y
- Lavín A., Sivasamy R., Mosquera E., Morel M.J. High proportion ZnO/CuO nanocomposites: Synthesis, structural and optical properties, and their photocatalytic behavior. *Surfaces and Interfaces*, 2019, vol. 17, pp. 100367. https://doi.org/10.1016/j. surfin.2019.100367
- Shelemanov A., Tincu A., Evstropiev S., Nikonorov N., Vasilyev V. Cu-doped porous ZnO-ZnAl₂O₄ nanocomposites synthesized by polymer-salt method for photocatalytic water purification. *Journal of Composites Science*, 2023, vol. 7, no. 7, pp. 263. https://doi. org/10.3390/jcs7070263
- 25. Wang C., Wang X., Xu B.Q., Zhao J.C., Mai B.X., Peng P., Sheng G.Y., Fu H.M. Enhanced photocatalytic performance of nanosized coupled ZnO/SnO₂ photocatalysts for methyl orange degradation. *Journal of Photochemistry and Photobiology A: Chemistry*, 2004, vol. 168, no. 1-2, pp. 47–52. https://doi. org/10.1016/j.jphotochem.2004.05.014

- Li B., Wang Y.F. Facile synthesis and photocatalytic activity of ZnO– CuO nanocomposite // Superlattices and Microstructures. 2010. V. 47. N 5. P. 615–623. https://doi.org/10.1016/j.spmi.2010.02.005
- Liu Y., Huang J., Feng X., Li H. Thermal-sprayed photocatalytic coatings for biocidal applications: a review // Journal of Thermal Spray Technology. 2021. V. 30. N 1-2. P. 1–24. https://doi. org/10.1007/s11666-020-01118-2
- Riaz N., Hassan M., Siddique M. Mahmood Q., Farooq U., Sarwar R., Khan M.S. Photocatalytic degradation and kinetic modeling of azo dye using bimetallic photocatalysts: effect of synthesis and operational parameters // Environmental Science and Pollution Research. 2020. V. 27. N 3. P. 2992–3006. https://doi. org/10.1007/s11356-019-06727-1
- Evstropiev S.K., Kislyakov I.M., Bagrov I.V., Belousova I.M. Stabilization of PbS quantum dots by high molecular polyvinylpyrrolidone // Polymers for Advanced Technologies. 2016. V. 27. N 3. P. 314–317. https://doi.org/10.1002/pat.3642
- 30. Дукельский К.В., Евстропьев С.К. Формирование защитных наноразмерных покрытий на основе Al₂O₃ (Al₂O₃-AlF₃) на поверхности стекол // Оптический журнал. 2011. Т. 78. № 2. С. 71–81.
- Gene S.A., Saion E., Shaari A.H., Kamarudin M.A., Al-Hada N.M., Kharazmi A. Structural, optical, and magnetic characterization of spinel zinc chromite nanocrystallines synthesised by thermal treatment method // Journal of Nanomaterials. 2014. V. 2014. P. 416765. https://doi.org/10.1155/2014/416765
- Bokuniaeva A.O., Vorokh A.S. Estimation of particle size using the Debye equation and the Scherrer formula for polyphasic TiO₂ powder // Journal of Physics: Conference Series. 2019. V. 1410. N 1. P. 012057. https://doi.org/10.1088/1742-6596/1410/1/012057
- Shannon R.D. Revised effective ionic radii and systematic studies of interatomic distances in halides and chalcogenides // Acta Crystallographica Section A. 1976. V. 32. N 5. P. 751–767. https:// doi.org/10.1107/S0567739476001551
- Wang X., Ahmad M., Sun H. Three-dimensional ZnO hierarchical nanostructures: solution phase synthesis and applications // Materials. 2017. V. 10. N 11. P. 1304. https://doi.org/10.3390/ma10111304
- Børseth T.M., Svensson B.G., Kuznetsov A.Yu., Klason P., Zhao Q.X., Willander M. Identification of oxygen and zinc vacancy optical signals in ZnO // Applied Physics Letters. 2006. V. 89. N 26. P. 262112. https://doi.org/10.1063/1.2424641
- 36. Родный П.А., Черненко К.А., Веневцев И.Д. Механизмы люминесценции ZnO в видимой области спектра // Журнал технической физики. 2018. Т. 125. № 3. С. 357–363. https://doi. org/10.21883/OS.2018.09.46551.141-18
- Cao B.Q., Cai W.P., Zeng H.B. Temperature-dependent shifts of three emission bands for ZnO nanoneedle arrays // Applied Physics Letters. 2006. V. 88. N 16. P. 161101. https://doi.org/10.1063/1.2195694
- Vempati S., Mitra J., Dawson P. One-step synthesis of ZnO nanosheets: a blue-white fluorophore // Nanoscale Research Letters. 2012. V. 7. P. 470. https://doi.org/10.1186/1556-276X-7-470
- Ghosh D., Dutta U., Haque A., Mordvinova N.E., Lebedev O.I., Pal K., Gayen A., Seikh M.M., Mahata P. Ultra-high sensitivity of luminescent ZnCr₂O₄ nanoparticles toward nitroaromatic explosives sensing // Dalton Transactions. 2018. V. 47. N 14. P. 5011–5018. https://doi.org/10.1039/C8DT00047F
- Nosaka Y., Daimon T., Nosaka A.Y., Murakami Y. Singlet oxygen formation in photocatalytic TiO₂ aqueous suspension // Physical Chemistry Chemical Physics. 2004. V. 6. N 11. P. 2917–2918. https:// doi.org/10.1039/b405084c
- Abbasi A., Hamadanian M., Salavati-Niasari M., Mortazavi-Derazkola S. Facile size-controlled preparation of highly photocatalytically active ZnCr₂O₄ and ZnCr₂O₄/Ag nanostructures for removal of organic contaminants // Journal of Colloid and Interface Science. 2017. V. 500. P. 276–284. https://doi.org/10.1016/j. jcis.2017.04.003
- 42. Dumitru R., Manea F., Păcurariu C., Lupa L., Pop A., Cioabla A., Surdu A., Ianculescu A. Synthesis, characterization of nanosized ZnCr₂O₄ and its photocatalytic performance in the degradation of humic acid from drinking water // Catalysts. 2018. V. 8. N 5. P. 210. https://doi.org/10.3390/catal8050210
- Khomutinnikova L., Evstropiev S., Meshkovskii I., Bagrov I., Kiselev V. Ceramic ZnO-SnO₂-Fe₂O₃ powders and coatings -effective photogenerators of reactive oxygen species // Ceramics. 2023. V. 6. N 2. P. 886–897. https://doi.org/10.3390/ceramics6020051

- Li B., Wang Y.F. Facile synthesis and photocatalytic activity of ZnO– CuO nanocomposite. *Superlattices and Microstructures*, 2010, vol. 47, no. 5, pp. 615–623. https://doi.org/10.1016/j. spmi.2010.02.005
- Liu Y., Huang J., Feng X., Li H. Thermal-sprayed photocatalytic coatings for biocidal applications: a review. *Journal of Thermal Spray Technology*, 2021, vol. 30, no. 1-2, pp. 1–24. https://doi.org/10.1007/ s11666-020-01118-2
- Riaz N., Hassan M., Siddique M. Mahmood Q., Farooq U., Sarwar R., Khan M.S. Photocatalytic degradation and kinetic modeling of azo dye using bimetallic photocatalysts: effect of synthesis and operational parameters. *Environmental Science and Pollution Research*, 2020, vol. 27, no. 3, pp. 2992–3006. https://doi. org/10.1007/s11356-019-06727-1
- Evstropiev S.K., Kislyakov I.M., Bagrov I.V., Belousova I.M. Stabilization of PbS quantum dots by high molecular polyvinylpyrrolidone. *Polymers for Advanced Technologies*, 2016, vol. 27, no. 3, pp. 314–317. https://doi.org/10.1002/pat.3642
- Dukel'skiĭ K.V., Evstrop'ev S.K. Forming protective nanoscale coatings based on Al₂O₃ (Al₂O₃-AlF₃) on a glass surface. *Journal of Optical Technology*, 2011, vol. 78, no. 2, pp. 137–144. https://doi. org/10.1364/JOT.78.000137
- Gene S.A., Saion E., Shaari A.H., Kamarudin M.A., Al-Hada N.M. Kharazmi A. Structural, optical, and magnetic characterization of spinel zinc chromite nanocrystallines synthesised by thermal treatment method // *Journal of Nanomaterials*, 2014, vol. 2014, pp. 416765. https://doi.org/10.1155/2014/416765
- Bokuniaeva A.O., Vorokh A.S. Estimation of particle size using the Debye equation and the Scherrer formula for polyphasic TiO₂ powder. *Journal of Physics: Conference Series*, 2019, vol. 1410, no. 1, pp. 012057. https://doi.org/10.1088/1742-6596/1410/1/012057
- 33. Shannon R.D. Revised effective ionic radii and systematic studies of interatomic distances in halides and chalcogenides. *Acta Crystallographica Section A*, 1976, vol. 32, no. 5, pp. 751–767. https://doi.org/10.1107/S0567739476001551
- Wang X., Ahmad M., Sun H. Three-dimensional ZnO hierarchical nanostructures: solution phase synthesis and applications. *Materials*, 2017, vol. 10, no. 11, pp. 1304. https://doi.org/10.3390/ma10111304
- Børseth T.M., Svensson B.G., Kuznetsov A.Yu., Klason P., Zhao Q.X., Willander M. Identification of oxygen and zinc vacancy optical signals in ZnO. *Applied Physics Letters*, 2006, vol. 89, no. 26, pp. 262112. https://doi.org/10.1063/1.2424641
- Rodnyi P.A., Chernenko K.A., Venevtsev I.D. Mechanisms of ZnO Luminescence in the Visible Spectral Region. *Optics and Spectroscopy*, 2018, vol. 125, no. 3, pp. 372–378. https://doi. org/10.1134/S0030400X18090205
- Cao B.Q., Cai W.P., Zeng H.B. Temperature-dependent shifts of three emission bands for ZnO nanoneedle arrays. *Applied Physics Letters*, 2006, vol. 88, no. 16, pp. 161101. https://doi.org/10.1063/1.2195694
- Vempati S., Mitra J., Dawson P. One-step synthesis of ZnO nanosheets: a blue-white fluorophore. *Nanoscale Research Letters*, 2012, vol. 7, pp. 470. https://doi.org/10.1186/1556-276X-7-470
- Ghosh D., Dutta U., Haque A., Mordvinova N.E., Lebedev O.I., Pal K., Gayen A., Seikh M.M., Mahata P. Ultra-high sensitivity of luminescent ZnCr₂O₄ nanoparticles toward nitroaromatic explosives sensing. *Dalton Transactions*, 2018, vol. 47, no. 14, pp. 5011–5018. https://doi.org/10.1039/C8DT00047F
- Nosaka Y., Daimon T., Nosaka A.Y., Murakami Y. Singlet oxygen formation in photocatalytic TiO₂ aqueous suspension. *Physical Chemistry Chemical Physics*, 2004, vol. 6, no. 11, pp. 2917–2918. https://doi.org/10.1039/b405084c
- Abbasi A., Hamadanian M., Salavati-Niasari M., Mortazavi-Derazkola S. Facile size-controlled preparation of highly photocatalytically active ZnCr₂O₄ and ZnCr₂O₄/Ag nanostructures for removal of organic contaminants. *Journal of Colloid and Interface Science*, 2017, vol. 500, pp. 276–284. https://doi.org/10.1016/j.jcis.2017.04.003
- Dumitru R., Manea F., Păcurariu C., Lupa L., Pop A., Cioabla A., Surdu A., Ianculescu A. Synthesis, characterization of nanosized ZnCr₂O₄ and its photocatalytic performance in the degradation of humic acid from drinking water. *Catalysts*, 2018, vol. 8, no. 5, pp. 210. https://doi.org/10.3390/catal8050210
- Khomutinnikova L., Evstropiev S., Meshkovskii I., Bagrov I., Kiselev V. Ceramic ZnO-SnO₂-Fe₂O₃ powders and coatings -effective photogenerators of reactive oxygen species. *Ceramics*, 2023, vol. 6, no. 2, pp. 886–897. https://doi.org/10.3390/ceramics6020051

- Schweitzer C., Schmidt R. Physical Mechanisms of Generation and Deactivation of Singlet Oxygen // Chemical Reviews. 2003. V. 103. N 5. P. 1685–1758. https://doi.org/10.1021/cr010371d
- Konstantinou I.K., Albanis T.A. TiO2-assisted photocatalytic degradation of azo dyes in aqueous solution: kinetic and mechanistic investigations: a review // Applied Catalysis B: Environmental. 2004. V. 49. N 1. P. 1–14. https://doi.org/10.1016/j.apcatb.2003.11.010

Авторы

Евстропьев Сергей Константинович — доктор химических наук, профессор, начальник отдела, Научно-производственное объединение Государственный оптический институт им. С.И. Вавилова, Санкт-Петербург, 192171, Российская Федерация; профессор, Санкт-Петербургский государственный технологический институт (технический университет), Санкт-Петербург, 190013, Российская Федерация; ведущий инженер, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, вс 6507317768, https://orcid.org/0000-0002-0160-8443, evstropiev@bk.ru

Шелеманов Андрей Александрович — ассистент, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, sc 57292759900, https://orcid.org/0000-0001-5854-9475, Shelemanov@ mail.ru

Никоноров Николай Валентинович — доктор физико-математических наук, профессор, директор научно-исследовательского центра оптического материаловедения, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, № 7003772604, https:// orcid.org/0000-0002-1341-067X, nikonorov@oi.ifmo.ru

Караваева Анна Владимировна — кандидат биологических наук, доцент, старший научный сотрудник, Санкт-Петербургский государственный химико-фармацевтический университет, Санкт-Петербург, 197022, Российская Федерация, вс 6602806968, https://orcid.org/0000-0001-8231-6364, anna.karavaeva@pharminnotech.com

Дукельский Константин Владимирович — доктор технических наук, доцент, первый заместитель генерального директора, Научнопроизводственное объединение Государственный оптический институт им. С.И. Вавилова, Санкт-Петербург, 192171, Российская Федерация, вс 6602633236, https://orcid.org/0000-0002-1627-7499, kdukel@mail.ru

Полишук Григорий Сергеевич — кандидат технических наук, генеральный директор, Научно-производственное объединение Государственный оптический институт им. С.И. Вавилова, Санкт-Петербург, 192171, Российская Федерация, sc 25926044900, https:// orcid.org/0009-0005-9358-6040, g.polishchuk@optrotech.ru

Гаврилова Марианна Александровна – инженер, Санкт-Петербургский государственный технологический институт (технический университет), Санкт-Петербург, 190013, Российская Федерация, SC 57983736200, https://orcid.org/0000-0001-6163-9316, amonrud@ vandex.ru

Портнова Ксения Александровна — ассистент, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, https://orcid.org/0009-0004-3825-1291, ksiu.san@gmail.com

Багров Игорь Викторович — кандидат физико-математических наук, старший научный сотрудник, Научно-производственное объединение Государственный оптический институт им. С.И. Вавилова, Санкт-Петербург, 192171, Российская Федерация, https://orcid. org/0009-0005-2174-6412, i.bagrov2@gmail.com

Статья поступила в редакцию 27.01.2025 Одобрена после рецензирования 17.02.2025 Принята к печати 26.03.2025

- Schweitzer C., Schmidt R. Physical Mechanisms of Generation and Deactivation of Singlet Oxygen. *Chemical Reviews*, 2003, vol. 103, no. 5, pp. 1685–1758. https://doi.org/10.1021/cr010371d
- 45. Konstantinou I.K., Albanis T.A. TiO₂-assisted photocatalytic degradation of azo dyes in aqueous solution: kinetic and mechanistic investigations: a review. *Applied Catalysis B: Environmental*, 2004, vol. 49, no. 1, pp. 1–14. https://doi.org/10.1016/j.apcatb.2003.11.010

Authors

Sergey K. Evstropyev — D.Sc. (Chemistry), Professor, Chief of Department, JSC S.I. Vavilov State Optical Institute (SOI), Saint Petersburg, 192171, Russian Federation; Professor, Saint Petersburg State Technological Institute (Technical University), Saint Petersburg, 190013, Russian Federation; Leading Engineer, ITMO University, Saint Petersburg, 197101, Russian Federation, Sc 6507317768, https://orcid. org/0000-0002-0160-8443, evstropiev@bk.ru

Andrey A. Shelemanov — Assistant, ITMO University, Saint Petersburg, 197101, Russian Federation, SC 57292759900, https://orcid.org/0000-0001-5854-9475, Shelemanov@mail.ru

Nikolay V. Nikonorov — D.Sc. (Physics & Mathematics), Professor, Head of the Research Center for Optical Material Engineering, ITMO University, Saint Petersburg, 197101, Russian Federation, sc 7003772604, https://orcid.org/0000-0002-1341-067X, nikonorov@oi.ifmo.ru

Anna V. Karavaeva — PhD (Biology), Associate Professor, Senior Researcher, Saint Petersburg Chemical-Pharmaceutical University, Saint Petersburg, 197022, Russian Federation, Sc 6602806968, https://orcid. org/0000-0001-8231-6364, anna.karavaeva@pharminnotech.com

Konstantin V. Dukelskii – D.Sc., Associate Professor, First Deputy General Director, JSC S.I. Vavilov State Optical Institute (SOI), Saint Petersburg, 192171, Russian Federation, sc 6602633236, https://orcid. org/0000-0002-1627-7499, kdukel@mail.ru

Grigorii S. Polischuk — PhD, CEO, JSC S.I. Vavilov State Optical Institute (SOI), Saint Petersburg, 192171, Russian Federation, SC 25926044900, https://orcid.org/0009-0005-9358-6040, g.polishchuk@ optrotech.ru

Marianna A. Gavrilova — Engineer, Saint Petersburg State Technological Institute (Technical University), Saint Petersburg, 190013, Russian Federation, sc 57983736200, https://orcid.org/0000-0001-6163-9316, amonrud@yandex.ru

Ksenia A. Portnova — Assistant, ITMO University, Saint Petersburg, 197101, Russian Federation, https://orcid.org/0009-0004-3825-1291, ksiu.san@gmail.com

Igor V. Bagrov — PhD, Senior Researcher, JSC S.I. Vavilov State Optical Institute (SOI), Saint Petersburg, 192171, Russian Federation, https://orcid.org/0009-0005-2174-6412, i.bagrov2@gmail.com

Received 27.01.2025 Approved after reviewing 17.02.2025 Accepted 26.03.2025



Работа доступна по лицензии Creative Commons «Attribution-NonCommercial» **I/İTMO**

НАУЧНО-ТЕХНИЧЕСКИЙ ВЕСТНИК ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ, МЕХАНИКИ И ОПТИКИ март–апрель 2025 Том 25 № 2 http://ntv.ifmo.ru/ SCIENTIFIC AND TECHNICAL JOURNAL OF INFORMATION TECHNOLOGIES, MECHANICS AND OPTICS March–April 2025 Vol. 25 No 2 http://ntv.ifmo.ru/en/ ISSN 2226-1494 (print) ISSN 2500-0373 (online)

ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ, МЕХАНИКИ И ОПТИКИ

университет итмо

doi: 10.17586/2226-1494-2025-25-2-212-221

Analysis of the influence of defocused laser beam on uneven material surface processing based on mathematical model and simulation approach Muhamad Albani Rizki¹, Yuri V. Fedosov²

1,2 ITMO University, Saint Petersburg, 197101, Russian Federation

¹ muhamadalbanirizki@gmail.com^{\Box}, https://orcid.org/0000-0001-7502-1699

² Yf01@yandex.ru, https://orcid.org/0000-0003-1869-0081

Abstract

Laser-processing technology has advanced precision surface material processing, but challenges remain in maintaining the laser beam waist position on uneven surfaces. Surface irregularities cause defocus and non-perpendicular alignment leading to distortions in beam spot size and shape, which reduce processing quality. This study develops a mathematical model and simulation framework to analyze beam waist positioning errors during surface processing. Using MATLAB Partial Differential Equation (PDE) and finite element method, the simulation evaluates how variables like laser incidence angle and focal distance affect beam spot characteristics. Results reveal that defocus and misalignment enlarge and distort the laser beam spot, with higher incidence angles causing elliptical deformation. The simulation is critical in advancing the understanding of laser-material interactions under suboptimal conditions such as defocus and misalignment. It provides critical insights into the geometrical of laser beam, enabling the development of precise error detection methods for beam spot irregularities. Furthermore, these findings lay the groundwork for designing adaptive mechanisms that enhance the precision and reliability of laser-based surface material processing, addressing challenges posed by uneven workpiece surfaces. This approach aims to optimize laser processing quality and expand its applicability in high-precision manufacturing.

Keywords

laser processing technology, defocus laser beam, mathematical model, simulation of laser beam spot, laser beam spot geometry, beam projection on the surface, material surface processing

For citation: Rizki M.A., Fedosov Yu.V. Analysis of the influence of defocused laser beam on uneven material surface processing based on mathematical model and simulation approach. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 2025, vol. 25, no. 2, pp. 212–221. doi: 10.17586/2226-1494-2025-25-2-212-221

УДК 681.5.073

Математический и имитационный анализ влияния расфокусировки лазерного пучка на процесс обработки неровной поверхности Мухамад Албани Ризки1[™], Юрий Валерьевич Федосов²

1,2 Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация

¹ muhamadalbanirizki@gmail.com^{\Box}, https://orcid.org/0000-0001-7502-1699

² Yf01@yandex.ru, https://orcid.org/0000-0003-1869-0081

Аннотация

Введение. Технология лазерной обработки позволила повысить точность обработки поверхности заготовки. При этом сохраняются проблемы, связанные с поддержанием положения перетяжки пучка на неровных поверхностях. Неровности поверхности вызывают расфокусировку и неперпендикулярное выравнивание, что приводит к искажениям размера и формы лазерного пятна, снижающим качество обработки. Метод. Разработана математическая модель и схема моделирования для анализа ошибок позиционирования лазерного пучка при обработке поверхностей. Используя средства решения дифференциальных уравнений MATLAB и метод конечных элементов, моделирование позволило рассчитать величины изменения угла падения лазерного излучения и фокусного расстояния, влияющие на характеристики проекции луча на поверхность. Основные

© Rizki M.A., Fedosov Yu.V., 2025

результаты. Полученные результаты показывают, что расфокусировка и несоосность увеличивают и искажают проекцию пучка на поверхность, а большие углы падения вызывают эллиптическую деформацию. **Обсуждение.** Моделирование имеет решающее значение для понимания взаимодействия лазерного излучения и материала в неоптимальных условиях, таких как расфокусировка и несоосность. Оно дает критическое представление о геометрических искажениях лазерного пучка, позволяя разработать точные методы обнаружения ошибок проекции пучка на поверхность. Кроме того, эти результаты закладывают основу для разработки адаптивных механизмов, повышающих точность и надежность лазерной обработки поверхности материала и решающих проблемы, связанные с неровной поверхностью заготовок. Этот подход направлен на оптимизацию качества лазерной обработки и расширение ее применимости в высокоточном производстве.

Ключевые слова

технология лазерной обработки, расфокусировка лазерного пучка, математическая модель, моделирование, геометрическая форма лазерного пятна, проекция пучка на поверхность, обработка поверхности материала

Ссылка для цитирования: Ризки М.А., Федосов Ю.В. Математический и имитационный анализ влияния расфокусировки лазерного пучка на процесс обработки неровной поверхности // Научно-технический вестник информационных технологий, механики и оптики. 2025. Т. 25, № 2. С. 212–221 (на англ. яз.). doi: 10.17586/2226-1494-2025-25-2-212-221

Introduction

In recent years, laser-processing technology has been widely developed with the aim of solving various needs, especially in surface material processing and micromachining. This is inseparable from the advantages of laser-processing, namely, high flexibility of processing, high processing accuracy, non-contact processing, strong adaptability of processing materials, and high energy density. In addition, the main advantage of laser processing that cannot be achieved by conventional machining technology is that laser processing technology makes it possible to process materials on a micro scale or commonly referred to as micromachining. Because of these advantages, laser-processing technology is widely applied to various surface material processing, such as annealing [1]; ablation [2–4]; marking [5]; texturing [6–9]; etching [10, 11]; patterning [12]; and polishing [13, 14].

Surface material processing is a material processing that requires a high level of accuracy or commonly referred to as high precision machining. In laser processing, high precision machining is greatly influenced by the distance between the laser focus (beam waist) and the surface of the material or workpiece being processed. During laser processing, it is very important to ensure the stability of the focus beam waist position right on the surface of the material or workpiece. This is because an unstable focus beam position or defocus will cause the laser energy density in the irradiated area to be unstable and reduce the precision and quality of the material or workpiece being processed [15]. But in reality maintaining the stability of the focus beam position in laser processing is not something that is easy in material processing, especially micromachining and surface material processing.

The main factor that causes laser defocus in surface processing is the uneven surface of the material or workpiece. Uneven surfaces result in the angle of laser beam incidence not being perpendicular to the material being processed, which results in changes in the shape and size of the laser beam spot thereby reducing the intensity of the laser energy. It is also known that the greater the inclination of the laser beam angle to the surface of material or workpiece, the wider the irradiated area will be and the intensity of the laser energy will also decrease [16]. Therefore, to be able to solve the problem of laser defocus and to maintain the stability of the focus beam on the uneven surface of the workpiece, research is needed on the development of laser beam position error compensation methods, especially in surface processing materials.

Several studies have explored methods to optimize laser processing and minimize distortions. Research [17] investigated how laser scanning patterns influence residual stress and distortions in laser-processed components, emphasizing that proper beam path control significantly improves dimensional accuracy. Research [18] introduced a Digital Micromirror Device for beam shaping, demonstrating that optimized beam control can reduce thermal distortions. However, while these studies focus on beam shaping and scanning, they do not directly address the effects of a defocused laser beam on uneven surfaces.

In the field of laser welding and bending, distortion remains a critical issue. Research [19] developed a model to predict residual stresses and distortions in Laser Beam Welding of aluminum lap joints, showing how thermal effects influence stress distribution. Similarly, research [20] studied edge effects and longitudinal distortions in the laser bending process, revealing how uneven heat distribution leads to deformation inaccuracies. Research [21] expanded this investigation by examining nonconventional laser beam geometries and their impact on stress distribution in laser bending of tubes. These studies collectively highlight the thermal effects of laser processing, but they focus on structured materials rather than surface processing of uneven materials.

More relevant to defocused laser beam effects, research [22] introduced a hybrid model that integrates Monte Carlo simulations and Shack-Hartman wavefront sensing to analyze laser beam distortion. Their approach provides a foundation for understanding how beam deviations affect processing accuracy. While these studies explore beam deviations and correction strategies, they primarily focus on beam optimization; they did not specifically analyze the effects of defocused laser beam on uneven surfaces.

Given this gap in existing research, in this study, a comprehensive mathematical model and simulation framework are proposed to investigate the characteristics of laser beam spots during material surface processing, particularly under conditions where positioning errors, such as defocus and non-perpendicular alignment, occur. The developed model provides a detailed analysis of how these errors affect critical parameters, including the size and shape of the laser beam spot on uneven material surfaces. By incorporating factors such as the laser incidence angle and focal distance, the simulation evaluates the impact of surface irregularities on processing accuracy and quality.

The findings from this study can also serve as a foundation for future research aimed at developing compensation mechanisms to correct laser beam spot defocused. By leveraging the insights gained from the mathematical model and simulation results, future work can explore strategies to improve machining precision and mitigate the effects of beam defocused on material processing quality.

Laser Beam Positioning Errors on Material Surface Processing

Material surface processing, commonly known as high-precision machining, demands an exceptional level of accuracy. In the realm of laser processing, achieving high precision is intricately linked to the precise distance between the laser focus (beam waist) and the surface of the material or workpiece undergoing precise. The beam waist, also referred to as laser focus or focus beam, designates the region where the laser beam is at its most constricted, exhibiting a minimal beam radius [23]. Ensuring the stability of the beam waist position precisely on the material or workpiece surface is paramount during laser processing. An unstable focus or beam waist position or defocus can result in fluctuating laser energy density in the irradiated area, subsequently compromising the precision and quality of the processed material or workpiece. Two types of defocus are observed in laser processing: positive and negative defocus. Positive defocus occurs when

the laser beam focuses above the workpiece or material surface, aligning with the position of the focus beam. Conversely, negative defocus transpires when the focus is below the workpiece or material surface, or the focus beam is positioned beneath it. For a detailed visual representation, refer to Fig. 1.

Ensuring the stability of the beam waist position in laser processing proves to be a challenging endeavor, particularly in material processing realms such as micromachining and surface material processing. The primary culprit inducing laser defocus in surface processing is the irregularity of the material or workpiece surface. These uneven surfaces disrupt the perpendicular alignment of the laser beam incidence to the material being processed, triggering alterations in the shape and size of the laser beam spot and consequently diminishing the overall intensity of the laser energy. It is further established that a higher inclination of the laser beam angle to the material or workpiece surface leads to a broader irradiated area accompanied by a concurrent reduction in the intensity of the laser energy. For more details please refer to Fig. 2.

To effectively address the challenges of laser defocus and ensure the stability of the focus beam on uneven material or workpiece surfaces, dedicated research is essential for developing laser beam position error compensation methods, particularly tailored for surface processing materials. Confronting issues arising from dynamic nature of the laser beam spot characterized by fluctuations in shape, size, or errors on irregular materials or workpieces surfaces, demands the implementation of a method that ensures the steadfast stability. In the context of this study, an innovative method is currently under development, featuring adaptive capabilities within the laser beam system. This method empowers the laser to dynamically adapt to the ever-changing materials or workpieces surfaces, ensuring a consistently perpendicular orientation as shown in Fig. 3.



Fig. 1. Scheme of focus and defocus beam position in laser processing: focus (*a*); positive defocus (*b*); negative defocus (*c*); Δf is the distance between the beam waist and the material surface


Fig. 2. Scheme of laser beam spot on material processing with: even/flat material surface and perpendicular beam — focus laser beam spot with round shape and high laser energy intensity (*a*); uneven material surface and inclined beam — defocus laser beam spot with eclipse shape and low laser energy intensity (*b*)



Fig. 3. Conceptual framework for an adaptive laser system designed to stabilize the focus of laser beams and compensate for positioning errors on uneven material surfaces

Such adaptability not only maintains precise beam waist positioning but sustains it with meticulous accuracy throughout the entire laser processing procedure. Consequently, the proposed method stands as a promising solution to either eliminate or compensate for laser beam position errors during material surface processing.

Mathematical Model of Laser Beam Positioning Errors on Material Surface Processing

A mathematical model serves as a systematic approach for representing and elucidating real-world systems and phenomena by employing mathematical formulas and descriptive methodologies. Within the scope of this study, the mathematical model is designed to meticulously investigate, analyze, and forecast the behavior of a laser beam when interacting with both uneven and inclined surfaces of a material. The primary objective is to rectify or eliminate positional errors exhibited by the laser beam on the material's surface, thereby enhancing precision and eliminating discrepancies in its positioning.

Laser beam exhibits a strong correlation with Gaussian beams [24]. This connection arises from the fact that the majority of lasers are engineered to function in the lowest transverse mode known as Gaussian beams. This deliberate design choice aims to achieve the most focused beam possible, a critical factor in applications like material processing. Consequently, the mathematical modeling undertaken in this study is grounded in the principles of Gaussian beam modeling.

The mathematical model employed in this study is rooted in the surface heat source model outlined in reference [16]. The primary objective of this model is to elucidate the distribution of energy intensity across irregular or inclined material surfaces. Understanding the intricacies of the laser beam spot surface heat source model allows for insight into how uneven or inclined material surface influences the laser beam spot shape, size, and energy intensity. Illustration in Fig. 4 is a schematic representation of the laser beam interacting with the material surface. Here, α denotes the incidence angle of the laser beam due to the inclined surface, *h* represents the distance to the focal plane or laser beam focus, and w(z) indicates the radius where intensity values decrease to $1/e^2$ of their axial values. Additionally, w_0 signifies the radius of the laser beam focus or beam waist radius, while θ represents the divergence half-angle.

In order to ascertain the intensity distribution of the laser beam spot on the material perpendicular surface, mathematical equations are utilized. Within this framework, the parameter w_1 represents the laser beam radius specifically on the perpendicular surface – 1 (defocus).

$$q_1(x, y) = q_{axis} \exp\left[-2\left(\frac{x^2 + y^2}{w_1^2}\right)\right],$$
 (1)

where q_{axis} is the on-axis intensity value. Obtaining the total power in the laser beam is determined by applying equation:

$$P\eta = q_{axis} \int_{-a_y - a_y}^{a_y - a_x} \exp\left[-2\left(\frac{x^2 + y^2}{w_1^2}\right)\right] dxdy,$$
(2)

where $P\eta$ is the power or energy of the laser beam within a specific area.

By merging equations (1) and (2), the resulting heat source model aligns with the formulation presented in the following equation:

$$q_1(x, y) = \frac{2P\eta}{\pi w_1^2} \exp\left[-2\left(\frac{x^2 + y^2}{w_1^2}\right)\right].$$
 (3)

The value of w_1 from equation (3) is obtained using the following equation:

$$w_1^2 = w_0^2 + \theta^2 h^2. \tag{4}$$

Meanwhile, the intensity distribution of the laser beam spot on the inclined surface of the material can be determined by consulting the following equation:

$$q_2(x, y, z) = q_{axis} \exp\left[-2\left(\frac{x^2 + y^2}{w_2^2(z)}\right)\right].$$
 (5)

The laser beam radius on the inclined surface — 2 (defocus), denoted as w_2 , is defined by the equation below which has the same principle with the equation (4).

$$w_2^2(z) = w_0^2 + \theta^2 z^2 = w_0^2 + \theta^2 (h + x \tan \alpha)^2.$$
 (6)

Transforming the coordinate system from the laser beam to the material surface can be expressed through the following equation:

$$x = X \cos \alpha, y = Y, \tag{7}$$

where x and y are coordinates in the material surface reference frame. X and Y are coordinates in the laser beam reference frame.

Combining equations (6) and (7) into equation (5) leads to the derived equation:

$$q_2(X, Y) = q_{axis} \exp\left[-2\left(\frac{(X\cos\alpha)^2 + Y^2}{w_0^2 + \theta^2(h + X\sin\alpha)^2}\right)\right].$$
 (8)



Fig. 4. Schematic of laser beam interaction on material surface; r_T is minor axis of ellipse; r_L is major axis of ellipse

Similar to equation (2), the total power of the laser beam can be computed using the following equation:

$$P\eta = \int_{-a_y - a_y}^{a_y - a_x} \exp\left[-2\left(\frac{(X\cos\alpha)^2 + Y^2}{w_0^2 + \theta^2(h + X\sin\alpha)^2}\right)\right] dxdy.$$
(9)

For simplifying the calculation, new parameter D is introduced, as defined in the following equation:

$$D = \int_{-3r_T - 3r_L}^{3r_L} \exp\left[-2\left(\frac{(X\cos\alpha)^2 + Y^2}{w_0^2 + \theta^2(h + X\sin\alpha)^2}\right)\right] dxdy, \quad (10)$$

where r_T and r_L are obtained according to the equations below:

$$r_T = \sqrt{w_0^2 + \theta^2 h^2 + \frac{(\theta^2 h \sin \alpha)^2}{\cos^2 \alpha - \cos^2 \sin^2 \alpha}},$$
$$r_L = \sqrt{\frac{w_0^2 + \theta^2 + h^2}{\cos^2 \alpha - \theta^2 \sin^2 \alpha}} + \frac{(\theta^2 h \sin \alpha)^2}{(\cos^2 \alpha - \theta^2 \sin^2 \alpha)^2}.$$

By combining equation (8), (9), and (10), the heat source model equation can be calculated with the equation below:

$$q_2(X, Y) = \frac{P\eta}{D} \exp\left[-2\left(\frac{(X\cos\alpha)^2 + Y^2}{w_0^2 + \theta^2(h + X\sin\alpha)^2}\right)\right].$$

The mathematical model describing the laser beam spot on the material surface serves as a valuable tool for analyzing the characteristics of the laser beam spot. It facilitates the development of simulations to study variations in shape, size, and energy intensity of the laser beam spot on materials with uneven surfaces.

Development of Laser Beam Spot Simulation on Material Surface for Analyzing the Laser Beam Positioning Errors

This simulation is designed to investigate the impact of laser beam positioning errors on material or workpiece surfaces, particularly on those with irregularities. The underlying mathematical model serves as the foundation for the simulation development. Beyond assessing changes in the shape and size of the laser beam spot (defocus) due to positioning errors, this simulation also serves the crucial purpose of gauging the laser beam energy intensity and heat distribution is facilitated by employing a Partial Differential Equation (PDE), as denoted in the following equation within the computation domain:

$$\rho C \frac{\partial T}{\partial t} - \nabla (k \nabla T) = Q,$$

where ρ is the density, *C* is the heat capacity, *T* is the temperature, *t* is the time, *k* is the coefficient of heat conduction, ∇ is partial derivative operator, and *Q* is the heat source.

By combining the parameters of material properties and laser beam parameters, the simulation of surface heat

Table 1. Surface heat treatment parameters list

Constant	Value	Description
<i>w</i> ₀ , mm	1	Beam waist radius
θ, rad	0.033	Divergence half-angle
<i>P</i> , W	260	Laser power
<i>C</i> , J/(kg·K)	0.54	Heat capacity
$k, J/(mm \cdot K)$	0.32	Coefficient of heat conduction
ρ, g/mm ³	0.078	Density

treatment can be developed. The parameters are described in Table 1.

Moreover, leveraging these parameters enables the creation of simulation aimed at elucidating the impact of laser beam positioning errors on materials characterized by uneven surfaces. These simulations incorporate variables such as the incidence angle of the laser beam resulting from the inclined surface α and encompass the representation of the distance to the focal plane or laser beam focus *h*.

The simulation is executed using the PDE solver within the MATLAB software, employing the Finite Element method. This methodology ensures a swift and efficient solution to the surface heat treatment problem. The focus of this simulation is on surface heat treatment, aiming to ascertain the dimensions, shape, and the laser beam spot energy intensity or heat distribution on the material surface. Introducing the element of laser beam positioning error on uneven surfaces, the key variables under scrutiny are the incidence angle of the laser beam due to the inclined surface α and the representation of the distance to the focal plane or laser beam focus *h*. To enhance clarity, the conducted tests are detailed in Table 2.

Ten distinct conditions align with the variables α and h. The synthesis of parameters from Table 1 and incorporation of variables from Table 2 facilitate the development of a comprehensive simulation for surface heat treatment of the laser beam spot on materials with uneven surfaces. For more in-depth information about results in this simulation, refer to Fig. 5, 6, and 7.

The outcomes from the laser beam simulation conducted on the material surface underscore a significant impact of the material or workpiece surface on the shape and size of the laser beam spot. This alteration in the laser beams characteristics on the material surface is termed as the laser beam positioning error. The ongoing emphasis is



Fig. 5. Surface heat treatment of positive laser beam spot defocus on even/flat material surface with variable: $\alpha = 0$, h = 0 mm (*a*); $\alpha = 0$, h = 25 mm (*b*); $\alpha = 0$, h = 50 mm (*c*); $\alpha = 0$, h = 75 mm (*d*)

Conditions	Inclined surface α, rad	Distance to focus <i>h</i> , mm	Beam spot radius — 1 r_T , mm	Beam spot radius — 2 r_L , mm
1	0	0	1.000	1.000
2	0	25	1.296	1.296
3	0	50	1.929	1.929
4	0	75	2.669	2.669
5	0.436	0	1.000	1.104
6	0.873	0	1.000	1.557
7	1.309	0	1.000	3.893
8	0.436	25	1.296	1.431
9	0.873	50	1.930	3.006
10	1.309	75	2.687	10.461

Table 2. Variables in surface heat treatment simulation



Fig. 6. Surface heat treatment of laser beam spot on inclined or uneven material surface with variable: $\alpha = 0$, h = 0 mm (*a*); $\alpha = 0.436$ rad, h = 0 mm (*b*); $\alpha = 0.873$ rad, h = 0 mm (*c*); $\alpha = 1.309$ rad, h = 0 mm (*d*)



Fig. 7. Surface heat treatment of positive laser beam spot defocus on inclined or uneven material surface with variable: $\alpha = 0, h = 0 \text{ mm}(a); \alpha = 0.436 \text{ rad}, h = 25 \text{ mm}(b);$ $\alpha = 0.873 \text{ rad}, h = 50 \text{ mm}(c); \alpha = 1.309 \text{ rad}, h = 75 \text{ mm}(d)$

placed on delving deeper into this laser beam positioning error, with the ultimate goal of compensating or eliminating this discrepancy through further study and analysis.

Analysis of Simulation Result of Incidence Angle and Focal Plane Distance of Laser Beam on Material Surface

The exploration of laser beam spot characteristics through surface heat treatment simulations is geared towards understanding the impact of material surface processing on the laser beam spot size and shape. In developing the simulation, two pivotal variables, namely, the incidence angle of the laser beam α and the distance from the focus beam or beam waist to the material surface h — contribute to the creation of 10 distinct conditions meticulously outlined in Table 2.

Condition 1 serves as a reference, situating the laser beam spot precisely at the beam waist or focus beam on the material surface, with $\alpha = 0$ rad and h = 0 mm. This baseline condition facilitates a comparative analysis, enabling a clearer evaluation of changes in size and shape across other laser beam spots.

Conditions 2 to 4 involve scenarios where the laser beam position is perpendicular to the material surface, yet the focus beam or beam waist is positioned above, creating a positive defocus situation. Positive defocus leads to an enlarged laser beam spot, with the size increasing proportionally to the distance between the beam waist and the material surface as depicted in Fig. 5 and Fig. 8 above.



Fig. 8. Comparison of changes in the shape and size of the laser beam spot based on conditions 2 to 4



Fig. 9. Comparison of changes in the shape and size of the laser beam spot based on conditions 5 to 7



Fig. 10. Comparison of changes in the shape and size of the laser beam spot based on conditions 8 to 10

For conditions 2 to 4, this distance is set at 25, 50, and 75 mm, respectively.

Moving to conditions 5 through 7, the laser beam position on the material surface becomes uneven or inclined. This inclination induces an angle altering the beam perpendicularity to the material surface, consequently changing the shape of the laser beam spot. As illustrated in Fig. 6 and 9 above, the shape transforms into an ellipse with increasing inclination angles set at 0.436, 0.873, and 1.309 rad for conditions 5 to 7.

Conditions 8 to 10 represent a combination of conditions 2 to 7, introducing positive defocus and an uneven or inclined material surface. This amalgamation results in size and shape variations compared to condition 1, with the laser beam spot enlarging and adopting a more ellipse shape as the distance and inclination angle increase. In these conditions, the distance from the beam waist or focus beam to the material surface is set at 25, 50, and 75 mm, while the inclination angles are 0.436, 0.873, and 1.309 rad, respectively, as depicted in Fig. 7 and Fig. 10 above.

Conclusion

This research endeavors to address and rectify laser beam position errors during surface processing of materials. The specific error involves the laser beam not being perpendicular to the workpiece surface, coupled with the beam waist or focus beam not aligning precisely with the surface. Uneven workpiece or material surfaces contribute to this discrepancy, inducing changes in the laser beam spot shape and size. These alterations negatively impact the quality of laser processing outcomes. To combat such laser position errors, this research introduces a method aimed at endowing the laser beam with adaptability to various workpiece surfaces, ensuring a constant perpendicular orientation and maintaining beam waist stability on the surface.

The proposed method is currently in the developmental phase, progressing through distinct stages to optimize its efficacy. In the ongoing research stage, a simulation has been devised to identify and validate the characteristics of laser beam positioning errors on material surfaces. Simulation results affirm that deviations in the laser beam spot shape and size are directly influenced by the non-perpendicular alignment of the laser to the workpiece surface and the defocused beam waist or focus. Additionally, it was observed that an increased distance between the focus beam or beam waist and the material surface correlates with a larger laser beam spot. Moreover, greater inclination angles introduced by the laser beam against the workpiece surface lead to an increasingly ellipse-shaped laser beam spot.

Subsequently, the data gleaned from the simulation serves as a foundational reference for real-world testing and comparison with experimental data. This empirical validation aims to enhance our understanding of laser beam position errors on material surfaces and serve as a pivotal point for the ongoing development of methods designed to detect and identify such errors in laser beam spots.

References

- Leniart A.A., Pula P., Sitkiewicz A., Majewski P.W. Macroscopic alignment of block copolymers on silicon substrates by laser annealing. ACS Nano, 2020, vol. 14, no. 4, pp. 4805–4815. https:// doi.org/10.1021/acsnano.0c00696
- Téllez H., Vadillo J.M., Chater R.J., Laserna J.J., McPhail D.S. Focused ion beam imaging of laser ablation sub-surface effects on layered materials. *Applied Surface Science*, 2008, vol. 255, no. 5, part 1, pp. 2265–2269. https://doi.org/10.1016/J. APSUSC.2008.07.082
- Shaheen M.E., Gagnon J.E., Fryer B.J. Studies on laser ablation of silicon using near IR picosecond and deep UV nanosecond lasers. *Optics and Lasers Engineering*, 2019, vol. 119, pp. 18–25. https://doi. org/10.1016/j.optlaseng.2019.02.003
- Žemaitis A., Gaidys M., Brikas M., Gečys P., Račiukaitis G., Gedvilas M. Advanced laser scanning for highly-efficient ablation and ultrafast surface structuring: experiment and model. *Scientific Reports*, 2018, vol. 8, pp. 17376. https://doi.org/10.1038/s41598-018-35604-z
- Sobotova L., Badida M. Laser marking as environment technology. *Open Engineering*, 2017, vol. 7, no. 1, pp. 303–316. https://doi. org/10.1515/eng-2017-0030
- Indrišiūnas S., Gedvilas M. Control of the wetting properties of stainless steel by ultrashort laser texturing using multi-parallel beam processing. *Optics & Laser Technology*, 2022, vol. 153, pp. 108187. https://doi.org/10.1016/j.optlastec.2022.108187
- Obilor A.F., Pacella M., Wilson A., Silberschmidt V.V. Microtexturing of polymer surfaces using lasers: a review. *The International Journal of Advanced Manufacturing Technology*, 2022, vol. 120, no. 1-2, pp. 103–135. https://doi.org/10.1007/s00170-022-08731-1
- Sugar P., Sugarova J., Frncik M. Laser surface texturing of tool steel: textured surfaces quality evaluation. *Open Engineering*, 2016, vol. 6, no. 1, pp. 90–97. https://doi.org/10.1515/eng-2016-0012
- Mao B., Siddaiah A., Liao Y., Menezes P.L. Laser surface texturing and related techniques for enhancing tribological performance of engineering materials: a review. *Journal of Manufacturing Processes*, 2020, vol. 53, pp. 153–173. https://doi.org/10.1016/j. jmapro.2020.02.009
- Yang Y., Zhao Y., Wang L., Zhao Y. Application of femtosecond laser etching in the fabrication of bulk SiC accelerometer. *Journal of Materials Research and Technology*, 2022, vol. 17, pp. 2577–2586. https://doi.org/10.1016/j.jmrt.2022.02.012
- Ehrhardt M., Lorenz P., Bauer J., Heinke R., Hossain M.A., Han B., Zimmer K. Dry etching of germanium with laser induced reactive micro plasma. *Lasers in Manufacturing and Materials Processing*, 2021, vol. 8, no. 3, pp. 237–255. https://doi.org/10.1007/s40516-021-00147-1
- Mulko L., Soldera M., Lasagni A.F. Structuring and functionalization of non-metallic materials using direct laser interference patterning: a review. *Nanophotonics*, 2022, vol. 11, no. 2, pp. 203–240. https://doi. org/10.1515/nanoph-2021-0591
- Krishnan A., Fang F. Review on mechanism and process of surface polishing using lasers. *Frontiers of Mechanical Engineering*, 2019, vol. 14, no. 3, pp. 299–319. https://doi.org/10.1007/s11465-019-0535-0
- Xiao H., Chen Y., Liu M., Zhou Y., Du C., Zhang W. Influence of laser additive manufacturing and laser polishing on microstructures and mechanical properties of high-strength maraging steel metal materials. *Applied Sciences*, 2022, vol. 12, no. 20, pp. 10340. https://doi. org/10.3390/app122010340
- Rizki M.A., Fedosov Yu.V. Development of adaptive laser head for compensating error of beam waist position during processing materials using laser beam spot detection method. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 2023, vol. 23, no. 5, pp. 859–870. https://doi. org/10.17586/2226-1494-2023-23-5-859-870
- Evdokimov A., Jasiewicz F., Doynov N., Ossenbrink R., Michailov V. Simulation of surface heat treatment with inclined laser beam. *Journal* of Manufacturing Processes, 2022, vol. 81, pp. 107–114. https://doi. org/10.1016/j.jmapro.2022.06.051
- Dai K., Shaw L. Distortion minimization of laser-processed components through control of laser scanning patterns. *Rapid Prototyping Journal*, 2002, vol. 8, no. 5, pp. 270–276. https://doi. org/10.1108/13552540210451732

Литература

- Leniart A.A., Pula P., Sitkiewicz A., Majewski P.W. Macroscopic alignment of block copolymers on silicon substrates by laser annealing // ACS Nano. 2020. V. 14. N 4. P. 4805–4815. https://doi. org/10.1021/acsnano.0c00696
- Téllez H., Vadillo J.M., Chater R.J., Laserna J.J., McPhail D.S. Focused ion beam imaging of laser ablation sub-surface effects on layered materials // Applied Surface Science. 2008. V. 255. N 5. Part 1. P. 2265–2269. https://doi.org/10.1016/J.APSUSC.2008.07.082
- Shaheen M.E., Gagnon J.E., Fryer B.J. Studies on laser ablation of silicon using near IR picosecond and deep UV nanosecond lasers // Optics and Lasers Engineering. 2019. V. 119. P. 18–25. https://doi. org/10.1016/j.optlaseng.2019.02.003
- Žemaitis A., Gaidys M., Brikas M., Gečys P., Račiukaitis G., Gedvilas M. Advanced laser scanning for highly-efficient ablation and ultrafast surface structuring: experiment and model // Scientific Reports. 2018. V. 8. P. 17376. https://doi.org/10.1038/s41598-018-35604-z
- Sobotova L., Badida M. Laser marking as environment technology // Open Engineering. 2017. V. 7. N 1. P. 303–316. https://doi. org/10.1515/eng-2017-0030
- Indrišiūnas S., Gedvilas M. Control of the wetting properties of stainless steel by ultrashort laser texturing using multi-parallel beam processing // Optics & Laser Technology. 2022. V. 153. P. 108187. https://doi.org/10.1016/j.optlastec.2022.108187
- Obilor A.F., Pacella M., Wilson A., Silberschmidt V.V. Microtexturing of polymer surfaces using lasers: a review // The International Journal of Advanced Manufacturing Technology. 2022. V. 120. N 1-2. P. 103–135. https://doi.org/10.1007/s00170-022-08731-1
- Sugar P., Sugarova J., Frncik M. Laser surface texturing of tool steel: textured surfaces quality evaluation // Open Engineering. 2016. V. 6. N 1. P. 90–97. https://doi.org/10.1515/eng-2016-0012
- Mao B., Siddaiah A., Liao Y., Menezes P.L. Laser surface texturing and related techniques for enhancing tribological performance of engineering materials: a review // Journal of Manufacturing Processes. 2020. V. 53. P. 153–173. https://doi.org/10.1016/j.jmapro.2020.02.009
- Yang Y., Zhao Y., Wang L., Zhao Y. Application of femtosecond laser etching in the fabrication of bulk SiC accelerometer // Journal of Materials Research and Technology. 2022. V. 17. P. 2577–2586. https://doi.org/10.1016/j.jmrt.2022.02.012
- Ehrhardt M., Lorenz P., Bauer J., Heinke R., Hossain M.A., Han B., Zimmer K. Dry etching of germanium with laser induced reactive micro plasma // Lasers in Manufacturing and Materials Processing. 2021. V. 8. N 3. P. 237–255. https://doi.org/10.1007/s40516-021-00147-1
- Mulko L., Soldera M., Lasagni A.F. Structuring and functionalization of non-metallic materials using direct laser interference patterning: a review // Nanophotonics. 2022. V. 11. N 2. P. 203–240. https://doi. org/10.1515/nanoph-2021-0591
- Krishnan A., Fang F. Review on mechanism and process of surface polishing using lasers // Frontiers of Mechanical Engineering. 2019. V. 14. N 3. P. 299–319. https://doi.org/10.1007/s11465-019-0535-0
- Xiao H., Chen Y., Liu M., Zhou Y., Du C., Zhang W. Influence of laser additive manufacturing and laser polishing on microstructures and mechanical properties of high-strength maraging steel metal materials // Applied Sciences. 2022. V. 12. N 20. P. 10340. https://doi. org/10.3390/app122010340
- Rizki M.A., Fedosov Yu.V. Development of adaptive laser head for compensating error of beam waist position during processing materials using laser beam spot detection method // Scientific and Technical Journal of Information Technologies, Mechanics and Optics. 2023. V. 23. N 5. P. 859–870. https://doi.org/10.17586/2226-1494-2023-23-5-859-870
- Evdokimov A., Jasiewicz F., Doynov N., Ossenbrink R., Michailov V. Simulation of surface heat treatment with inclined laser beam // Journal of Manufacturing Processes. 2022. V. 81. P. 107–114. https:// doi.org/10.1016/j.jmapro.2022.06.051
- Dai K., Shaw L. Distortion minimization of laser-processed components through control of laser scanning patterns // Rapid Prototyping Journal. 2002. V. 8. N 5. P. 270–276. https://doi. org/10.1108/13552540210451732
- Messaoudi H., Thiemicke F., Falldorf C., Bergmann R.B., Vollertsen F. Distortion-free laser beam shaping for material processing using a

- Messaoudi H., Thiemicke F., Falldorf C., Bergmann R.B., Vollertsen F. Distortion-free laser beam shaping for material processing using a digital micromirror device. *Production Engineering*, 2017, vol. 11, no. 3, pp. 365–371. https://doi. org/10.1007/s11740-017-0722-y
- Moraitis G.A., Labeas G.N. Residual stress and distortion calculation of laser beam welding for aluminum lap joints. *Journal of Materials Processing Technology*, 2008, vol. 198, no. 1-3, pp. 260–269. https:// doi.org/10.1016/j.jmatprotec.2007.07.013
- Zahrani E.G., Marasi A. Experimental investigation of edge effect and longitudinal distortion in laser bending process. *Optics and Laser Technology*, 2013, vol. 45, no. 1. pp. 301–307. https://doi. org/10.1016/j.optlastec.2012.06.031
- Safdar S., Li N., Sheikh M.A., Liu Z. The effect of nonconventional laser beam geometries on stress distribution and distortions in laser bending of tubes. *Journal of Manufacturing Science and Engineering*, 2007, vol. 129, no. 3, pp. 592–600. https://doi.org/10.1115/1.2716715
- Galaktionov I., Sheldakova J., Nikitin A., Toporovsky V., Kudryashov A. A hybrid model for analysis of laser beam distortions using monte carlo and Shack-Hartmann techniques: numerical study and experimental results. *Algorithms*, 2023, vol. 16, no. 7, pp. 337. https:// doi.org/10.3390/a16070337
- Marom D.M. Optical communications. *Comprehensive Microsystems*, 2007, vol. 3, pp. 219–265. https://doi.org/10.1016/B978-044452190-3.00035-5
- Paschotta R. Laser Beams. *RP Photonics Encyclopedia*, 2005, https:// doi.org/10.61835/ggt

Authors

Muhamad Albani Rizki — PhD Student, ITMO University, Saint Petersburg, 197101, Russian Federation, sc 58038476200, https://orcid. org/0000-0001-7502-1699, muhamadalbanirizki@gmail.com

Yuri V. Fedosov — PhD, Head of Laboratory, ITMO University, Saint Petersburg, 197101, Russian Federation, SC 57194080548, https://orcid. org/0000-0003-1869-0081, Yf01@yandex.ru

Received 19.11.2024 Approved after reviewing 18.02.2025 Accepted 22.03.2025 digital micromirror device // Production Engineering. 2017. V. 11. N 3. P. 365–371. https://doi.org/10.1007/s11740-017-0722-y

- Moraitis G.A., Labeas G.N. Residual stress and distortion calculation of laser beam welding for aluminum lap joints // Journal of Materials Processing Technology. 2008. V. 198. N 1-3. P. 260–269. https://doi. org/10.1016/j.jmatprotec.2007.07.013
- Zahrani E.G., Marasi A. Experimental investigation of edge effect and longitudinal distortion in laser bending process // Optics and Laser Technology. 2013. V. 45. N 1. P. 301–307. https://doi.org/10.1016/j. optlastec.2012.06.031
- Safdar S., Li N., Sheikh M.A., Liu Z. The effect of nonconventional laser beam geometries on stress distribution and distortions in laser bending of tubes // Journal of Manufacturing Science and Engineering. 2007. V. 129. N 3. P. 592–600. https://doi. org/10.1115/1.2716715
- 22. Galaktionov I., Sheldakova J., Nikitin A., Toporovsky V., Kudryashov A. A hybrid model for analysis of laser beam distortions using monte carlo and Shack–Hartmann techniques: numerical study and experimental results // Algorithms. 2023. V. 16. N 7. P. 337. https://doi.org/10.3390/a16070337
- Marom D.M. Optical communications // Comprehensive Microsystems. 2007. V. 3. P. 219–265. https://doi.org/10.1016/B978-044452190-3.00035-5
- Paschotta R. Laser Beams // RP Photonics Encyclopedia. 2005. https://doi.org/10.61835/ggt

Авторы

Ризки Мухамад Албани — аспирант, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, so 58038476200, https:// orcid.org/0000-0001-7502-1699, muhamadalbanirizki@gmail.com Федосов Юрий Валерьевич — кандидат технических наук, заведующий лабораторией, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, sc 57194080548, https://orcid.org/0000-0003-1869-0081, Yf01@yandex.ru

Статья поступила в редакцию 19.11.2024 Одобрена после рецензирования 18.02.2025 Принята к печати 22.03.2025



Работа доступна по лицензии Creative Commons «Attribution-NonCommercial» **I/İTMO**

НАУЧНО-ТЕХНИЧЕСКИЙ ВЕСТНИК ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ, МЕХАНИКИ И ОПТИКИ март–апрель 2025 Том 25 № 2 http://ntv.ifmo.ru/ SCIENTIFIC AND TECHNICAL JOURNAL OF INFORMATION TECHNOLOGIES, MECHANICS AND OPTICS March–April 2025 Vol. 25 No 2 http://ntv.ifmo.ru/en/ ISSN 2226-1494 (print) ISSN 2500-0373 (online)

ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ, МЕХАНИКИ И ОПТИКИ

университет итмо

doi: 10.17586/2226-1494-2025-25-2-222-228 УДК 681.7.068

Повышение чувствительности рефрактивного волоконно-оптического датчика на основе поверхностного плазмонного резонанса и MMF-SMF-MMF структуры с помощью диэлектрических покрытий

Аделия Алексеевна Зыкина¹[∞], Семён Алексеевич Плясцов², Игорь Аркадьевич Гладских³

1,2,3 Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация

² s.plyastsov@gmail.com, https://orcid.org/0000-0002-5764-0960

³ 138020@mail.ru, https://orcid.org/0000-0001-6389-1611

Аннотация

Введение. Представлены теоретические и экспериментальные исследования возможности повышения чувствительности рефрактивного волоконно-оптического датчика, принцип работы которого основан на эффекте поверхностного плазмонного резонанса. Конструкция чувствительного элемента состоит из последовательного соединения многомодового (MMF), одномодового (SMF) и многомодового волокна со структурой MMF-SMF-MMF. Для создания эффекта поверхностного плазмонного резонанса на одномодовый участок наносятся последовательно покрытия из металла (меди (Cu)) и диэлектрика (оксида алюминия (Al₂O₃), диоксида титана (TiO₂)), что позволяет получить более узкие резонансные пики. В результате упрощается процесс детектирования изменения длины волны и повышается чувствительность датчика. Метод. Математическое моделирование чувствительного элемента с многослойной поверхностной структурой выполнено с применением характеристических матриц. Матрицы составляются для каждого слоя чувствительного элемента в отдельности, после чего формируется характеристическая матрица всей структуры и рассчитывается коэффициент пропускания. По результатам моделирования выбираются варианты диэлектрического покрытия, определяются толщины слоев, позволяющие получить наименьшую ширину резонансного пика. Основные результаты. Для апробации результатов моделирования изготовлены образцы чувствительных элементов с диэлектрическим покрытием Al₂O₃ толщиной 60 и 100 нм и с покрытием TiO₂ толщиной 50 и 100 нм. Получены спектры пропускания чувствительных элементов на воздухе, в воде и спирте. Показано, что покрытие TiO₂ толщиной 50 нм позволяет повысить чувствительность волоконно-оптического датчика на основе поверхностного плазмонного резонанса в три раза по сравнению с чувствительным элементом без диэлектрического покрытия. Обсуждение. Предложенный подход к усовершенствованию конструкции волоконно-оптического датчика дает возможность сместить получаемые резонансные максимумы в инфракрасную область спектра. Показана возможность применения более доступных волоконно-оптических элементов для работы с исследуемым латчиком.

Ключевые слова

поверхностный плазмонный резонанс, гетеросердцевинная структура, оптоволоконный датчик, повышение чувствительности волоконно-оптических датчиков, многослойные покрытия

Ссылка для цитирования: Зыкина А.А., Плясцов С.А., Гладских И.А. Повышение чувствительности рефрактивного волоконно-оптического датчика на основе поверхностного плазмонного резонанса и MMF-SMF-MMF структуры с помощью диэлектрических покрытий // Научно-технический вестник информационных технологий, механики и оптики. 2025. Т. 25, № 2. С. 222–228. doi: 10.17586/2226-1494-2025-25-2-222-228

[©] Зыкина А.А., Плясцов С.А., Гладских И.А., 2025

Increasing the sensitivity of a refractive fiber-optic sensor based on surface plasmon resonance and MMF-SMF-MMF fiber using dielectric coatings

Adeliia A. Zykina^{1⊠}, Semyon A. Plyastsov², Igor A. Gladskikh³

1,2,3 ITMO University, Saint Petersburg, 197101, Russian Federation

¹ adeliya.zykina@gmail.com^{\boxdot}, https://orcid.org/0000-0002-8021-0678

² s.plyastsov@gmail.com, https://orcid.org/0000-0002-5764-0960

³ 138020@mail.ru, https://orcid.org/0000-0001-6389-1611

Abstract

This paper presents theoretical and experimental research on improving the sensitivity of a refractive fiber-optic sensor operating on the principle of surface plasmon resonance (SPR). The sensitive element consists of a multimode-singlemode-multimode (MMF-SMF-MMF) fiber structure. To induce the SPR effect, the single-mode fiber section is sequentially coated with metal (Cu) and dielectric layers (Al₂O₃, TiO₂), which results in narrower resonance peaks. This enhances wavelength shift detection and increases sensor sensitivity. Mathematical modeling of the sensitive element with a multilayer surface structure was conducted using characteristic matrices. Each layer of the sensitive element was individually characterized, followed by the formulation of the overall characteristic matrix to calculate the transmission coefficient. Based on the simulation results, optimal dielectric coatings and layer thicknesses were selected to achieve the narrowest resonance peak. To validate the simulation findings, sensitive element samples with dielectric coatings of Al₂O₃ (60 nm and 100 nm) and TiO₂ (50 nm and 100 nm) were fabricated. Transmission spectra were obtained in air, water, and ethanol. The results demonstrate that the proposed coating increases the sensitivity of the fiber-optic SPR sensor threefold compared to an uncoated sensitive element. The proposed approach not only enhances sensor sensitivity but also shifts the resonance peaks into the infrared spectral region. Additionally, the study highlights the feasibility of using more accessible fiber-optic components for the investigated sensor system.

Keywords

surface plasmon resonance, heterocore structure, fiber-optic sensor, increasing the sensitivity of fiber-optic sensors, multilayer coatings

For citation: Zykina A.A., Plyastsov S.A., Gladskikh I.A. Increasing the sensitivity of a refractive fiber-optic sensor based on surface plasmon resonance and MMF-SMF-MMF fiber using dielectric coatings. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 2025, vol. 25, no. 2, pp. 222–228 (in Russian). doi: 10.17586/2226-1494-2025-25-2-222-228

Введение

В настоящее время волоконно-оптические датчики (ВОД) на основе поверхностного плазмонного резонанса (ППР) активно исследуются для различных применений в сфере медицины и фармакологии [1]. Например, с их помощью возможно измерение уровня концентрации в крови сывороточного белка Цистатин С, что необходимо для определения исходного функционального состояния почек и последующего его мониторинга [2]. Другим важным перспективным применением ВОД на основе ППР является количественное определение остатков антибактериальных препаратов в пищевой промышленности [3]. Такие ВОД отличаются высокой чувствительностью, селективностью, воспроизводимостью результатов и отсутствием необходимости предварительной обработки исследуемых образцов [1].

Для опроса чувствительного элемента (ЧЭ) в ППРсистемах используются несколько вариантов подходов: фазовый метод опроса, опрос по интенсивности, измерение угла отражения излучения или длины волны пика поглощения ППР [4]. В первом случае получают синусоидальные спектральные интерференционные полосы с помощью детектирования фиксированной разности оптического пути между *s*- и *p*-поляризациями. Такой метод опроса наиболее чувствительный (10⁻⁸–10⁻⁹ RIU (Refractive Index Unit)), однако его существенным недостатком является значительная сложность и стоимость всей системы, так как для демодуляции фазы ППР требуется как минимум один электронный модуль, что также увеличивает вероятность зашумления сигнала [5]. При применении метода опроса по интенсивности для достижения наибольшей чувствительности (10⁻⁵) необходимо, чтобы угол падения и длина волны излучения были оптимизированы и строго зафиксированы [6]. В режиме углового опроса длина волны излучения фиксирована, и смещение провала ППР можно отслеживать путем непрерывного сканирования углов. По сравнению с режимом опроса по интенсивности, в режиме углового опроса можно устранить шум источника света, что приводит к относительно более высокой чувствительности [7]. При опросе по длине волны пика поглощения ППР угол отражения остается фиксированным, а сканирование происходит путем изменения длины волны излучения.

Существуют различные конструкции ВОД, основанные на эффекте ППР [8]. Один из вариантов — ЧЭ на основе гетеросердцевинной структуры, созданный путем сращивания оптических волокон разных типов. Такие датчики отличаются простотой изготовления и высокой механической прочностью. Наиболее распространенная структура — последовательная сварка MMF-SMF-MMF (Multi-Mode fiber, MMF, Single-Mode fiber, SMF), которая используется в различных сенсорах для измерения показателя преломления [9], уровня жидкости [10], температуры [11], акустических волн [12] и т. д.

Одним из наиболее перспективных вариантов повышения чувствительности ВОД на основе ППР является усовершенствование конструкции ЧЭ с помощью внедрения многослойного покрытия с чередующимся диэлектриком. В основе данной конструкции лежит взаимодействие поверхностных плазмон поляритонов, которые образуются в металле, с волноводными модами и распространяются в дополнительном диэлектрическом покрытии. Таким образом, образуется «зеркальное» покрытие, которое усиливает ППР и значительно сужает результирующий спектр.

В качестве «зеркального» покрытия может выступать один слой диэлектрика, либо набор из нескольких слоев с большой разницей в показателе преломления каждого из материалов отдельного слоя. Это могут быть оксид магния (MgO), оксид цинка (ZnO), оксид алюминия (Al₂O₃), диоксид титана (TiO₂) и диоксид кремния (SiO₂) [13]. От толщины и количества диэлектрических слоев будет зависеть получаемый спектр ППР, поэтому данные параметры предварительно рассчитываются в зависимости от необходимого результата для конкретных применений.

С другой стороны, такая конструкция из диэлектрических слоев может не только уменьшить ширину резонансной кривой, но и сдвинуть ее в другую область спектра, например в инфракрасную [14]. Благодаря этому можно использовать перестраиваемый лазер с вертикально излучающим резонатором (Vertical-cavity surface-emitting laser, VCSEL) для опроса с помощью изменения длины волны. Использование VCSEL позволит повысить чувствительность датчика, так как обладает достаточно узким спектром излучения. Однако рабочая длина волны излучения VCSEL лежит в инфракрасной области спектра, в то время как спектры ППРметаллов находятся в видимом диапазоне. Для решения данного несоответствия рабочих диапазонов длин волн предлагается сдвинуть спектр ППР в инфракрасную область путем нанесения на оптическое волокно дополнительного диэлектрического покрытия [15, 16].

Математическое моделирование конструкции чувствительного элемента

Для расчета параметров диэлектрического покрытия было выполнено математическое моделирование с помощью программного обеспечения Mathcad на основе характеристических матриц [17] и классической схемы Кретчманна. На рис. 1 приведена схема моделируемой части ЧЭ, коэффициент отражения которой рассчитывался. Полная схема ЧЭ была представлена и подробно описана в работе [18].





Слой под номером 0 обозначает оптическое волокно, которым была заменена призма из схемы Кретчманна, 1 — слой хрома (Cr), который наносится на оптическое волокно для лучшего нанесения металла, 2 — слой меди (Cu) толщиной 50 нм, 3 — слой диэлектрика, толщину и состав которого необходимо рассчитать, и 4 — исследуемая среда с известным показателем преломления, например воздух или вода. В качестве оптического волокна использовалось стандартное телекоммуникационное одномодовое волокно Corning SMF-28® Ultra с диаметром сердцевины 9 мкм и оболочки 125 мкм, со ступенчатым профилем показателя преломления и апертурой 0,14.

Для составления характеристической матрицы **М**_i для каждого слоя с номером *i* рассчитаем с помощью уравнений Селлмейера [19]:

показатель преломления слоя n_i

$$n_{i}(\lambda) = \sqrt{1 + \frac{a_{1}\lambda^{2}}{\lambda^{2} - b_{1}} + \frac{a_{2}\lambda^{2}}{\lambda^{2} - b_{2}} + \frac{a_{3}\lambda^{2}}{\lambda^{2} - b_{3}}}, \qquad (1)$$

— комплексную диэлектрическую проницаемость ε_i

$$\varepsilon_i(\lambda) = n_i^2(\lambda), \tag{2}$$

— коэффициент поглощения слоя q_i

$$q_i(\lambda, \theta) = \frac{\sqrt{\varepsilon_i(\lambda) - \varepsilon_{i-1}(\lambda) \sin^2 \theta}}{\varepsilon_i(\lambda)},$$
(3)

— коэффициент нелинейного поглощения слоя β_i

$$\beta_i(\lambda, \theta) = \frac{2\pi d_i}{\lambda} \sqrt{\varepsilon_i(\lambda) - \varepsilon_{i-1}(\lambda) \sin^2 \theta}.$$
 (4)

Выражения (1)-(4) подставим в уравнение:

$$\mathbf{M}_{\mathbf{i}}(\lambda, \theta) = \begin{vmatrix} \cos\beta_i & \frac{-i}{q_i} \sin\beta_i \\ -iq_i \sin\beta_i & \cos\beta_i \end{vmatrix},$$
(5)

где λ — длина волны излучения; $a_{1,2,3}$ и $b_{1,2,3}$ — значения коэффициентов Селлмейера; θ — угол падения излучения; d_i — толщина слоя.

Ввиду того, что модель является упрощенной и направлена на прогнозирование диапазона длин волн ППР спектра, для расчета показателя преломления оптического волокна были взяты наборы коэффициентов Селлмейера для чистого кварца. Показатель преломления оболочки волокна отличается от сердцевины в третьем знаке после запятой, поэтому этой разницей в рамках данной модели можно пренебречь.

Далее была составлена характеристическая матрица **М** всей структуры ЧЭ и рассчитан ее коэффициент отражения *r*:

$$\mathbf{I}(\lambda, \theta) = \mathbf{M}_1(\lambda, \theta) \mathbf{M}_2(\lambda, \theta) \mathbf{M}_3(\lambda, \theta), \tag{6}$$

$$\mathbf{f}(\lambda,\theta) = \begin{vmatrix} (\mathbf{M}(\lambda,\theta)_{0,0} + \mathbf{M}(\lambda,\theta)_{0,1}q_4(\lambda,\theta))q_0(\lambda,\theta) - \\ - \frac{(\mathbf{M}(\lambda,\theta)_{1,0} + \mathbf{M}(\lambda,\theta)_{1,1}q_4(\lambda,\theta))}{(\mathbf{M}(\lambda,\theta)_{0,0} + \mathbf{M}(\lambda,\theta)_{0,1}q_4(\lambda,\theta))q_0(\lambda,\theta) + \\ + (\mathbf{M}(\lambda,\theta)_{1,0} + \mathbf{M}(\lambda,\theta)_{1,1}q_4(\lambda,\theta)) \end{vmatrix}, (7)$$

где q_0 и q_4 — коэффициенты отражения слоев 0 и 4.

Научно-технический вестник информационных технологий, механики и оптики, 2025, том 25, № 2 Scientific and Technical Journal of Information Technologies, Mechanics and Optics, 2025, vol. 25, no 2 Используя приближения геометрической оптики, рассчитаем интенсивность отраженного излучения *R*

$$R(\lambda) = \frac{\int_{78}^{89} r(\lambda, \theta) d\theta}{\int_{78}^{89} 1 d\theta},$$
(8)

где нижний предел был получен на основе числовой апертуры многомодового волокна (0,22), из которого излучение попадает в одномодовую часть, а верхний предел соответствует прямо проходящему лучу.

Моделирование проводилось для двух конструкций ЧЭ с различным составом слоя 3: Al_2O_3 толщиной 0–100 нм и TiO₂ толщиной 0–80 нм. В качестве исследуемой среды 4 использовалась вода.

Теоретически полученные спектры ППР (рис. 2) показали меньшую ширину на полувысоте и более резкий пик поглощения по сравнению с модельными спектрами без диэлектрического покрытия [18].



Рис. 2. Модельные спектры поверхностного плазмонного резонанса для разных толщин диэлектрического покрытия из Al_2O_3 (0–100 нм) (*a*) и TiO_2 (0–80 нм) (*b*) для воды *Fig. 2.* Model SPR spectra for different thicknesses of dielectric coating made of Al_2O_3 (0–100 nm) (*a*) and TiO_2 (0–80 nm) (*b*) for water



Puc. 3. Зависимость длины волны от толщины слоя $\rm Al_2O_3$ и $\rm TiO_2$

Fig. 3. Dependence of the wavelength vs. the thickness of the Al_2O_3 and TiO_2 layer

При увеличении толщины покрытия ППР сдвигается в длинноволновую область спектра (рис. 3). Для сдвига спектра до 1550 нм, что соответствует рабочей длине волны VCSEL, необходима толщина слоя покрытия $Al_2O_3 - 100$ нм, а TiO₂ - 60 нм.

Экспериментальная апробация результатов моделирования

Для апробации результатов моделирования были изготовлены четыре образца ЧЭ с различным составом и толщиной диэлектрического покрытия. Полная конструкция ЧЭ представляет собой структуру MMF-SMF-MMF, которая создается путем последовательной сварки одномодового участка (SMF) волокна длиной 15-17 мм между длинными участками многомодового волокна (MMF) [18]. Сварка производилась с помощью сварочного аппарата Fujikura FSM 100P. Сверху одномодового участка с помощью вакуумного напыления в установке Kurt J. Lesker PVD 75 были последовательно нанесены слои Cr толщиной 2 нм и Cu — 50 нм. Далее на первый образец нанесено покрытие из Al₂O₃ толщиной 60 нм, на второй — Al₂O₃ (100 нм), на третий -TiO₂ (50 нм) и на четвертый — TiO₂ (100 нм). Чистота напыляемых материалов была 99,99 %. Напыление производилось со скоростью до 0,1 нм/с при давлении остаточных паров не более 5.10-6 Торр. Во время напыления образец вращался со скоростью 30 об/мин.

Эксперимент проводился следующим образом: ЧЭ через участок MMF волокна подключался к стабилизированному вольфрамово-галогенному источнику света Thorlabs с оптоволоконным выходом и широким диапазоном длин волн 360–2600 нм. Спектры пропускания измерялись с помощью оптического анализатора спектра Yokogawa AQ6370C. Экспериментальные измерения проводились в диапазоне длин волн 700–1700 нм, что соответствует диапазону спектроанализатора. В качестве исследуемой среды использовались: воздух, дистиллированная вода и этиловый спирт. Полученные спектры нормировались на спектр воздуха (рис. 4).



Рис. 4. Спектры поверхностного плазмонного резонанса, нормированные на спектр воздуха, при толщинах диэлектрических покрытий: из Al₂O₃ для 60 нм (*a*) и 100 нм (*b*); из TiO₂ для 50 нм (*c*) и 100 нм (*d*)

Fig. 4. Surface plasmon resonance spectra normalized to the air spectrum for the following dielectric coating thicknesses: Al_2O_3 for 60 nm (*a*) and 100 nm (*b*); TiO₂ for 50 nm (*c*) and 100 nm (*d*)

Экспериментальные результаты качественно отражают теоретические, однако, сдвиг ППР спектров в инфракрасный диапазон происходит на меньшую величину, чем теоретически рассчитано, с разницей в 300–400 нм. Первой причиной такого несоответствия может быть то, что теоретическая модель была рассчитана в достаточно упрощенном виде. Она не учитывает ни температуру, ни изгиб волоконного, ни потери излучения на сварках оптического волокна. Другим важным аспектом является отсутствие возможности точно контролировать толщину напыленных слоев, что также может влиять на результирующие экспериментальные спектры.

Для расчета коэффициента чувствительности *S* была использована следующая формула [20]:

$$S = \frac{\delta \lambda_{\text{IIIP}}}{\delta n_s},\tag{9}$$

где $\delta\lambda_{\Pi\Pi P}$ — сдвиг длины волны ППР; δn_s — соответствующее изменение показателя преломления исследуемой среды. При анализе характеристик различных датчиков необходимо учитывать не только чувствительность, но также точность измерения длины волны.

В данном контексте удобно использовать показатель качества (Figure OF Merit, FOM) устройства, который определяется отношением коэффициента чувствительности к сдвигу длины волны и ширине линии ППР:

$$FOM = \frac{s}{FWHM},$$
 (10)

где *FWHM* (Full Width at Half Maximum) — ширина на полувысоте. Учитывая, что определение местоположения узкого ППР проще, чем широкого, FOM позволяет более объективно оценить эффективность датчика [20].

В таблице приведены рассчитанные характеристики для всех экспериментальных образцов ЧЭ. Несмотря на меньший сдвиг спектров ППР в длинноволновую область, в отличие от теоретически рассчитанных значений, чувствительность датчика для всех образцов значительно повысилась по сравнению с ЧЭ без дополнительного диэлектрического покрытия с чувствитель-

Материал диэлектрического слоя	Толщина слоя, нм	Коэффициент чувствительности S, нм/RIU	FWHM	Показатель качества FOM
Al ₂ O ₃	60	2800	185	около 15
	100	3400	207	около 16
TiO ₂	50	4400	117	около 37
	100	около 5600	около 303	около 18

Таблица. Характеристики экспериментальных образцов чувствительного элемента *Table.* Characteristics of experimental designs of the sensing element

ностью 1800 нм/RIU [18]. Сравнение показателя качества ЧЭ показал, что диэлектрическое покрытие из TiO₂ толщиной 50 нм является наилучшим, так как спектр плазмонного резонанса значительно уже, чем при других конструкциях. Однако большая чувствительность наблюдается у ЧЭ с толщиной TiO₂ 100 нм, поэтому необходимы дальнейшие исследования оптимальной толщины диэлектрического покрытия.

Заключение

В работе представлен метод повышения чувствительности волоконно-оптического датчика на основе поверхностного плазмонного резонанса с использованием диэлектрических покрытий. Предложена мате-

Литература

- Yadav P.K., Srivastava R., Chaurasiya N. Optical fiber sensor: review and applications // International Journal of Creative Research Thoughts. 2018. V. 6. N 1. P. 1358–1362.
- Simonsen O., Grubb A., Thysell H. The blood serum concentration of cystatin C (γ-trace) as a measure of the glomerular filtration rate // Scandinavian Journal of Clinical and Laboratory Investigation. 1985. V. 45. N 2. P. 97–101. https://doi.org/10.3109/00365518509160980
- Алсовэйди К.М., Караваева О.А., Гулий О.И. Методы и подходы для определения антибиотиков // Антибиотики и химиотерапия. 2022. Т. 67. № 1-2. С. 53–61. https://doi.org/10.37489/0235-2990-2022-67-1-2-53-61
- Deng S., Wang P., Yu X. Phase-sensitive Surface Plasmon Resonance sensors: recent progress and future prospects // Sensors. 2017. V. 17. N 12. P. 2819. https://doi.org/10.3390/s17122819
- Zeng Y., Wang X., Zhou J., Miyan R., Qu J., Ho H.P., Zhou K., Gao B., Shao Y. Phase interrogation SPR sensing based on white light polarized interference for wide dynamic detection range // Optics Express. 2020. V. 28. N 3. P. 3442–3450. https://doi.org/10.1364/ OE.382242
- Zeng Y., Hu R., Wang L., Gu D., He J., Wu S., Ho H.P., Li X., Qu J., Gao B.Z., Shao Y. Recent advances in surface plasmon resonance imaging: detection speed, sensitivity, and portability // Nanophotonics. 2017. V. 6. N 5. P. 1017–1030. https://doi.org/10.1515/ nanoph-2017-0022
- Huang Y.H., Ho H.P., Wu S.Y., Kong S.K. Detecting phase shifts in Surface Plasmon Resonance: a review // Advances in Optical Technologies. 2012. V. 2012. P. 471957. https://doi. org/10.1155/2012/471957
- Caucheteur C., Guo T., Albert J. Review of plasmonic fiber optic biochemical sensors: improving the limit of detection // Analytical and Bioanalytical Chemistry. 2015. V. 407. N 14. P. 3883–3897. https://doi.org/10.1007/s00216-014-8411-6
- Xu F., Chen D., Peng B., Xu J., Wu G. All-fiber refractometer based on core mismatch structure // Laser Physics. 2012. V. 22. N 10. P. 1577–1580. https://doi.org/10.1134/s1054660x12100271
- Sun T., Liu Z., Liu Y., Zhang Y., Jing Z., Peng W. All-fiber liquid-level sensor based on in-line MSM fiber structure // Photonic Sensors. 2021. V. 11. N 3. P. 291–297. https://doi.org/10.1007/s13320-020-0586-1

матическая модель на основе характеристических матриц для расчета толщины диэлектрического покрытия. Экспериментальная апробация продемонстрировала, что предложенный метод может быть успешно применен для повышения чувствительности волоконно-оптического датчика на основе поверхностного плазмонного резонанса, однако указала на недостатки математического моделирования.

В дальнейших исследованиях при математическом моделировании необходимо учесть такие важные факторы, как влияние температуры, изгиба волоконного чувствительного элемента, потери на сварках различных типов оптического волокна. Также следует разработать метод контроля толщины напыляемых диэлектрических покрытий.

References

- 1. Yadav P.K., Srivastava R., Chaurasiya N. Optical fiber sensor: review and applications. *International Journal of Creative Research Thoughts*, 2018, vol. 6, no. 1, pp. 1358–1362.
- Simonsen O., Grubb A., Thysell H. The blood serum concentration of cystatin C (γ-trace) as a measure of the glomerular filtration rate. *Scandinavian Journal of Clinical and Laboratory Investigation*, 1985, vol. 45, no. 2, pp. 97–101. https://doi.org/10.3109/00365518509160980
- Alsowaidi A.K.M., Karavaeva O.A., Guliy O.I. Methods and approaches for antibiotics detection. *Antibiotiki i Khimioterapiya*, 2022, vol. 67, no. 1-2, pp. 53–61. (in Russian). https://doi. org/10.37489/0235-2990-2022-67-1-2-53-61
- Deng S., Wang P., Yu X. Phase-sensitive Surface Plasmon Resonance sensors: recent progress and future prospects. *Sensors*, 2017, vol. 17, no. 12, pp. 2819. https://doi.org/10.3390/s17122819
- Zeng Y., Wang X., Zhou J., Miyan R., Qu J., Ho H.P., Zhou K., Gao B., Shao Y. Phase interrogation SPR sensing based on white light polarized interference for wide dynamic detection range. *Optics Express*, 2020, vol. 28, no. 3, pp. 3442–3450. https://doi.org/10.1364/ OE.382242
- Zeng Y., Hu R., Wang L., Gu D., He J., Wu S., Ho H.P., Li X., Qu J., Gao B.Z., Shao Y. Recent advances in surface plasmon resonance imaging: detection speed, sensitivity, and portability. *Nanophotonics*, 2017, vol. 6, no. 5, pp. 1017–1030. https://doi.org/10.1515/ nanoph-2017-0022
- Huang Y.H., Ho H.P., Wu S.Y., Kong S.K. Detecting phase shifts in Surface Plasmon Resonance: a review. Advances in Optical Technologies, 2012, vol. 2012, pp. 471957. https://doi. org/10.1155/2012/471957
- Caucheteur C., Guo T., Albert J. Review of plasmonic fiber optic biochemical sensors: improving the limit of detection. *Analytical and Bioanalytical Chemistry*, 2015, vol. 407, no. 14, pp. 3883–3897. https://doi.org/10.1007/s00216-014-8411-6
- Xu F., Chen D., Peng B., Xu J., Wu G. All-fiber refractometer based on core mismatch structure. *Laser Physics*, 2012, vol. 22, no. 10, pp. 1577–1580. https://doi.org/10.1134/s1054660x12100271
- Sun T., Liu Z., Liu Y., Zhang Y., Jing Z., Peng W. All-fiber liquid-level sensor based on in-line MSM fiber structure. *Photonic Sensors*, 2021, vol. 11, no. 3, pp. 291–297. https://doi.org/10.1007/s13320-020-0586-1

- Nguyen L.V., Hwang D., Moon S., Moon D.S., Chung Y. High temperature fiber sensor with high sensitivity based on core diameter mismatch // Optics Express. 2008. V. 16. N 15. P. 11369–11375. https://doi.org/10.1364/OE.16.011369
- Sun A., Wu Z., Wan C., Yang C. All-fiber optic acoustic sensor based on multimode-single mode-multimode structure // Optik. 2012. V. 123. N 13. P. 1138–1139. https://doi.org/10.1016/j. ijleo.2011.07.040
- Sharma A.K., Jha R., Gupta B.D. Fiber-optic sensors based on Surface Plasmon Resonance: a comprehensive review // IEEE Sensors Journal. 2007. V. 7. N 8. P. 1118–1129. https://doi.org/10.1109/ jsen.2007.897946
- Bao S., Li H., Zheng G. Concentration sensor with multilayer thin film-coupled surface plasmon resonance // Optoelectronics Letters. 2021. V. 17. N 5. P. 289–293. https://doi.org/10.1007/s11801-021-0088-4
- Ma Y., Liu F., Ren Q., Zhang A. FOM enhancement of a D-shaped SPR fiber sensor based on Al₂O₃-graphene-platinum grating // Journal of Optics. 2024. V. 53. N 1. P. 197–205. https://doi. org/10.1007/s12596-022-01076-8
- Takagi K., Sasaki H., Seki A., Watanabe K. Surface plasmon resonances of a curved hetero-core optical fiber sensor // Sensors and Actuators A: Physical. 2010. V. 161. N 1-2. P. 1–5. https://doi. org/10.1016/j.sna.2010.03.009
- Виноградов С.В., Кононов М.А. Расчет параметров многослойной структуры при резонансном возбуждении поверхностных плазмонов // Успехи прикладной физики. 2016. Т. 4. № 4. С. 343– 348.
- 18. Ивойлов К.А., Гагаринова Д.О., Зыкина А.А., Мешковский И.К., Плясцов С.А. Модель чувствительного элемента рефрактивного волоконно-оптического сенсора на основе MMF-SMF-MMF структуры с использованием поверхностного плазмонного резонанса // Научно-технический вестник информационных технологий, механики и оптики. 2023. Т. 23, № 3. С. 448–454. https:// doi.org/10.17586/2226-1494-2023-23-3-448-454
- Malitson I.H. Interspecimen comparison of the refractive index of fused silica // Journal of the Optical Society of America. 1965. V. 55. N 10. P. 1205–1209. https://doi.org/10.1364/JOSA.55.001205
- Vikas, Verma R.K. High figure of merit fiber optic surface plasmon resonance sensor with topological insulator (BSTS) // Optical and Quantum Electronics. 2022. V. 54. N 1. P. 55. https://doi.org/10.1007/ s11082-021-03435-4

Авторы

Зыкина Аделия Алексеевна — инженер, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, вс 58527719900, https://orcid.org/0000-0002-8021-0678, adeliya.zykina@gmail.com Плясцов Семён Алексеевич — кандидат технических наук, заведующий лабораторией, ассистент, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, вс 57195587476, https://

orcid.org/0000-0002-5764-0960, s.plyastsov@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, 138020@mail.ru

Статья поступила в редакцию 05.02.2024 Одобрена после рецензирования 13.02.2025 Принята к печати 21.03.2025



- Nguyen L.V., Hwang D., Moon S., Moon D.S., Chung Y. High temperature fiber sensor with high sensitivity based on core diameter mismatch. *Optics Express*, 2008, vol. 16, no. 15, pp. 11369–11375. https://doi.org/10.1364/OE.16.011369
- Sun A., Wu Z., Wan C., Yang C. All-fiber optic acoustic sensor based on multimode-single mode-multimode structure. *Optik*, 2012, vol. 123, no. 13, pp. 1138–1139. https://doi.org/10.1016/j. ijleo.2011.07.040
- Sharma A.K., Jha R., Gupta B.D. Fiber-optic sensors based on Surface Plasmon Resonance: a comprehensive review. *IEEE Sensors Journal*, 2007, vol. 7, no. 8, pp. 1118–1129. https://doi.org/10.1109/ jsen.2007.897946
- Bao S., Li H., Zheng G. Concentration sensor with multilayer thin film-coupled surface plasmon resonance. *Optoelectronics Letters*. 2021. vol. 17, no. 5, pp. 289–293. https://doi.org/10.1007/s11801-021-0088-4
- Ma Y., Liu F., Ren Q., Zhang A. FOM enhancement of a D-shaped SPR fiber sensor based on Al₂O₃–graphene–platinum grating. *Journal* of Optics, 2024, vol. 53, no. 1, pp. 197–205. https://doi.org/10.1007/ s12596-022-01076-8
- Takagi K., Sasaki H., Seki A., Watanabe K. Surface plasmon resonances of a curved hetero-core optical fiber sensor. *Sensors and Actuators A: Physical*, 2010, vol. 161, no. 1-2, pp. 1–5. https://doi. org/10.1016/j.sna.2010.03.009
- Vinogradov S.V., Kononov M.A. Calculation of parameters of a multilayered structure at the surface plasmon resonance. *Uspekhi Prikladnoi Fiziki*, 2016, vol. 4, no. 4, pp. 343–348. (in Russian)
- Ivoilov K.A., Gagarinova D.O., Zykina A.A., Meshkovskiy I.K., Plyastsov S.A. A model of a refractive fiber optic sensor sensing element based on MMF-SMF-MMF structure using surface plasmon resonance. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 2023, vol. 23, no. 3, pp. 448– 454. (in Russian). https://doi.org/10.17586/2226-1494-2023-23-448-454
- Malitson I.H. Interspecimen comparison of the refractive index of fused silica. *Journal of the Optical Society of America*, 1965, vol. 55, no. 10, pp. 1205–1209. https://doi.org/10.1364/JOSA.55.001205
- Vikas, Verma R.K. High figure of merit fiber optic surface plasmon resonance sensor with topological insulator (BSTS). *Optical and Quantum Electronics*, 2022, vol. 54, no. 1, pp. 55. https://doi. org/10.1007/s11082-021-03435-4

Authors

Adeliia A. Zykina — Engineer, ITMO University, Saint Petersburg, 197101, Russian Federation, sc 58527719900, https://orcid.org/0000-0002-8021-0678, adeliya.zykina@gmail.com

Semyon A. Plyastsov — PhD, Head of Laboratory, Assistant, ITMO University, Saint Petersburg, 197101, Russian Federation, SC 57195587476, https://orcid.org/0000-0002-5764-0960, s.plyastsov@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, 138020@mail. ru

Received 05.02.2024 Approved after reviewing 13.02.2025 Accepted 21.03.2025

Работа доступна по лицензии Creative Commons «Attribution-NonCommercial» **VİTMO**

НАУЧНО-ТЕХНИЧЕСКИЙ ВЕСТНИК ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ, МЕХАНИКИ И ОПТИКИ март–апрель 2025 Том 25 № 2 http://ntv.ifmo.ru/ SCIENTIFIC AND TECHNICAL JOURNAL OF INFORMATION TECHNOLOGIES, MECHANICS AND OPTICS March–April 2025 Vol. 25 No 2 http://ntv.ifmo.ru/en/ ISSN 2226-1494 (print) ISSN 2500-0373 (online)

ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ, МЕХАНИКИ И ОПТИКІ

НОВЫЕ МАТЕРИАЛЫ И НАНОТЕХНОЛОГИИ MATERIAL SCIENCE AND NANOTECHNOLOGIES

doi: 10.17586/2226-1494-2025-25-2-229-235 УДК 532.6:546.22

Исследование термодинамических характеристик наночастиц золота в процессах плавления и кристаллизации

с использованием метода молекулярной динамики Дамдин Галсанович Цыдыпов¹, Андрей Валерьевич Номоев², Баир Заятуевич Гармаев³

1,2,3 Институт физического материаловедения СО РАН, Улан-Удэ, 670047, Российская Федерация

¹ damdinkatsydypov@gmail.com[⊠], https://orcid.org/0009-0005-5571-5075

² nomoevav@mail.ru, https://orcid.org/0009-0000-6201-6743

³ bair.garmaev@gmail.com, https://orcid.org/0000-0001-6086-3658

Аннотация

Введение. Представлены результаты компьютерного моделирования процессов плавления и кристаллизации наночастиц золота сферической формы. Методом молекулярной динамики выполнен анализ термодинамических характеристик (температуры, теплоты, энтропии плавления и кристаллизации) наночастиц для различных скоростей нагрева и охлаждения. Такие исследование позволяют выбирать наиболее подходящие температурные диапазоны формирования нанокристаллических структур и предсказывать их размеры. Уменьшение размера наночастиц золота (менее 100 нм) приводит к значительному увеличению соотношения площади поверхности и объема частиц, в результате чего физические и химические характеристики материала существенно изменяются по сравнению с характеристиками этого же материала в объемной форме. Интерес к наночастицам золота обусловлен усиленной фотоэмиссией, высокими электро- и теплопроводностью, повышенной каталитической активностью поверхности. Наночастицы золота обладают сильным оптическим поглощением и рассеивающими свойствами в видимой области спектра вследствие поверхностных плазмонных колебаний свободных электронов. В известных из литературы исследованиях обнаружено, что с увеличением размера наночастиц гистерезис между температурой плавления и кристаллизации увеличивается, в то время как в теории макроскопические температуры плавления и кристаллизации должны быть одинаковыми. Новизна представленного в данной работе исследования состоит в выявлении ранее не наблюдаемой тенденции к сближению макроскопических температур, теплоты и энтропии плавления и кристаллизации при уменьшении скоростей нагрева и охлаждения. Метод. Для изучения термодинамических свойств наночастиц золота применен метод классической молекулярной динамики. Предметом моделирования выбраны наночастицы золота различных размеров сферической формы с гранецентрированной кубической решеткой. В процессе моделирования применялся межатомный потенциал, соответствующий методу «погруженного атома», который был разработан для золота с использованием улучшенной методологии «согласования сил». Моделировались нагрев и охлаждение наночастии при скоростях изменения температуры 0,1 ТК/с, 1 ТК/с, 3 ТК/с. Основные результаты. Путем анализа взаимосвязи потенциальной энергии наночастиц золота и температуры выявлена зависимость температуры плавления и температуры кристаллизации наночастиц от их размера. Установлена связь между размером наночастиц, теплотой, энтропией плавления и кристаллизации при различных скоростях нагрева и охлаждения. Показано, что при уменьшении скоростей нагрева и охлаждения от 3 ТК/с до 0,1 ТК/с происходит «сближение» макроскопических значений температур плавления и кристаллизации (уменьшение разницы от 467 К до 158 К), макроскопических значений теплоты плавления и кристаллизации (от 4,24 кДж/моль до 0,67 кДж/моль), энтропии плавления и кристаллизации (от 1,99 Дж/(моль К) до 0,16 Дж/(моль К)). Сделано предположение, что это связано с уменьшением доли образующихся наноструктур, отличных от кубических гранецентрированных. Обсуждение. Предсказание температурных режимов плавления и кристаллизации наночастиц золота позволяет управлять фазовыми переходами при изготовлении нанокристаллов с заданными свойствами. Это явление может найти применение в микроэлектронике для формирования тонких пленок с высокой степенью однородности, в катализе для формирования наночастиц с нужными структурами и свойствами.

Ключевые слова

наночастица золота, температуры плавления и кристаллизации, теплота плавления и кристаллизации, энтропия плавления и кристаллизации, метод молекулярной динамики

© Цыдыпов Д.Г., Номоев А.В., Гармаев Б.З., 2025

Благодарности

Работа выполнена Институтом физического материаловедения СО РАН по проекту государственного задания (FWSF-2024-0013) «Разработка физических основ создания функциональных композитных наноструктур, материалов и построение моделей фазовых диаграмм многокомпонентных систем».

Ссылка для цитирования: Цыдыпов Д.Г., Номоев А.В., Гармаев Б.З. Исследование термодинамических характеристик наночастиц золота в процессах плавления и кристаллизации с использованием метода молекулярной динамики // Научно-технический вестник информационных технологий, механики и оптики. 2025. Т. 25, № 2. С. 229–235. doi: 10.17586/2226-1494-2025-25-2-229-235

Study of the thermodynamic characteristics of gold nanoparticles in the processes of melting and crystallization by the molecular dynamics method Damdin G. Tsydypov¹²⁷, Andrey V. Nomoev², Bair Z. Garmaev³

1.2.3 Institute of Physical Material Science of the Siberian Branch of the RAS, Ulan-Ude, 670047, Russian Federation

¹ damdinkatsydypov@gmail.com^{\Box}, https://orcid.org/0009-0005-5571-5075

² nomoevav@mail.ru, https://orcid.org/0009-0000-6201-6743

³ bair.garmaev@gmail.com, https://orcid.org/0000-0001-6086-3658

Abstract

The article presents the results of computer modeling of melting and crystallization processes of spherical gold nanoparticles. The molecular dynamics method was used to analyze the thermodynamic characteristics (temperature, heat, entropy of melting, and crystallization) of nanoparticles for different heating and cooling rates. Such studies allow choosing the most suitable temperature ranges for the formation of nanocrystalline structures and predicting their sizes. Reducing the size of gold nanoparticles (less than 100 nm) leads to a significant increase in the ratio of the surface area to the volume of particles, as a result of which the physical and chemical characteristics of the material change significantly compared to materials of the same material in bulk form. Interest in gold nanoparticles is due to enhanced photoemission, high electrical and thermal conductivity, and increased catalytic activity of the surface. Gold nanoparticles have strong optical absorption and scattering properties in the visible region of the spectrum due to surface plasmon oscillations of free electrons. In the known studies from the literature, it was found that with an increase in the size of nanoparticles, the hysteresis between the melting and crystallization temperatures increases, while in theory the macroscopic melting and crystallization temperatures should be the same. The novelty of the study presented in this paper is to reveal a previously unobserved tendency for macroscopic temperatures, heat and entropy of melting and crystallization to converge with a decrease in the heating and cooling rates. The classical molecular dynamics method was used to study the thermodynamic properties of gold nanoparticles. The subject of modeling was gold nanoparticles of various sizes, spherical in shape with a face-centered cubic lattice. In the modeling process, the interatomic potential was used corresponding to the embedded atom method which was developed for gold, using an improved force matching methodology. Heating and cooling of nanoparticles were modeled at temperature change rates of 0.1 TK/s, 1 TK/s, 3 TK/s. By analyzing the relationship between the potential energy of gold nanoparticles and temperature, the dependences of the melting and crystallization temperatures of nanoparticles on their size were revealed. A relationship was established between the size of nanoparticles, heat, entropy of melting and crystallization at different heating and cooling rates. It was shown that with a decrease in the heating and cooling rates from 3 TK/s to 0.1 TK/s, there is a convergence of the macroscopic values of the melting and crystallization temperatures (a decrease in the difference from 467 K to 158 K), macroscopic values of the heat of melting and crystallization (from 4.24 kJ/mol to 0.67 kJ/mol), entropy of melting and crystallization (from $1.99 \text{ J/(mol \cdot K)}$) to $0.16 \text{ J/(mol \cdot K)}$). It was assumed that this is due to a decrease in the proportion of the formed nanostructures other than face-centered cubic ones. Prediction of temperature regimes of melting and crystallization of gold nanoparticles allows controlling phase transitions in the production of nanocrystals with specified properties. This phenomenon can find application in microelectronics for the formation of thin films with a high degree of homogeneity and in catalysis for the formation of nanoparticles with the required structures and properties. The obtained data can be used to verify the results of real experiments on phase transitions and to adjust molecular dynamics models.

Keywords

gold nanoparticle, melting and crystallization temperatures, heats of melting and crystallization, entropies of melting and crystallization, molecular dynamics method

Acknowledgements

The presented study is performed within the state task of Institute of Physical Materials Science of the Siberian Branch of the Russian Academy of Sciences (FWSF-2024-0013) "Development of physical foundations for the creation of functional composite nanostructures, materials and the construction of models of phase diagrams of multicomponent systems".

For citation: Tsydypov D.G., Nomoev A.V., Garmaev B.Z. Study of the thermodynamic characteristics of gold nanoparticles in the processes of melting and crystallization by the molecular dynamics method. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 2025, vol. 25, no. 2, pp. 229–235 (in Russian). doi: 10.17586/2226-1494-2025-25-2-229-235

Введение

Компьютерное моделирование активно применяется для анализа физико-химических свойств металлических наночастиц, так как современные вычислительные мощности компьютеров позволяют проводить подробное их изучение. Использование методов компьютерного моделирования дает возможность прогнозировать различные свойства наночастиц без проведения реальных экспериментов, что является важным при решении прикладных технологических задач.

В работах [1-4] было проведено изучение зависимостей температур плавления и кристаллизации наночастиц от размера. Согласно описанной теории, явление гистерезиса плавления-кристаллизации наночастиц было предсказано В.П. Скриповым и В.П. Ковердой [1]. Как указано в [1], с возрастанием размера частиц увеличивается ширина гистерезиса плавления-кристаллизации. В работе [2] исследовались нанокластеры золота с использованием потенциала Гупта [5]. Также отметим работу [3], которая проведена для кластеров меди и никеля с использованием потенциала сильной связи [6]. В работах [2, 3] наблюдалось увеличение ширины гистерезиса плавления-кристаллизации при увеличении размера нанокластеров в отличие от теории, по которой макроскопические температуры плавления и кристаллизации должны совпадать. В настоящей работе исследовались зависимости температур плавления и кристаллизации наночастиц золота сферической формы от размера при разных скоростях нагрева и охлаждения. Также при разных скоростях нагрева и охлаждения изучались размерные зависимости теплоты плавления и кристаллизации, а также энтропии соответствующих фазовых переходов. Для этого было выполнено компьютерное моделирование с использованием метода молекулярной динамики (МД).

В работе [4] высказано предположение, что максимальные температуры плавления и кристаллизации будут достигнуты при значительно больших размерах частиц, чем те, которые можно исследовать в лабораторных условиях или с помощью компьютерного моделирования. Также в [4] допущено, что при достижении определенного размера частицы размерные зависимости температур плавления и кристаллизации сойдутся скачкообразно.

Отметим, что различие между макроскопическими температурами плавления и кристаллизации является следствием высоких значений скоростей нагрева и охлаждения, при которых в наночастице в процессах плавления и кристаллизации образуются кристаллические структуры (гексагональные плотноупакованные, икосаэдрические, декаэдрические), отличные от гранецентрированных кубических (ГЦК) [7]. В результате этих процессов исследуемая наночастица имеет отличные от наночастицы с ГЦК структурой температуры плавления и кристаллизации. Можно предположить, что если скорости нагрева и охлаждения будут низкими, то расхождение макроскопических значений будет минимальным.

Новизна настоящей работы состоит в выявлении тенденции к сближению температур, теплоты, энтро-

пии плавления и кристаллизации в «макроскопическом пределе» при уменьшении скоростей нагрева и охлаждения от 3 ТК/с до 0,1 ТК/с.

Реализация моделирования процессов плавления и кристаллизации наночастиц золота

В настоящей работе проведено моделирование процессов плавления и кристаллизации наночастиц золота сферической формы с ГЦК решеткой с помощью МД-метода. В ходе моделирования использовался межатомный потенциал, соответствующий методу «погруженного атома» [8], разработанный для золота с использованием улучшенной методологии «согласования сил».

Для моделирования процессов релаксации, нагрева и охлаждения частицы до необходимой температуры применялся термостат Нозе–Гувера (канонический ансамбль NVT), который считается одним из лучших методов для МД-моделирования. Работа выполнялась при временном шаге, равном 1 фс.

Первоначально проводилась релаксация наночастицы золота при температуре 300 К длительностью в 500 пс. Затем моделировался нагрев наночастицы от температуры 300 К до 1800 К при скоростях нагрева 0,1 ТК/с, 1 ТК/с, 3 ТК/с. После нагрева вновь выполнялась релаксация при температуре 1800 К в течение 500 пс. Далее моделировалось охлаждение от 1800 К до 300 К при скоростях охлаждения 0,1 ТК/с, 1 ТК/с, 3 ТК/с.

Результаты и их обсуждение

В работе [9] изучены зависимости температуры плавления наночастиц серебра от их диаметра с применением различных межатомных потенциалов, а также определены макроскопические температуры плавления для каждого из используемых потенциалов. В [10] для повышения точности определения макроскопических значений температуры плавления проанализированы зависимости температуры плавления наночастиц серебра от обратного радиуса частицы с использованием тех же межатомных потенциалов, что и в [9]. Также в [10] исследованы зависимости теплоты плавления наночастиц серебра от их размера при применении различных потенциалов. В результате в [9, 10] сделано утверждение, что необходимо обращать внимание на выбор потенциала для наночастиц, так как он оказывает значительное влияние на результат моделирования.

Известно, что с уменьшением размера наночастиц их температура плавления также уменьшается. Зависимость температуры плавления T_m наночастицы от обратного радиуса (1/r) является линейной и выражается формулой Томсона [11]:

$$T_m = T_0 \left(1 - \frac{2\sigma_{sl}\upsilon_s}{\lambda_0} \frac{1}{r} \right),$$

где T_0 — макроскопическая температура плавления; λ_0 — макроскопическая удельная теплота плавления;

 σ_{sl} — межфазное натяжение на границе раздела кристалл-расплав; υ_s — удельный объем твердой фазы; r — радиус наночастицы.

В настоящей работе получены температурные зависимости потенциальной энергии для наночастиц золота различных размеров с использованием межатомного потенциала [8]. Температуры плавления и кристаллизации определялись по резкому изменению температурной зависимости потенциальной энергии наночастицы. За температуры плавления и кристаллизации принимались значения, расположенные в середине интервала между началом и завершением фазового перехода.

На рис. 1 представлены зависимости температур плавления $T_m(N^{-1/3})$ (N — количество атомов в частице, $r \sim N^{1/3}$) и кристаллизации $T_c(N^{-1/3})$ наночастицы от размера с использованием потенциала [8] для различных скоростей нагрева и охлаждения.

Путем линейной экстраполяции зависимостей температур плавления и кристаллизации частицы от ее размера к $N^{1/3} \rightarrow \infty$, были получены макроскопические значения температур плавления $T_m^{(\infty)}$ и кристаллизации $T_c^{(\infty)}$. В результате моделирования для скорости изменения температуры 3 ТК/с получены значения $T_m^{(\infty)} = 1295$ К и $T_c^{(\infty)} = 828$ К, для 1 ТК/с — 1 ТК/с $T_m^{(\infty)} = 1227$ К, $T_c^{(\infty)} = 921$ К, для 0,1 ТК/с $-T_m^{(\infty)} = 1276$ К. Таким образом, при снижении скоростей нагрева и охлаждения от 3 ТК/с до 0,1 ТК/с разница между $T_m^{(\infty)}$ и $T_c^{(\infty)}$ сокращается от 467 К до 158 К, т. е. наблюдается тенденция к слиянию значений $T_m^{(\infty)}$ и $T_c^{(\infty)}$.



Рис. 1. Зависимости температур плавления и кристаллизации наночастицы золота от размера для различных скоростей нагрева и охлаждения при использовании потенциала [8]. Температуры плавления и кристаллизации для скоростей нагрева и охлаждения: $T_m^{(1)}$ и $T_c^{(1)}$ для 3 ТК/с; $T_m^{(2)}$ и $T_c^{(2)}$ для 1 ТК/с; $T_m^{(3)}$ и $T_c^{(3)}$ для 0,1 ТК/с.

Штриховая линия — справочное значение макроскопической температуры плавления золота $T_0 = 1336$ К [12]

Fig. 1. Dependences of melting and crystallization temperatures of gold nanoparticles vs. their size for different heating and cooling rates using the potential [8]. Melting and crystallization temperatures for heating and cooling rates: $T_m^{(1)}$ and $T_c^{(1)}$ for 3 TK/s; $T_m^{(2)}$ and $T_c^{(2)}$ for 1 TK/s; $T_m^{(3)}$ and $T_c^{(3)}$ for 0.1 TK/s. Dashed line — reference value of the macroscopic melting temperature of gold $T_0 = 1336$ K [12] На рис. 2 представлены зависимости температур плавления и кристаллизации нанокластеров золота от размера, рассчитанные в работе [2] с использованием потенциала Гупта [5] при скоростях нагрева (10,36 ТК/с) и охлаждения (1,04 ТК/с) нанокластеров.

Рассчитаны значения макроскопических температур плавления частицы $T_m^{(\infty)} = 1568$ К и кристаллизации $T_c^{(\infty)} = 1199$ К. Разница между значениями $T_m^{(\infty)}$ и $T_c^{(\infty)}$ довольно большая (369 К) в отличие от разницы, полученной в настоящей работе для скоростей нагрева и охлаждения 0,1 ТК/с (158 К).

Также важными термодинамическими характеристиками наночастиц являются теплота и энтропия плавления и кристаллизации.

Метод определения теплоты плавления с использованием зависимости потенциальной энергии наночастицы от температуры подробно описан в работах [13–16].

На рис. 3 представлены зависимости теплоты плавления $H_m(N^{-1/3})$ и кристаллизации $H_c(N^{-1/3})$ наночастицы золота от размера при различных скоростях нагрева и охлаждения при применении потенциала [8].

В работах [13–17] установлено, что зависимость теплоты плавления от обратного радиуса (1/*r*) имеет почти линейный характер, что позволяет выполнить линейную экстраполяцию зависимостей теплоты плавления и кристаллизации частицы от размера к $N^{1/3} \rightarrow \infty$ ($r \sim N^{1/3}$) для определения макроскопических значений теплоты плавления $H_m^{(\infty)}$ и кристаллизации $H_c^{(\infty)}$. При скорости нагрева и охлаждения 3 ТК/с получены результаты $H_m^{(\infty)} = 7,96$ кДж/моль и $H_c^{(\infty)} = 3,72$ кДж/моль; для 0,1 ТК/с — $H_m^{(\infty)} = 6,47$ кДж/моль и $H_c^{(\infty)} = 5,8$ кДж/моль. Таким образом, при уменьшении скоростей нагрева и охлаждения от 3 ТК/с до 0,1 ТК/с раз-



Рис. 2. Зависимости температур плавления и кристаллизации нанокластеров золота от размера, рассчитанные в [2].

Штриховая линия — справочное значение макроскопической температуры плавления золота *T*₀ = 1336 К [12]

Fig. 2. Size dependences of melting and crystallization temperatures of gold nanoclusters, calculated in [2]. Dashed line — reference value of the macroscopic melting temperature of gold $T_0 = 1336$ K [12]

Научно-технический вестник информационных технологий, механики и оптики, 2025, том 25, № 2 Scientific and Technical Journal of Information Technologies, Mechanics and Optics, 2025, vol. 25, no 2 ница между $H_m^{(\infty)}$ и $H_c^{(\infty)}$ уменьшается от 4,24 кДж/моль до 0,67 кДж/моль. При скорости нагрева и охлаждения 0,1 ТК/с значения $H_m^{(\infty)}$ и $H_c^{(\infty)}$ практически совпадают.

На рис. 4 представлены графики зависимостей теплоты плавления и кристаллизации нанокластеров золота от размера, на основании расчетов [17] при применении потенциала Гупта [5]. Отметим, что в [17] зависимости были получены для высоких скоростей нагрева (10,36 ТК/с) и охлаждения (1,04 ТК/с), при значениях $H_m^{(\infty)} = 17,16$ кДж/моль, $H_c^{(\infty)} = 13,08$ кДж/моль.

Разница между $H_m^{(\infty)}$ и $H_c^{(\infty)}$ большая (4,08 кДж/моль) в отличие от текущей работы, в которой значения $H_m^{(\infty)}$ и $H_c^{(\infty)}$ почти совпадают при скорости нагрева и охлаждения 0,1 ТК/с.

На рис. 5 представлены рассчитанные с использованием потенциала [8] зависимости энтропии плавления $\Delta S_m = H_m (N^{-1/3})/T_m (N^{-1/3})$ и кристаллизации $\Delta S_c = H_c (N^{-1/3})/T_c (N^{-1/3})$ наночастицы золота от размера для различных скоростей нагрева и охлаждения.

При скорости нагрева и охлаждения 3 ТК/с получены значения $\Delta S_m^{(\infty)} = 6,47 \ \text{Дж/(моль·K)}, \ \Delta S_c^{(\infty)} =$ = 4,48 Дж/(моль·K), при 0,1 ТК/с – $\Delta S_m^{(\infty)} =$ = 5,54 Дж/(моль·К), $\Delta S_c^{(\infty)} = 5,7 \ \text{Дж/(моль·K)}$. Таким образом, при уменьшении скорости от 3 ТК/с до 0,1 ТК/с разница между $\Delta S_m^{(\infty)}$ и $\Delta S_c^{(\infty)}$ уменьшается от 1,99 Дж/(моль·К) до 0,16 Дж/(моль·К). При скорости нагрева и охлаждения 0,1 ТК/с зависимости энтропии плавления и кристаллизации частицы от размера почти сливаются в «макроскопическом пределе».



Рис. 3. Зависимости теплоты плавления и кристаллизации наночастицы золота от размера для различных скоростей нагрева и охлаждения при использовании потенциала [8].
 Теплота плавления и кристаллизации для скоростей нагрева и охлаждения: H_m⁽¹⁾ и H_c⁽¹⁾ для 3 ТК/с; H_m⁽²⁾ и H_c⁽²⁾ для 0,1 ТК/с.
 Штриховая линия — справочное значение макроскопической

птриховая линия — справочное значение макроскопической теплоты плавления золота *H*₀ = 12,6 кДж/моль [12]

Fig. 3. Dependences of the heat of melting and crystallization of gold nanoparticles vs. the size for different heating and cooling rates using the potential [8]. Heats of melting and crystallization for heating and cooling rates: $H_m^{(1)}$ and $H_c^{(1)}$ for 3 TK/s; $H_m^{(2)}$ and $H_c^{(2)}$ for 0.1 TK/s.

Dashed line — reference value of the macroscopic heat of melting of gold $H_0 = 12.6$ kJ/mol [12]



Рис. 4. Зависимости теплоты плавления и кристаллизации нанокластеров золота от размера, рассчитанные в [17].

Штриховая линия — справочное значение макроскопической теплоты плавления золота H₀ = 12,6 кДж/моль [12]

Fig. 4. Dependences of the heat of melting and heat of crystallization of gold nanoclusters vs. size, calculated in [17]. Dashed line — reference value of the macroscopic heat of melting of gold $H_0 = 12.6$ kJ/mol [12]

На рис. 6 показаны рассчитанные в [18] зависимости энтропии плавления и кристаллизации нанокластеров золота от размера. Зависимости были рассчита-



Рис. 5. Зависимости энтропии плавления и кристаллизации наночастицы золота от размера для различных скоростей нагрева и охлаждения при использовании потенциала [8]. Энтропия плавления и кристаллизации для скоростей нагрева и охлаждения: $\Delta S_m^{(1)}$ и $\Delta S_c^{(1)}$ для 3 ТК/с; $\Delta S_m^{(2)}$ и $\Delta S_c^{(2)}$ для 0,1 ТК/с.

Штриховая линия — справочное значение макроскопической энтропии плавления золота $\Delta S_0 = 9,43 \text{ Дж/(моль·K)}$ [12]

Fig. 5. Dependences of the entropies of melting and crystallization of a gold nanoparticle vs. its size for different heating and cooling rates using the potential [8]. Entropies of melting and crystallization for heating and cooling rates: $\Delta S_m^{(1)}$ and $\Delta S_c^{(1)}$ for 3 TK/s; $\Delta S_m^{(2)}$ and $\Delta S_c^{(2)}$ for 0.1 TK/s.

Dashed line — reference value of the macroscopic entropy of melting of gold $\Delta S_0 = 9.43 \text{ J/(mol·K)}$ [12]





Fig. 6. Dependences of the entropies of melting and crystallization of gold nanoclusters vs. size, calculated in [18]. Dashed line — reference value of the macroscopic entropy of melting of gold $\Delta S_0 = 9.43 \text{ J/(mol·K)}$ [12]

ны с помощью полученных в [2] и [17] зависимостей температур плавления и кристаллизации кластеров от размера, теплоты плавления и кристаллизации кластеров от размера. Таким образом, рассчитанные в [18] зависимости также соответствуют высоким скоростям нагрева и охлаждения. Получены значения $\Delta S_m^{(\infty)} = 11,54 \text{ Дж/(моль·K)}, \Delta S_c^{(\infty)} = 8,45 \text{ Дж/(моль·K)}. Разница между <math>\Delta S_m^{(\infty)}$ и $\Delta S_c^{(\infty)}$ большая (3,09 Дж/(моль·K)) в отличие от настоящей работы, в которой наблюдается слияние зависимостей энтропии плавления и кристаллизации частицы от размера в «макроскопическом пределе» при скорости нагрева и охлаждения 0,1 ТК/с.

Литература

- Скрипов В.П., Коверда В.П. Спонтанная кристаллизация переохлажденных жидкостей: Зарождение кристаллов в жидкостях и аморфных твердых телах. М.: Наука, 1984. 232 с.
- Сдобняков Н.Ю., Комаров П.В., Соколов Д.Н., Самсонов В.М. Исследование термодинамических характеристик нанокластеров золота с использованием многочастичного потенциала Гупта // Физика металлов и металловедение. 2011. Т. 111. № 1. С. 15–22.
- Самсонов В.М., Харечкин С.С., Гафнер С.Л., Редель Л.В., Гафнер Ю.Я. Молекулярно-динамическое исследование плавления и кристаллизации наночастиц // Кристаллография. 2009. Т. 54. № 3. С. 563–569.
- Сдобняков Н.Ю., Соколов Д.Н., Базулев А.Н., Самсонов В.М., Зыков Т.Ю., Антонов А.С. О взаимосвязи между размерными зависимостями температур плавления и кристаллизации для металлических наночастиц // Расплавы. 2012. № 5. С. 88–94.
- Gupta R.P. Lattice relaxation at a metal surface // Physical Review B. 1981. V. 23. N 12. P. 6265–6270. https://doi.org/10.1103/ PhysRevB.23.6265
- Cleri F., Rosato V. Tight-binding potentials for transition metals and alloys // Physical Review B. 1993. V. 48. N 1. P. 22–33. https://doi. org/10.1103/PhysRevB.48.22
- Самсонов В.М., Васильев С.А., Талызин И.В., Рыжков Ю.А. О причинах гистерезиса плавления и кристаллизации наночастиц //

Заключение

В работе получены термодинамические характеристики наночастицы золота в процессах плавления и кристаллизации с помощью компьютерного моделирования молекулярно-динамическим методом в случае применения потенциала [8] для различных скоростей нагрева и охлаждения. Установлено, что с увеличением размера наночастиц гистерезис между плавлением и кристаллизацией расширяется несмотря на то, что в теории макроскопические температуры плавления и кристаллизации должны быть одинаковыми. Разницу между макроскопическими температурами кристаллизации и плавления можно объяснить высокими скоростями нагрева и охлаждения, при которых в наночастице в процессах плавления и кристаллизации происходит образование кристаллических структур, отличных от гранецентрированных кубических. По этой причине наночастица обладает отличными от гранецентрированной кубической структуры температурами плавления и кристаллизации. При снижении скоростей нагрева и охлаждения от 3 ТК/с до 0,1 ТК/с разница между $T_m^{(\infty)}$ и $T_c^{(\infty)}$ сократилась от 467 К до 158 К, т. е. наблюдается тенденция к сближению макроскопических значений $T_m^{(\infty)}$ и $T_c^{(\infty)}$. При снижении скорости нагрева и охлаждения от 3 ТК/с до 0,1 ТК/с разница между $T_m^{(\infty)}$ и $T_c^{(\infty)}$ уменьшилась от 4,24 кДж/моль до 0,67 кДж/моль, а между $H_m^{(\infty)}$ и $H_c^{(\infty)}$ — от 1,99 Дж/(моль·К) до 0,16 Дж/(моль·К). Можно сделать вывод, что при скорости 0,1 ТК/с зависимости теплоты плавления и кристаллизации частицы от размера и энтропии плавления и кристаллизации частицы от размера почти сливаются в «макроскопическом пределе». Предположительно, это обусловлено низкими скоростями нагрева и охлаждения, при которых в наночастице меньше образуются кристаллические структуры, отличные от гранецентрированных кубических.

References

- Skripov V.P., Koverda V.P. Spontaneous Crystallization of Supercooled Liquids: Nucleation of Crystals in Liquids and Amorphous Solids. Moscow, Nauka Publ., 1984. 232 p. (in Russian)
- Sdobnyakov N.Yu., Komarov P.V., Sokolov D.N., Samsonov V.M. Study of the thermodynamic characteristics of gold nanoclusters using a Gupta many-body potential, The Physics of Metals and Metallography, 2011, vol. 111, no. 1, pp. 13–20. https://doi.org/10.1134/ S0031918X11010121
- Samsonov V.M., Kharechkin S.S., Gafner S.L., Redel' L.V., Gafner Yu.Ya. Molecular dynamics study of the melting and crystallization of nanoparticles, *Crystallography Reports*, 2009, vol. 54, no. 3, pp. 526– 531. https://doi.org/10.1134/S1063774509030250
- Sdobnyakov N.Y., Sokolov D.N., Bazulev A.N., Samsonov V.M., Zykov T.Y., Antonov A.S. Relation between the size dependences of the melting and crystallization temperatures of metallic nanoparticles. *Russian Metallurgy (Metally)*, 2013, vol. 2013, no. 2, pp. 100–105. https://doi.org/10.1134/S0036029513020110
- Gupta R.P. Lattice relaxation at a metal surface. *Physical Review B*, 1981, vol. 23, no. 12, pp. 6265–6270. https://doi.org/10.1103/ PhysRevB.23.6265
- Cleri F., Rosato V. Tight-binding potentials for transition metals and alloys. *Physical Review B*, 1993, vol. 48, no. 1, pp. 22–33. https://doi. org/10.1103/PhysRevB.48.22

Письма в Журнал экспериментальной и теоретической физики. 2016. Т. 103. № 1-2. С. 100–105.

- Grochola G., Russo S.P., Snook I.K. On fitting a gold embedded atom method potential using the force matching method // The Journal of Chemical Physics. 2005. V. 123. N 20. P. 204719. https://doi. org/10.1063/1.2124667
- Цыдыпов Д.Г., Номоев А.В. Вычисление зависимости температуры плавления наночастицы серебра от размера наночастицы с использованием различных потенциалов погруженного атома // Вестник Бурятского государственного университета. Химия. Физика. 2022. № 2-3. С. 22–30. https://doi.org/10.18101/2306-2363-2022-2-3-22-30
- Цыдыпов Д.Г., Номоев А.В., Гармаев Б.З. Определение теплофизических свойств и функции радиального распределения наночастицы серебра с применением численных методов // Доклады Академии наук высшей школы Российской Федерации. 2023. № 3 (60). С. 20–31. https://doi.org/10.17212/1727-2769-2023-3-20-31
- Thomson W. LX. On the equilibrium of vapour at a curved surface of liquid // The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science. 1871. V. 42. N 282. P. 448–452. https://doi. org/10.1080/14786447108640606
- Бабичев А.П. Физические величины: Справочник. М.: Энергоатомиздат, 1991. 1232 с.
- Соколов Д.Н., Сдобняков Н.Ю., Комаров П.В. Расчет размерных зависимостей теплоты плавления наночастиц металлов // Физикохимические аспекты изучения кластеров, наноструктур и наноматериалов. 2011. № 3. С. 229–238.
- Сдобняков Н.Ю., Комаров П.В., Колосов А.Ю., Новожилов Н.В., Соколов Д.Н., Кульпин Д.А. Расчет размерных зависимостей теплоты плавления наночастиц металлов // Конденсированные среды и межфазные границы. 2013. Т. 15. № 3. С. 337–344.
- Самсонов В.М., Сдобняков Н.Ю., Васильев С.А., Соколов Д.Н. О размерной зависимости теплоты плавления металлических нанокластеров // Известия Российской академии наук. Серия физическая. 2016. Т. 80. № 5. С. 547–550. https://doi.org/10.7868/ S0367676516050161
- Самсонов В.М., Васильев С.А., Бембель А.Г., Самсонов Т.Е., Скопич В.Л. Молекулярно-динамическое исследование размерной зависимости теплоты плавления металлических нанокластеров // Физика твёрдого тела. 2014. Т. 56. № 12. С. 2289–2292.
- Сдобняков Н.Ю., Соколов Д.Н., Мясниченко В.С., Базулев А.Н. Расчет размерных зависимостей теплот плавления и кристаллизации наночастиц металлов // Физико-химические аспекты изучения кластеров, наноструктур и наноматериалов. 2014. № 6. С. 342–348.
- Sdobnyakov N.Yu., Veselov A.D., Ershov P.M., Sokolov D.N., Samsonov V.M., Vasilyev S.A., Myasnichenko V.S. Size dependence of the entropies of melting and crystallisation of metal nanoparticles // Computational Materials Science. 2018. V. 153. P. 153–158. https:// doi.org/10.1016/j.commatsci.2018.06.037

Авторы

Цыдыпов Дамдин Галсанович — аспирант, Институт физического материаловедения СО РАН, Улан-Удэ, 670047, Российская Федерация, https://orcid.org/0009-0005-5571-5075, damdinkatsydypov@gmail.com

Номоев Андрей Валерьевич — доктор физико-математических наук, доцент, заведующий лабораторией, Институт физического материаловедения СО РАН, Улан-Удэ, 670047, Российская Федерация, sc 6505488442, https://orcid.org/0009-0000-6201-6743, nomoevav@ mail.ru

Гармаев Баир Заятуевич — кандидат физико-математических наук, старший научный сотрудник, Институт физического материаловедения СО РАН, Улан-Удэ, 670047, Российская Федерация, sc 50561272800, https://orcid.org/0000-0001-6086-3658, bair.garmaev@gmail.com

Статья поступила в редакцию 19.10.2024 Одобрена после рецензирования 29.01.2025 Принята к печати 20.03.2025

- Samsonov V.M., Vasilyev S.A., Talyzin I.V., Ryzhkov Y.A. On reasons for the hysteresis of melting and crystallization of nanoparticles. *JETP Letters*, 2016, vol. 103, no. 2, pp. 97–99. https:// doi.org/10.1134/S0021364016020119
- Grochola G., Russo S.P., Snook I.K. On fitting a gold embedded atom method potential using the force matching method. *The Journal of Chemical Physics*, 2005, vol. 123, no. 20, pp. 204719. https://doi. org/10.1063/1.2124667
- Tsydypov D.G., Nomoev A.V. Calculation of the dependence of the melting point temperature of a silver nanoparticle on the nanoparticle size using different embedded atom potentials. *BSU bulletin. Chemistry. Physics*, 2022, no. 2-3, pp. 22–30. (in Russian). https:// doi.org/10.1063/1.2124667
- Tsydypov D.G., Nomoev A.V., Garmaev B.Z. Determination of thermophysical properties and the radial distribution function of a silver nanoparticle using numerical methods. *Proceedings of the Russian Higher School Academy of Sciences*, 2023, no. 3 (60), pp. 20– 31. (in Russian). https://doi.org/10.17212/1727-2769-2023-3-20-31
- Thomson W. LX. On the equilibrium of vapour at a curved surface of liquid. *The London, Edinburgh, and Dublin Philosophical Magazine* and Journal of Science, 1871, vol. 42, no. 282, pp. 448–452. https:// doi.org/10.1080/14786447108640606
- 12. Babichev A.P. *Physical Quantities*. Handbook. Moscow, Energoatomizdat Publ., 1991. 1232 p. (in Russian)
- Sokolov D.N., Sdobnyakov N.Yu., Komarov P.V. The calculation of the size dependencies of the heat of fusion for metal nanoparticles. *Physical and Chemical Aspects of the Study of Custers, Nanostructures* and Nanomaterials, 2011, no. 3, pp. 229–138. (in Russian)
- Sdobnyakov N.Yu., Komarov P.V., Kolosov A.Yu., Novozhilov N.V., Sokolov D.N., Kulpin D.A. The calculation of the size dependencies of the heat of fusion for metal nanoparticles. *Condensed Matter and Interphases*, 2013, vol. 15, no. 3, pp. 334–344. (in Russian)
- Samsonov V.M., Sdobnyakov N.Yu., Vasilyev S.A., Sokolov D.N. On the size dependence of the heats of melting of metal nanoclusters. *Bulletin of the Russian Academy of Sciences: Physics*, 2016, vol. 80, no. 5, pp. 494–496. https://doi.org/10.3103/S1062873816050166
- Samsonov V.M., Vasilyev S.A., Bembel A.G., Samsonov T.E., Skopich V.L. Molecular dynamics investigation of the size dependence of the heat of melting of metal nanoclusters. *Physics of the Solid State*, 2014, vol. 56, no. 12, pp. 2369–2373. https://doi. org/10.1134/S1063783414120270
- Sdobnyakov N.Yu., Sokolov D.N., Myasnichenko V.S., Bazulev A.N. Calculation of the heat of fusion and the heat of freezing size dependencies for metal nanoparticles. *Physical and Chemical Aspects of the Study of Custers, Nanostructures and Nanomaterials*, 2014, no. 6, pp. 342–348. (in Russian)
- Sdobnyakov N.Yu., Veselov A.D., Ershov P.M., Sokolov D.N., Samsonov V.M., Vasilyev S.A., Myasnichenko V.S. Size dependence of the entropies of melting and crystallisation of metal nanoparticles. *Computational Materials Science*, 2018, vol. 153, pp. 153–158. https://doi.org/10.1016/j.commatsci.2018.06.037

Authors

Damdin G. Tsydypov — PhD Student, Institute of Physical Material Science of the Siberian Branch of the RAS, Ulan-Ude, 670047, Russian Federation, https://orcid.org/0009-0005-5571-5075, damdinkatsydypov@gmail.com

Andrey V. Nomoev — D.Sc. (Physics & Mathematics), Associate Professor, Head of Laboratory, Institute of Physical Material Science of the Siberian Branch of the RAS, Ulan-Ude, 670047, Russian Federation, Sc 6505488442, https://orcid.org/0009-0000-6201-6743, nomoevav@ mail.ru

Bair Z. Garmaev — PhD (Physics & Mathematics), Senior Researcher, Institute of Physical Material Science of the Siberian Branch of the RAS, Ulan-Ude, 670047, Russian Federation, sc 50561272800,

https://orcid.org/0000-0001-6086-3658, bair.garmaev@gmail.com

Received 19.10.2024 Approved after reviewing 29.01.2025 Accepted 20.03.2025



Работа доступна по лицензии Creative Commons «Attribution-NonCommercial» **I/İTMO**

НАУЧНО-ТЕХНИЧЕСКИЙ ВЕСТНИК ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ, МЕХАНИКИ И ОПТИКИ март–апрель 2025 Том 25 № 2 http://ntv.ifmo.ru/ SCIENTIFIC AND TECHNICAL JOURNAL OF INFORMATION TECHNOLOGIES, MECHANICS AND OPTICS March–April 2025 Vol. 25 No 2 http://ntv.ifmo.ru/en/ ISSN 2226-1494 (print) ISSN 2500-0373 (online)

ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ, МЕХАНИКИ И ОПТИКИ

университет итмо

doi: 10.17586/2226-1494-2025-25-2-236-242 УДК 621.382.2/.3

Исследование геометрических параметров кремниевых структур в приборном слое при изготовлении чувствительных элементов микромеханических акселерометров Никита Сергеевич Каранин[⊠]

AO «Концерн «ЦНИИ «Электроприбор», Санкт-Петербург, 197046, Российская Федерация Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация karanin.ns@gmail.com[⊠], https://orcid.org/0009-0007-5086-0833

Аннотация

Введение. Исследованы технологические процессы изготовления инерциальных чувствительных элементов изделий микроэлектромеханических систем. Изучено влияние площади травления на кинетические параметры процесса глубокого реактивно-ионного травления, позволяющего формировать кремниевые структуры с высоким аспектным соотношением для изготовления микромеханических акселерометров и гироскопов. Метод. Инерциальные чувствительные элементы микромеханических акселерометров были изготовлены на подложке диаметром 150 мм по усовершенствованному технологическому процессу, позволяющему минимизировать площадь травления на этапе формирования приборного слоя, который состоит из инерционной массы, упругого подвеса, управляющих и измерительных электродов, изоляционной рамки. Значения геометрических параметров кремниевых структур приборного слоя были получены путем анализа профилей инерциальных чувствительных элементов на сканирующем электронном микроскопе. Исследованы элементы приборного слоя по всему диаметру в радиальном и тангенциальном направлениях подложки для определения разброса геометрических параметров инерциальных чувствительных элементов. Основные результаты. Представлен технологический процесс изготовления инерциальных чувствительных элементов, позволяющий снизить площадь травления на этапе формирования приборного слоя, за счет альтернативного способа вскрытия области вывода контактов. Основываясь на измерениях геометрических параметров кремниевых структур приборного слоя, установлено, что размеры элементов и их отклонения изменяются в радиальном направлении от центра подложки к краю. Разброс значений геометрических параметров кремниевых структур инерциальных чувствительных элементов, изготовленных по усовершенствованному технологическому процессу на подложке диаметром 150 мм, был снижен до 0,4 мкм, а разброс их отклонений уменьшен до 0,2 мкм. Обсуждение. Предложенный технологический процесс может быть использован для повышения выхода годных при изготовлении инерциальных чувствительных элементов и однородности функциональных характеристик изделий микроэлектромеханических систем, таких как акселерометры и гироскопы. Результаты работы могут быть применены при проектировании технологических процессов изготовления новых инерциальных чувствительных элементов.

Ключевые слова

глубокое реактивно-ионное травление, Bosch-процесс, приборный слой, инерциальный чувствительный элемент, микромеханический акселерометр

Ссылка для цитирования: Каранин Н.С. Исследование геометрических параметров кремниевых структур в приборном слое при изготовлении чувствительных элементов микромеханических акселерометров // Научнотехнический вестник информационных технологий, механики и оптики. 2025. Т. 25, № 2. С. 236–242. doi: 10.17586/2226-1494-2025-25-2-236-242

© Каранин Н.С., 2025

Investigation of geometric parameters of silicon structures in device layer during manufacture of sensitive elements of micromechanical accelerometers

Nikita S. Karanin⊠

Concern CSRI Electropribor, JSC, Saint Petersburg, 197046, Russian Federation ITMO University, Saint Petersburg, 197101, Russian Federation karanin.ns@gmail.com[⊠], https://orcid.org/0009-0007-5086-0833

Abstract

The technological processes of fabrication inertial elements of devices of microelectromechanical systems are investigated. The influence of etch square on critical parameters of the process of deep reactive ion etching, allowing to etch silicon structures with high aspect ratios for fabrication micromechanical accelerometers and gyroscopes, is studied. Inertial sensitive elements of micromechanical accelerometers were manufactured on a 150 mm wafer diameter within the framework of an advanced technological process with minimized square etch area on stage of formation of the device layer consisting of an inertial mass, an elastic suspension, control and measuring electrodes, and insulating frame. Values of geometric parameters of silicon structural layers of the device were obtained by analyzing the profiles of inertial visible elements on a scanning electron microscope. Elements of device layer were studied both in the radial and tangential directions of a substrate with a diameter of 150 mm to determine the spread of geometric parameters of inertial sensitive elements. The technological process of fabrication inertial sensitive elements to reduce square of etch area at the stage of device layer formation using an alternative opening the area of the contacts is shown. Based on measurements of the geometric parameters of the silicon structures of the device layer, it was found that the dimensions of the elements and their deviations change in the radial direction from the center of the substrate to the edge. The spread of the geometric parameters of the silicon structures of inertial sensitive elements manufactured according to the advanced technological process on a 150 mm diameter substrate was reduced to 0.4 µm, and the spread of their deviations was reduced to 0.2 µm. The proposed technological process can be used to increase the yield of devices goods during manufacture of inertial sensitive elements and to obtain more uniform characteristics of microelectromechanical systems, such as accelerometers and gyroscopes. The work results can be used in the design of technological processes for the manufacture of new inertial sensitive elements.

Keywords

deep reactive ion etching, Bosch process, device layer, inertial sensitive elements, micromechanical accelerometer

For citation: Karanin N.S. Investigation of geometric parameters of silicon structures in device layer during manufacture of sensitive elements of micromechanical accelerometers. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 2025, vol. 25, no. 2, pp. 236–242 (in Russian). doi: 10.17586/2226-1494-2025-25-2-236-242

Введение

Инерциальные чувствительные элементы изделий микроэлектромеханических систем, такие как микромеханические акселерометры и гироскопы, активно используются в составе микромодулей в различных отраслях науки и техники [1], в том числе в навигационных целях [2]. Главными преимуществами инерциальных микромодулей являются низкая стоимость и компактность [1], что достигается за счет технологии изготовления чувствительных элементов [3].

На подложках из монокристаллического кремния или кварцевого стекла формируют поэтапно слои с определенной геометрией одновременно для множества будущих чувствительных элементов. Число элементов, размещаемых на одной подложке, варьируется от нескольких десятков до нескольких тысяч в зависимости от диаметра подложки. Конструкционно инерциальные чувствительные элементы состоят из пластины-основания, приборного слоя и пластины-крышки [4, 5], что позволяет обеспечить герметичность объема, внутри которого размещают подвижную структуру акселерометра или гироскопа [6]. На рис. 1 представлено изображение, полученное на растровом электронном микроскопе (РЭМ), инерциального чувствительного элемента микромеханического акселерометра.

Технологический процесс создания инерциальных чувствительных элементов характеризуется следующими этапами [7, 8]. Этап 1. Изготовление пластины-основания *1* с полостями под подвижную структуру.

Этап 2. Формирование приборного слоя 2 посредством сращивания через слой диэлектрика [9] или металла [10] с пластиной-основанием *1*.

Этап 3. Глубокое реактивно-ионное травление приборного слоя 2 для формирования подвижной и неподвижных структур и области под контакты [11].



Рис. 1. РЭМ-изображение инерциального чувствительного элемента микромеханического акселерометра:

I — кремниевая пластина-основание; 2 — кремниевый приборный слой; 3 — кремниевая пластина-крышка; 4 — внешняя контактная область

Fig. 1. SEM image of the inertial sensitive element of a micromechanical accelerometer, where *1* is the silicon main wafer, *2* is the silicon device layer, *3* is the silicon cover wafer, *4* is the outer contact area

Этап 4. Сращивание с пластиной-крышкой 3 посредством эвтектической пайки [12], на которой предварительно выполняют металлические слои для контактных дорожек и зоны герметизации.

Этап 5. Травление контактной области для обеспечения доступа к контактам.

На характеристики инерциальных микромодулей наибольшее влияние оказывает процесс формирования структур в приборном слое, так как акселерометры и гироскопы чаще всего являются емкостными преобразователями, а жесткость конструкции определяется геометрическими параметрами упругого подвеса [7, 13]. В связи с этим размеры элементов в приборном слое являются критичными и на них устанавливают жесткие требования.

Элементы в приборном слое образуются за счет глубокого реактивно-ионного травления Bosch-процессом через защитную маску [11]. Процесс травления является прецизионным и позволяет формировать структуры в кремнии с высоким аспектным соотношением и вертикальными стенками [14]. Одним из ключевых параметров процесса является скорость травления и ее неравномерность по подложке, которая зависит в большей мере от соотношения площадей травления и подложки [15]. За счет необходимости травления одновременно и контактной области, и полостей вокруг подвижной структуры отношение площади травления областей незащищенной маской к площади подложки может составлять не менее 50 %. Оптимально для достижения минимальной неравномерности скорости травления отношение площади травления к площади подложки должно быть не более 20 %. Существует несколько способов снижения соотношения площади травления к площади подложки: увеличение габаритных размеров чувствительных элементов, что снижает общее число элементов на подложке; более сложные процессы изготовления за счет отсутствия необходимости травить область под контакты на этапе 2 формирования подвижной структуры, но большего числа слоев [4, 16, 17], изготовление открытых структур, которые являются менее надежными и негерметичными [8, 18].

Повышение надежности микромеханических акселерометров является актуальной задачей особенно в вопросах безопасности жизнедеятельности. Благодаря исследованиям надежности микромеханических акселерометров было выявлено множество механизмов отказа изделий [19, 20]. Такие механизмы, как слипание подвижной структуры с неподвижными элементами и механическое разрушение элементов упругого подвеса в результате ударных нагрузок, приводят к выходу изделия из строя за счет малого запаса прочности и жесткости элементов микромеханического акселерометра.

Целью настоящей работы является исследование снижения разброса геометрических параметров элементов приборного слоя микромеханических акселерометров путем уменьшения соотношения площади травления к площади подложки при сохранении общего числа изготавливаемых чувствительных элементов и их габаритных размеров.

Методы изготовления и исследования экспериментальных чувствительных элементов

Одним из наиболее технологичных процессов изготовления инерциальных чувствительных элементов является способ с обратным выводом контактов [21]. Этот способ позволяет достичь минимального количества слоев и операций для изготовления инерциального чувствительного элемента микромеханического акселерометра или гироскопа. Особенностью этого технологического процесса является прямой доступ к внешним контактам на пластине-крышке 3, что достигается за счет травления полостей в приборном слое 2 и пластине-основании 1 (рис. 1). Технологически это реализовано следующим образом [21]: изготавливается пластина-основание с полостями в кремнии и слое диоксида кремния (рис. 2, а), после чего пластина-основание сращивается с приборным слоем, на котором формируют глубоким реактивно-ионным травлением полости вокруг подвижной структуры и в области вывода контактов (рис. 2, b). Это необходимо для последующего удаления диоксида кремния и возможности вскрытия внешней контактной области со стороны пластины-основания. Для снижения площади травления на этапе 3 глубокого реактивно-ионного травления приборного слоя предлагается при изготовлении пластины-основании формировать полость в слое диоксида кремния в области вывода контактов (рис. 2, c), после чего в приборном слое необходимо вытравливать только полость вокруг подвижной структуры (рис. 2, *d*).

Реализовывалось изготовление пластины-основания по предлагаемому способу с помощью дополнительного этапа 6 фотолитографии и травления. После окисления кремниевой пластины на ее лицевой стороне выполнена фотолитография и протравка полости 4до слоя кремния (рис. 2, c). Затем последовательно вытравливался диоксид кремния и кремний через фоторезистивную маску, закрывающую полость 4, для формирования полостей 3 (рис. 2, c).

Для исследования геометрических параметров элементов приборного слоя были изготовлены инерциальные чувствительные элементы микромеханических акселерометров по приведенным технологическим процессам на подложках диаметром 150 мм, значения отношений площадей травления к площади подложки на этапе 2 формирования приборного слоя составляли 36 % и 20 % для стандартного [21] и улучшенного технологических процессов соответственно. Формирование структур в приборном слое проводилось на установке глубокого реактивно-ионного травления Plasma Therm Versaline Bosch-процессом по двухстадийному циклическому режиму, параметры стадий пассивации и травления которого приведены в таблице.

Исследование геометрических параметров производились путем определения размеров кремниевых структур на изображениях профилей, полученных на сканирующем электронном микроскопе Tescan Mira 3 (рис. 3). Наиболее существенными являются геометрические параметры элементов электродных структур, размеры которых были определены с помощью сканирующего электронного микроскопа



Рис. 2. Схематические изображения этапов технологического процесса изготовления инерциального чувствительного элемента: формирование пластины-основания (*a*) и структуры приборного слоя (*b*) стандартным технологическим процессом, формирование пластины-основания (*c*) и структуры приборного слоя (*d*) усовершенствованным технологическим процессом.

I — кремниевая пластина-основание; 2 — слой диоксида кремния; 3 — полость в кремниевой пластине-основании; 4 — полость в области вывода контактов в слое диоксида кремния; 5 — кремниевый приборный слой; 6 — подвижная структура в приборном слое; 7 — полость в области вывода контактов

Fig. 2. Schematic images of the stages of the technological process for fabrication an inertial sensitive element: formation of the main wafer (*a*) and the structure of the device layer (*b*) by the standard technological process, formation of the main wafer (*c*) and the structure of the device layer (*d*) by the advanced technological process; where *I* is a silicon main wafer, 2 is a silicon dioxide layer, 3 is a cavity in the silicon main wafer, 4 is a cavity in the area of the contact output in the silicon dioxide layer, 5 is a silicon device layer, 6 is a movable structure in the device layer, 7 is a cavity in the area of the contact output

Таблица. Параметры глубокого реактивно-ионного травления приборного слоя микромеханического акселерометра *Table*. Parameters of deep reactive ion etching of device layer of micromechanical accelerometer

Панациян	Значение		
Параметр	Пассивация	Травление	
Рабочее давление, Па	3,3	4,0	
Расход С ₄ F ₈ , см ³ /мин	150		
Расход SF ₆ , см ³ /мин		150	
Мощность разряда частотой 2 МГц, Вт	1500		
Напряжение смещения, В	10	400	
Продолжительность стадии, с	2,5	5	
Температура подложкодержателя, °С	-15		
Количество циклов 360		50	



Рис. 3. РЭМ-изображения профиля элементов электродной структуры (*a*) и чувствительного элемента микромеханического акселерометра (*b*)

Fig. 3. SEM images of the profile of the electrode structure elements (a) and the sensitive element of the micromechanical accelerometer (b)

(рис. 3, a). Важными геометрическими параметрами являются среднее значение размера и разброс значений размеров элементов вдоль толщины приборного слоя. Были определены значения толщин элементов электродных структур в крайних положениях и в центре (рис. 3, a).

Для определения характера распределения геометрических параметров элементов приборного слоя были исследованы чувствительные элементы как в радиальном направлении подложки, так и в тангенциальном.

Результаты и обсуждение

На основе полученных данных значений толщин элементов электродных структур были определены их средние значения (рис. 4, a) и разброс значений (рис. 4, b) для чувствительных элементов микромеханических акселерометров, изготовленных по стандартному [21] и усовершенствованному процессу на подложках диаметром 150 мм. Усовершенствованный технологический процесс позволил снизить разброс средних значений толщин элементов в приборном слое на 33 % (рис. 4, a). Также значительно большее количество чувствительных элементов имеет минимальный разброс значений толщин кремниевых структур в приборном слое (рис. 4, b) за счет характера распределения травления чувствительных элементов по подложке.

На рис. 5 представлено распределение средних значений толщин элементов приборного слоя чувствительных элементов по подложке диаметром 150 мм. Из распределения были исключены технологические элементы (белые прямоугольники), которые необходимы исключительно для выполнения определенных операций при изготовлении акселерометров. Преимущественно геометрические параметры элементов изменяются от центра к краю подложки, что связано с диффузией реагентов и продуктов реакций в процессе глубокого реактивно-ионного травления [15] и более высокой концентрацией радикалов и ионов в центре вакуумной камеры за счет расположения волновода источника высокочастотного газового разряда вокруг вакуумной камеры. Следует отметить, что распределение не является симметричным как относительно центрального чувствительного элемента, так и геометрического центра подложки. Такой характер распределения был выявлен на всех исследуемых подложках, что может быть связано в меньшей степени со смещением рисунка при выполнении фотолитографии из-за неидеально расположенных меток совмещения относительно центра подложки. А в большей степени с герметизацией подложки по краю на керамическом подложкодержателе, с образующейся внутренней полостью, в которую напускают и поддерживают гелий под давлением 500-2000 Па. Смещение центров подложки и подложкодержателя приводит к неравномерной утечке гелия из-под подложки и, как следствие, ее охлаждению, что и может оказывать влияние на характер распределения.

На основе анализа глубин травления зазора шириной 3 мкм было установлено, что увеличилось среднее значение скорости травления с 1,44 до 1,56 мкм/мин и снизилась ее неравномерность по подложке с 7,4 % до 5,9 % за счет уменьшения соотношения площади травления к площади подложки в 1,8 раз (с 36 % до 20 %), что согласуется с результатами в работе [15].

При травлении кремния вклад вносят две составляющие: ионная — за счет физического распыления материала и реакционная — за счет химического взаимодействия радикалов и ионов фтора с кремнием. Полимер, который осаждается на стадии пассивации и предотвращает травление боковых стенок, в первые секунды стадии травления удаляется ионной бомбардировкой со дна вытравливаемых полостей. Так как концентрация ионов больше в центре вакуумной камеры, полимер удаляется быстрее в центре подложки, за счет чего и продолжительность травления кремния





Fig. 4. Histograms of the thicknesses average values distribution of the electrode structures elements (*a*) and the spread of the thicknesses values of the electrode structures elements (*b*) with accounting the intersection of values (dark gray columns)



Рис. 5. Распределение средних значений толщин элементов электродных структур в чувствительных элементах микромеханических акселерометров по подложке диаметром 150 мм

Fig. 5. Distribution of the thicknesses average values of electrode structures elements in sensitive elements of micromechanical accelerometers on a 150 mm diameter substrate

Литература

- Barzegar M., Blanks S., Sainsbury B.-A., Timms W. MEMS technology and applications in geotechnical monitoring: a review // Measurement Science and Technology. 2022. V. 33. N 5. P. 052001. https://doi. org/10.1088/1361-6501/ac4f00
- Fitzgerald A.M. MEMS Inertial Sensors // Position, Navigation, and Timing Technologies in the 21st Century: Integrated Satellite Navigation, Sensor Systems, and Civil Applications. Wiley, 2021. P. 1435–1446. https://doi.org/10.1002/9781119458555.ch45
- Naumenko D., Tkachenko A., Lysenko I., Kovalev A. Development and research of the sensitive element of the MEMS gyroscope manufactured using SOI technology. Micromachines. 2023. V. 14. N 4. P. 895. https://doi.org/10.3390/mi14040895
- Torunbalci M., Alper S., Akin T. Advanced MEMS process for wafer level hermetic encapsulation of MEMS devices using SOI cap wafers with vertical feedthroughs // Journal of Microelectromechanical Systems. 2015. V. 24. N 3. P. 556–564. https://doi.org/10.1109/ JMEMS.2015.2406341
- Zoschke K., Mackowiak P., Kröhnert K., Oppermann H., Jürgensen N., Wietstruck M., Göritz A., Wipf ST., Kaynak M., Lang K.D. Cap fabrication and transfer bonding technology for hermetic and quasi hermetic wafer level MEMS packaging // Proc. of the IEEE 70th Electronic Components and Technology Conference (ECTC). 2020. P. 432–438. https://doi.org/10.1109/ECTC32862.2020.00076
- Torunbalci M., Gavcar H., Yesil F., Alper S., Akin T. An all-silicon process platform for wafer-level vacuum packaged MEMS devices // IEEE Sensors Journal. 2021. V. 21. N 13. P. 13958–13964. https://doi. org/10.1109/JSEN.2021.3073928
- Евстифеев М.И. Методы проектирования конструкций микромеханических гироскопов. Учебное пособие. СПб.: Университет ИТМО, 2018. 182 с.
- 8. Тимошенков С.П., Анчутин С.А., Зарянкин Н.М., Калугин В.В., Кочурина Е.С., Тимошенков А.С., Боев Л.Р. Проектирование и изготовление чувствительного элемента МЭМС-акселерометра // Нано- и микросистемная техника. 2021. Т. 23. № 2. С. 63–67. https://doi.org/10.17587/nmst.23.63-67

локально больше в пределах одного цикла, но скорость травления кремния начинает уменьшаться по мере удаления полимера в радиальном направлении подложки. Отметим, что при этом скорость травления принимает большие значения там, где продуктам проще покинуть зону реакций, т. е. ближе к краю подложки. За счет этого эффекта изменение соотношения площади травления к площади подложки влияет сильнее на размеры формируемых элементов, расположенных ближе к краю, чем к центру подложки.

Заключение

На основе полученных экспериментальных данных приведено распределение толщин элементов, характеризующееся изменением значений от центра к краю подложки. Показано, что снижение соотношения площади травления к площади подложки оказывает значительно большее влияние на среднюю толщину элементов и разброс их значений, чем на глубину формируемых полостей. Представленные результаты можно использовать для повышения выхода годных изделий и их надежности, а также при проектировании новых инерциальных чувствительных элементов микромеханических акселерометров и гироскопов.

References

- Barzegar M., Blanks S., Sainsbury B.-A., Timms W. MEMS technology and applications in geotechnical monitoring: a review. *Measurement Science and Technology*, 2022, vol. 33, no. 5, pp. 052001. https://doi. org/10.1088/1361-6501/ac4f00
- Fitzgerald A.M. MEMS Inertial Sensors. Position, Navigation, and Timing Technologies in the 21st Century: Integrated Satellite Navigation, Sensor Systems, and Civil Applications. Wiley, 2021, pp. 1435–1446. https://doi.org/10.1002/9781119458555.ch45
- Naumenko D., Tkachenko A., Lysenko I., Kovalev A. Development and research of the sensitive element of the MEMS gyroscope manufactured using SOI technology. *Micromachines*, 2023, vol. 14, no. 4, pp. 895. https://doi.org/10.3390/mi14040895
- Torunbalci M., Alper S., Akin T. Advanced MEMS process for wafer level hermetic encapsulation of MEMS devices using SOI cap wafers with vertical feedthroughs. *Journal of Microelectromechanical Systems*, 2015, vol. 24, no. 3, pp. 556–564. https://doi.org/10.1109/ JMEMS.2015.2406341
- Zoschke K., Mackowiak P., Kröhnert K., Oppermann H., Jürgensen N., Wietstruck M., Göritz A., Wipf ST., Kaynak M., Lang K.D. Cap fabrication and transfer bonding technology for hermetic and quasi hermetic wafer level MEMS packaging. *Proc. of the IEEE 70th Electronic Components and Technology Conference (ECTC)*, 2020, pp. 432–438. https://doi.org/10.1109/ECTC32862.2020.00076
- Torunbalci M., Gavcar H., Yesil F., Alper S., Akin T. An all-silicon process platform for wafer-level vacuum packaged MEMS devices. *IEEE Sensors Journal*, 2021, vol. 21, no. 13, pp. 13958–13964. https:// doi.org/10.1109/JSEN.2021.3073928
- 7. Evstifeev M.I. Design Methods for Micromechanical Gyroscope Structures. Study guide. St. Petersburg, ITMO, 2018. 182 p. (in Russian)
- Timoshenkov S.P., Anchutin S.A., Zarjankin N.M., Kalugin V.V., Kochurina E.S., Timoshenkov A.S., Boev L.R. Research and Development of MEMS Accelerometer's Sensor. *Nano- and Microsystems Technology*, 2021, vol. 23, no. 2, pp. 63–67. (in Russian). https://doi.org/10.17587/nmst.23.63-67

- Xu J., Ren Z., Dong B., Wang C., Tian Y., Lee C. Evolution of Wafer Bonding Technology and Applications from Wafer-Level Packaging to Micro/Nanofluidics-Enhanced Sensing // Advanced MEMS/NEMS Fabrication and Sensors. Springer, 2022. P. 187–215. https://doi. org/10.1007/978-3-030-79749-2
- Oggioni L., Garavaglia M., Seghizzi L. Wafer-to-Wafer Bonding // Silicon Sensors and Actuators: The Feynman Roadmap. Springer, 2022. P. 345–386. https://doi.org/10.1007/978-3-030-80135-9_11
- Karanin N.S. Deep reactive ion etching of device layer during manufacture micromechanical accelerometer // Proc. of the Conference of Russian Young Researchers in Electrical and Electronic Engineering (ElConRus). 2022. P. 962–965. https://doi.org/ 10.1109/ ElConRus54750.2022.9755694
- Li D., Shang Z., She Y., Wen Z. Investigation of Au/Si eutectic wafer bonding for MEMS accelerometers // Micromachines. 2017. V. 8. N 5. P. 158. https://doi.org/10.3390/mi8050158
- Kavitha S., Daniel R.J., Sumangala K. Design and analysis of MEMS comb drive capacitive accelerometer for SHM and seismic applications // Measurement. 2016. V. 93. P. 327–339. https://doi. org/10.1016/j.measurement.2016.07.029
- Zhang Y., Wu Y., Sun Q., Shen L., Lan J., Guo L., Shen Z., Wang X., Xiao J., Xu J. Inductively coupled plasma dry etching of silicon deep trenches with extremely vertical smooth sidewalls used in microoptical gyroscopes // Micromachines. 2023. V. 14. N 4. P. 846. https:// doi.org/10.3390/mi14040846
- Alnakhli Z., Liu Zh., AlQatari F., Cao H., Li X. UV-assisted nanoimprint lithography: the impact of the loading effect in silicon on nanoscale pattern of metalens // Nanoscale Advances. 2024. V. 6. N 11. P. 2954–2967. https://doi.org/10.1039/D4NA00120F
- Wang X., Bleiker S.J., Edinger P., Errando-Herranz C., Roxhed N., Stemme G., Gylfason B., Niklaus F. Wafer-level vacuum sealing by transfer bonding of silicon caps for small footprint and ultra-thin MEMS packages // Journal of Microelectromechanical Systems. 2019. V. 28. N 3. P. 460–471. https://doi.org/10.1109/ JMEMS.2019.2910985
- Liu J., Xia S., Peng C., Wu Z., Chu Z., Zhang Z., Lei H., Zheng F., Zhang W. Wafer-level vacuum-packaged electric field microsensor: structure design, theoretical model, microfabrication, and characterization // Micromachines. 2022. V. 13. N 6. P. 928. https:// doi.org/10.3390/mi13060928
- Belyaev Y.V., Belogurov A.A., Bocharov A.N., Kostygov D.V., Lemko I.V., Mihteeva A.A. Design of a micromechanical accelerometer // Proc. of the 25th International Conference on Integrated Navigation Systems (ICINS). 2018. P. 1–7. https://doi. org/10.23919/ICINS.2018.8405921
- Xu Y., Liu S., He C., Wu H., Cheng L., Yan G., Huang Q. Reliability of MEMS inertial devices in mechanical and thermal environments: a review // Heliyon. 2024. V. 10. N 5. P. e27481. https://doi. org/10.1016/j.heliyon.2024.e27481
- Peng T., You Z. Reliability of MEMS in shock environments: 2000– 2020 // Micromachines. 2021. V. 12. N 11. P. 1275. https://doi. org/10.3390/mi12111275
- Wenk B., Collet J., Gaff V. Technology platform for high performance Mems inertial & vibration sensors // Proc. of the IEEE International Symposium on Inertial Sensors and Systems (INERTIAL). 2024. P. 1–4. https://doi.org/10.1109/INERTIAL60399.2024.10502059

Автор

Каранин Никита Сергеевич — инженер-технолог, АО «Концерн «ЦНИИ «Электроприбор», Санкт-Петербург, 197046, Российская Федерация; аспирант, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, https://orcid.org/0009-0007-5086-0833, karanin. ns@gmail.com

Статья поступила в редакцию 07.12.2024 Одобрена после рецензирования 13.02.2025 Принята к печати 22.03.2025



- Xu J., Ren Z., Dong B., Wang C., Tian Y., Lee C. Evolution of Wafer Bonding Technology and Applications from Wafer-Level Packaging to Micro/Nanofluidics-Enhanced Sensing. *Advanced MEMS/NEMS Fabrication and Sensors*. Springer, 2022, pp. 187–215. https://doi. org/10.1007/978-3-030-79749-2_7
- Oggioni L., Garavaglia M., Seghizzi L. Wafer-to-Wafer Bonding. Silicon Sensors and Actuators: The Feynman Roadmap. Springer, 2022, pp. 345–386. https://doi.org/10.1007/978-3-030-80135-9_11
- Karanin N.S. Deep reactive ion etching of device layer during manufacture micromechanical accelerometer. Proc. of the Conference of Russian Young Researchers in Electrical and Electronic Engineering (ElConRus), 2022, pp. 962–965. https://doi.org/ 10.1109/ ElConRus54750.2022.9755694
- Li D., Shang Z., She Y., Wen Z. Investigation of Au/Si eutectic wafer bonding for MEMS accelerometers. *Micromachines*, 2017, vol. 8, no. 5, pp. 158. https://doi.org/10.3390/mi8050158
- Kavitha S., Daniel R.J., Sumangala K. Design and analysis of MEMS comb drive capacitive accelerometer for SHM and seismic applications. *Measurement*, 2016, vol. 93, pp. 327–339. https://doi. org/10.1016/j.measurement.2016.07.029
- Zhang Y., Wu Y., Sun Q., Shen L., Lan J., Guo L., Shen Z., Wang X., Xiao J., Xu J. Inductively coupled plasma dry etching of silicon deep trenches with extremely vertical smooth sidewalls used in microoptical gyroscopes. *Micromachines*, 2023, vol. 14, no. 4. pp. 846. https://doi.org/10.3390/mi14040846
- Alnakhli Z., Liu Zh., AlQatari F., Cao H., Li X. UV-assisted nanoimprint lithography: the impact of the loading effect in silicon on nanoscale pattern of metalens. *Nanoscale Advances*, 2024, vol. 6, no. 11, pp. 2954–2967. https://doi.org/10.1039/D4NA00120F
- Wang X., Bleiker S.J., Edinger P., Errando-Herranz C., Roxhed N., Stemme G., Gylfason B., Niklaus F. Wafer-level vacuum sealing by transfer bonding of silicon caps for small footprint and ultra-thin MEMS packages. *Journal of Microelectromechanical Systems*, 2019, vol. 28, no. 3, pp. 460–471. https://doi.org/10.1109/ JMEMS.2019.2910985
- Liu J., Xia S., Peng C., Wu Z., Chu Z., Zhang Z., Lei H., Zheng F., Zhang W. Wafer-level vacuum-packaged electric field microsensor: structure design, theoretical model, microfabrication, and characterization. *Micromachines*, 2022, vol. 13, no. 6, pp. 928. https:// doi.org/10.3390/mi13060928
- Belyaev Y.V., Belogurov A.A., Bocharov A.N., Kostygov D.V., Lemko I.V., Mihteeva A.A. Design of a micromechanical accelerometer. *Proc. of the 25th International Conference on Integrated Navigation Systems (ICINS)*, 2018, pp. 1–7. https://doi. org/10.23919/ICINS.2018.8405921
- Xu Y., Liu S., He C., Wu H., Cheng L., Yan G., Huang Q. Reliability of MEMS inertial devices in mechanical and thermal environments: a review. *Heliyon*, 2024, vol. 10, no. 5, pp. e27481. https://doi. org/10.1016/j.heliyon.2024.e27481
- Peng T., You Z. Reliability of MEMS in shock environments: 2000– 2020. *Micromachines*, 2021, vol. 12, no. 11, pp. 1275. https://doi. org/10.3390/mi12111275
- Wenk B., Collet J., Gaff V. Technology platform for high performance Mems inertial & vibration sensors. *Proc. of the IEEE International Symposium on Inertial Sensors and Systems (INERTIAL)*, 2024, pp. 1–4. https://doi.org/10.1109/INERTIAL60399.2024.10502059

Author

Nikita S. Karanin — Process Engineer, Concern CSRI Electropribor, JSC, Saint Petersburg, 197046, Russian Federation; PhD Student, ITMO University, Saint Petersburg, 197101, Russian Federation, https://orcid. org/0009-0007-5086-0833, karanin.ns@gmail.com

Received 07.12.2024 Approved after reviewing 13.02.2025 Accepted 22.03.2025

Работа доступна по лицензии Creative Commons «Attribution-NonCommercial» **I/İTMO**

НАУЧНО-ТЕХНИЧЕСКИЙ ВЕСТНИК ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ, МЕХАНИКИ И ОПТИКИ март–апрель 2025 Том 25 № 2 http://ntv.ifmo.ru/ SCIENTIFIC AND TECHNICAL JOURNAL OF INFORMATION TECHNOLOGIES, MECHANICS AND OPTICS March–April 2025 Vol. 25 No 2 http://ntv.ifmo.ru/en/ ISSN 2226-1494 (print) ISSN 2500-0373 (online)

ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ, МЕХАНИКИ И ОПТИКІ

АВТОМАТИЧЕСКОЕ УПРАВЛЕНИЕ И РОБОТОТЕХНИКА AUTOMATIC CONTROL AND ROBOTICS

doi: 10.17586/2226-1494-2025-25-2-243-252

Frequencies estimation in multisinusoidal time-varying parameter of the first order discrete linear system with the application to indirect adaptive control

Dang Hien Ngo¹, Dmitry N. Gerasimov²⊠

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

¹ danghien719@gmail.com, https://orcid.org/0000-0001-6673-921X

² gerasimovdn@mail.ru[™], dngerasimov@itmo.ru, https://orcid.org/0000-0001-8306-4138

Abstract

The paper addresses the problem of adaptive frequencies estimation for multisinusoidal Time-Varying (TV) parameter of a discrete linear system of the first order. It is assumed that the amplitudes, frequencies, and phases of the harmonics in the TV parameter are unknown, however the number of harmonics is known. The novelty of the proposed approach consists in the fact that the frequencies identification is possible even if the system output crosses zero when the information about the TV parameter is inaccessible. In this case, when the proposed solution is used in a problem of adaptive control of the system considered, the frequencies identification and the work of TV parameter observer are independent, what increases the rate and precision of controller parameters tuning. The problem is solved by transformation of the plant model into a regression model linear with respect to unknown frequencies and used for design of identification algorithms. In the paper, two identification algorithms are applied. The first one is the standard gradient algorithm, while the second one is the algorithm with improved parametric convergence achieved by accumulation of regressor over past period of time and referred to as algorithm with memory regressor extension. The problem of control is solved with the use of: certainty equivalence principle; internal model principle according to which the TV parameter is represented as the output of dynamic autonomous model (exosystem) and involving of this model into the structure of the control law; observer of the exosystem state; one of the proposed frequencies identifier; and a formula of recalculation of the frequencies estimates into the controller adjustable parameters. A procedure of transformation of the TV system into a regression linear with respect to unknown frequencies used for design of identification algorithms is represented. The obtained solution is applied to the problem of indirect (identification-based) adaptive control of the TV system considered in the paper. The main distinguishing feature of the solution proposed consists in independence of the obtained identifiers from the observation property of the TV parameter what increases the transient performance and precision of the indirect adaptive control algorithms designed for the considered class of TV systems. The proposed solution can be used in problems of control of parametric resonance systems.

Keywords

linear time-varying system; multisinusoidal time-varying parameter; uncertain periodic coefficients; adaptive identification; adaptation with improved convergence

Acknowledgements

Supported by the Ministry of Science and Higher Education of the Russian Federation (project no. FSER-2025-0002).

For citation: Ngo D.H., Gerasimov D.N. Frequencies estimation in multisinusoidal time-varying parameter of the first order discrete linear system with the application to indirect adaptive control. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 2025, vol. 25, no. 2, pp. 243–252. doi: 10.17586/2226-1494-2025-25-2-243-252

© Ngo D.H., Gerasimov D.N., 2025

УДК 618.51

Оценивание частот мультисинусоидального переменного параметра дискретной линейной системы первого порядка с применением в задаче непрямого адаптивного управления

Данг Хиен Нго¹, Дмитрий Николаевич Герасимов²⊠

1,2 Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация

¹ danghien719@gmail.com, https://orcid.org/0000-0001-6673-921X

² dngerasimov@itmo.ru[⊠], https://orcid.org/0000-0001-8306-4138

Аннотация

Введение. Рассмотрена задача адаптивного оценивания частот мультисинусоидального переменного во времени параметра дискретной линейной системы первого порядка с применением результата в задаче адаптивного управления. Предполагается, что амплитуды, частоты и фазы гармоник переменного параметра неизвестны, однако известно число этих гармоник. Новизна предлагаемого подхода заключается в том, что идентификация частот происходит даже при пересечении выходом системы нуля, когда недоступна информация о нестационарном параметре. В этом случае при использовании предложенного решения в задаче адаптивного управления рассматриваемым объектом идентификация частот и работа наблюдателя динамики переменного параметра происходят независимо, что повышает скорость и точность автонастройки параметров регулятора. Метод. Задача решается путем преобразования модели объекта к линейной по неизвестным частотам регрессии, на базе которой синтезируются алгоритмы идентификации. В работе применяется два алгоритма идентификации. Первый представляет собой стандартный градиентный алгоритм, в то время как второй — алгоритм адаптации с улучшенной сходимостью, обеспеченной за счет накопления предыдущих значений регрессора, и называется алгоритмом адаптации с памятью регрессора. Задача управления решена с использованием: принципа непосредственной компенсации; принципа внутренней модели, согласно которому переменный параметр представляется в виде выхода динамической автономной модели (экзосистемы) и внедрении этой модели в структуру закона управления; наблюдателя состояния экзосистемы; одного из двух предложенных идентификаторов частоты; формулы пересчета оценок частот в вектор настраиваемых параметров регулятора. Основные результаты. Приведена процедура сведения нестационарной системы к линейной по неизвестным частотам регрессионной модели, на базе которой построены алгоритмы идентификации. Полученное решение использовано в задаче непрямого (идентификационного) адаптивного управления рассматриваемой нестационарной системы. Обсуждение. Главной отличительной чертой предложенного решения является независимость полученных алгоритмов идентификации от свойства наблюдаемости нестационарного параметра, что повышает быстродействие и точность алгоритмов непрямого адаптивного управления рассматриваемым классом линейных нестационарных систем. Представленное решение может быть использовано при решении задач управления системами с параметрическим резонансом.

Ключевые слова

линейная нестационарная система, мультисинусоидальный переменный параметр, неопределенные периодические коэффициенты, адаптивная идентификация, адаптация с улучшенной сходимостью

Благодарности

Статья подготовлена при финансовой поддержке Министерства науки и высшего образования Российской Федерации, проект № FSER-2025-0002.

Ссылка для цитирования: Нго Д.Х., Герасимов Д.Н. Оценивание частот мультисинусоидального переменного параметра дискретной линейной системы первого порядка с применением в задаче непрямого адаптивного управления // Научно-технический вестник информационных технологий, механики и оптики. 2025. Т. 25, № 2. С. 243–252 (на англ. яз.). doi: 10.17586/2226-1494-2025-25-2-243-252

Introduction

In this paper, we develop an adaptive estimation technique for a class of discrete linear Time-Varying (TV) systems with uncertain multi-sinusoidal parameters and without disturbance.

Attempts of extension of identifiers and controllers design to TV systems made for the last decades are very challenging; however there exist potentials for increasing the scope of practical applications. One of the ways of such an extension is based on the Internal Model Principle (IMP) [1, 2] consisting in representation of the system TV parameters as the outputs of autonomous models (referred to as exosystems) with constant parameters and suitable incorporation of the structure of these models into the problem solution to completely compensate the effects caused by the TV parameters changing. In this case, as shown in [2], the exosystems can generate multisinusoidal functions of time, and, as a result, the corresponding design methods can be successfully used for systems with multisinusoidal TV parameters and then applied to different real-life systems [3–6].

From the practical point of view, assumption about the complete knowledge of the system TV parameters is hardly feasible, hence the interest of adaptive control and identification theory application. At the same time, as the literature survey showed, the overwhelming majority of identification based on direct adaptive controllers for systems with multisinusoidal or just periodic coefficients assume that the amplitudes and phases of the coefficients are unknown; however the periods or frequencies of the parameters variation are known [3, 7–12] or partially known (in [13] the frequency is factorized by an unknown integer and a known constant, in [10, 12] the frequency is unknown within a known interval). In [14], the authors propose an estimation technique for a linear system with sinusoidal parameters with completely unknown amplitudes, phases, and frequencies, however under assumptions that the plant state is bounded away from zero and the TV parameters contain only one harmonic. In [15, 16], the authors propose an approach of direct adaptive control for similar class of continuous time and discrete time linear systems with the state matrix containing multisinusoidal TV parameters on the main diagonal. The main drawback of the proposed method consists in dependence of controller parameters estimation (adaptation) on the plant state. So, if at least one element of the state vector crosses zero, the TV parameter is no longer to be observable and, as a result, the adaptive estimator being dependent on the TV parameters observers is not able to estimate the controller parameters.

Thus, in this paper we make an effort to overcome this problem by conversion of plant model into a linear regression with the regressor that is independent from the observer estimates and admits the state zero crossing. The regression is used for design of two frequencies estimators. The first estimator is based on the standard gradient adaptation algorithm, while the second one uses modified Kreisselemeier-like adaptation algorithm referred to as adaptation algorithm improved parametric convergence achieved by Memory Regressor Extension (MRE) [2, 17, 18]. Both estimators are applied to the proposed scheme of indirect (identification based) adaptive control.

In order to focus the readers' attention on the main idea and simplify the paper presentation, we start considering a discrete TV linear system of the first order with one multisinusoidal parameter in which, besides the unknown frequencies of the harmonics, the amplitudes and phases of these harmonics are assumed unknown. Then, it is shown how the proposed method can be straightforwardly extended to the more general class of linear systems considered in [16].

The remaining of the paper is organized as follows. In the second section, the problem of the frequencies identification is formulated. In the third section, the parameterization of plant model is presented. In the fourth section, the adaptive frequencies estimators are designed. In the fifth section, the identification algorithms are applied to the scheme of indirect adaptive control.

For the sake of completeness of the paper presentation, we represent the definition of the Persistent Excitation (PE) condition crucial for systems identification.

Definition 1. A bounded function $\varphi(k) \in \mathbb{R}^N$ is persistently exciting (i.e., $\in \mathcal{PE}$) if there exist such constants $\alpha, K \in \mathbb{R}_+$ such that:

$$\sum_{i=k+1}^{k+K} \boldsymbol{\varphi}(j) \boldsymbol{\varphi}^{\mathsf{T}}(j) \ge \alpha \mathbf{I}_n \ \forall k \ge 0.$$
(1)

Notations: z^{-1} is the delay operator; k is the sampling time; \mathcal{N} is the set of natural numbers; \mathbb{R}_+ is the set of positive real numbers; $\mathcal{P}\mathcal{E}$ is the set of persistently exciting functions satisfying (1); \mathbf{I}_n is the $n \times n$ identity matrix; \mathcal{L}_2 is the space of squared summable functions; \mathcal{L}_{∞} is the space of bounded functions; $C_N^i = \frac{N!}{i!(N-i)!}$ is the binomial coefficient.

Problem statement

Consider a discrete-time system

$$x(k+1) = \psi(k)x(k) + u(k),$$
 (2)

where $x(k) \in \mathbb{R}$ is the state with the initial condition $x(0), u(k) \in \mathbb{R}$ is the control signal, *k* is the sampling time, $\psi(k) \in \mathbb{R}$ is the uncertain time-varying parameter given by the function

$$\psi(k) = \sum_{i=1}^{N} a_i \sin(\omega_i kT + \varphi_i)$$
(3)

with a priori unknown constant amplitudes a_i , phases φ_i , and frequencies ω_i of harmonics, however known maximum number of harmonic N and the discrete time interval T^1 .

For the plant, we accept the following assumptions. **Assumption 1.**

- A1.1. The frequencies ω_i are different (distinguishable) and satisfy the conditions $\omega_i T \in (0; \pi]$;
- A1.2. None of the frequencies ω_i is zero, i.e., the signal
 (3) is not biased.

Assumption A1.1 allows us to identify the frequencies and is typical for harmonic identification problems. Assumption A1.2 is introduced for the sake of simplicity and can straightforwardly be extended to biased (i.e., with zero frequency) multisinusoidal parameters.

The objective is to design an identifier that under the PE condition generates estimates $\hat{\omega}_i(k)$ such that

$$\lim_{k \to \infty} (\omega_i - \hat{\omega}_i(k)) = 0.$$
(4)

It is necessary to note that compared to the identifier (adaptation algorithm) proposed in [16], this problem statement admits that the state x(k) can cross zero arbitrary number of times; therefore the solution presented below is less restrictive.

The solution proposed is based on linear parameterization of the plant model (2) superposition of measurable variables multiplied by constant parameters dependent from ω_i . Before to propose the general solution to the problem, we clarify the main idea of the parameterization considering the particular case of single harmonic parameter (3) with N = 1 and then extend this case to arbitrary N > 1.

Model parameterization

Single harmonic case

Assuming that

$$\psi(k) = a \sin(\omega kT + \varphi),$$

where *a*, ω , and φ are constant unknown parameters, and applying the IMP we represent $\psi(k)$ as the solution of the linear discrete equation

$$\beta(z)[\psi(k)] = 0 \tag{5}$$

with the operator $\beta(z) = 1 - cz^{-1} + z^{-2}$; the initial conditions $\psi(0) = a \sin(\varphi)$; $\psi(1) = a \sin(\omega T + \varphi)$ and the constant $c = 2\cos(\omega T)$.

¹ Hereafter, we will omit the dependence from k for the sake of brevity and if is not in contrary to the context.

From (2) we obtain

$$\psi(k) = \frac{x(k+1) - u(k)}{x(k)} \,. \tag{6}$$

Replacing $\psi(k)$ from the latter in (5), we have

$$\frac{x(k+1)-u(k)}{x(k)} - c \frac{x(k)-u(k-1)}{x(k-1)} + \frac{x(k-1)-u(k-2)}{x(k-2)} = 0$$

Multiplying this equality by x(k)x(k-1)x(k-2), applying the shift backward operator z^{-1} , and rearranging terms, we obtain the parameterized model of the plant in the form of linear regression

$$y(k) = \theta \varphi(k), \tag{7}$$

where $\theta = c = 2\cos(\omega T)$,

$$y(k) = x(k-2)x(k-3)(x(k) - u(k-1)) + x(k-1)x(k-2)(x(k-2) - u(k-3)),$$

$$\varphi(k) = x(k-1)x(k-3)(x(k-1) - u(k-2)).$$

Remark 1. The form (7) can be used for identification of the frequency estimate $\hat{\omega} = \frac{1}{T}a\cos\left(\frac{1}{2}\hat{\theta}\right)$, where $\hat{\theta}$ is the estimate of θ generated by an identifier, under Assumption 1.1 since the function $\cos(\cdot)$ is invertible in the interval $(0, \pi]$.

Now, extend the presented result to arbitrary number of harmonics.

Multiharmonic case

Now, assume that $\psi(k)$ is represented in the general form (3). In this case, the operator form of discrete equation generating $\psi(k)$ takes the form

$$\prod_{i=1}^{N} \beta_i(z)[\psi(k)] = 0 \tag{8}$$

with the operators $\beta_i(z) = 1 - c_i z^{-1} + z^{-2}$, the initial conditions $\psi(j) = \sum_{i=1}^{N} a_i \sin(\omega_i j T + \varphi_i)$ (j = 0, 1, ..., 2N - 1), and the constants $c_i = 2 \cos(\omega_i T)$.

Replacing $\psi(k)$ from (6) in (8) after straightforward but routine algebraic transformations we obtain the following result.

Lemma 1. The plant (2) with the parameter (3) can be transformed into the form of linear regression

$$y(k) = \mathbf{\Theta}^{\mathsf{T}} \mathbf{\phi}(k), \tag{9}$$

where $\boldsymbol{\varphi} = [\varphi_1, \varphi_2, ..., \varphi_N]^{\mathsf{T}}$ is the regressor with the elements

$$\varphi_{i}(k) = \sum_{j=0}^{N-i} \sum_{N-i}^{j} \left(\prod_{\substack{p=1\\p\neq 2j+i+1\\p\neq 2j+i+1\\-u(k-2j-i-1))}}^{2N+1} x(k-p) \right) (x(k-2j-i) - (10))$$

$$y(k) = \sum_{i=0N}^{N} C \left(\prod_{\substack{j=1\\ j\neq 2i+1}}^{2N+1} x(k-j) \right) (x(k-2i) - u(k-2i-1)), \quad (11)$$

is the regression output; $\mathbf{\theta} = \begin{bmatrix} \theta_1 & \theta_2 & \dots & \theta_N \end{bmatrix}^{\mathsf{T}}$ is the vector obtained using c_i via the Vieta's formulas

$$\begin{cases} \theta_1 = c_1 + c_2 + \dots + c_N, \\ \theta_2 = -(c_1 c_2 + c_1 c_3 + \dots + c_{N-1} c_N) \\ \vdots \\ \theta_N = (-1)^{N+1} c_1 c_2 \dots c_N. \end{cases}$$
(12)

The linear regression (9) can be used for identification of the parameters θ_i (i = 1, 2, ..., N) and, as a result, of the frequencies ω_i .

Adaptive frequencies identification

In order to estimate θ in (9), we use the gradient adaptation algorithm [19, 20] and the adaptation algorithm with improved parametric convergence based on the Kreisselmeier's like scheme and referred to as algorithm with MRE [17, 18].

Gradient algorithm

The algorithm takes the form

$$\hat{\boldsymbol{\theta}}(k+1) = \hat{\boldsymbol{\theta}}(k) + \gamma \frac{\boldsymbol{\varphi}(k)\boldsymbol{\varphi}^{\mathsf{T}}(k)}{1 + \gamma \boldsymbol{\omega}^{\mathsf{T}}(k)\boldsymbol{\omega}(k)} \boldsymbol{\varepsilon}(k), \qquad (13)$$

where $\varepsilon(k) = y(k) - \mathbf{\phi}^{\mathsf{T}}(k)\hat{\mathbf{\theta}}(k), \gamma > 0$ is adaptation gain. The algorithm (13) yields the parametric error model

$$\tilde{\boldsymbol{\theta}}(k+1) = \left(\mathbf{I}_N - \gamma \frac{\boldsymbol{\varphi}(k) \boldsymbol{\varphi}^{\mathsf{T}}(k)}{1 + \gamma \boldsymbol{\varphi}^{\mathsf{T}}(k) \boldsymbol{\varphi}(k)} \right) \tilde{\boldsymbol{\theta}}(k), \qquad (14)$$

where $\tilde{\theta}(k) = \theta - \hat{\theta}(k)$ is the parametric error vector. This model can be used for the proof of the following properties.

Proposition 1. The identification system closed by the algorithm (13) with $\boldsymbol{\varphi}$ defined in (10) and y defined by (11) *has the following properties:*

- $\operatorname{Pl}_{1} \tilde{1} \cdot \tilde{\theta}(k) \in \mathcal{L}_{\infty} \cap \mathcal{L}_{2}. If \, \varphi(k) \in \mathcal{L}_{\infty}, then \, \varepsilon(k), \, \hat{\theta}(k+1)$ $\begin{array}{l} -\hat{\theta}(k) \in \mathcal{L}_{\infty} \cap \mathcal{L}_{2}; \\ -\text{P1.2. If } \varphi(k) \in \mathcal{L}_{\infty}, \ then \ \varepsilon(k), \ \hat{\theta}(k+1) - \hat{\theta}(k) \to 0 \ as \end{array}$
- $k \to \infty$;
- P1.3. $\hat{\mathbf{0}}(k) \rightarrow 0$ exponentially fast as $k \rightarrow \infty$ if and only if $\varphi(k) \in \mathcal{PE}$. As a result, the objective (4) is achieved;
- P1.4. If $\tilde{\mathbf{\theta}}(k) \to 0$ as $k \to \infty$, then there exists an optimal value of γ for which the rate of convergence is maximum.

The proposition can be proved using the Lyapunov function

$$V(k) = \frac{1}{2\gamma} \tilde{\mathbf{\Theta}}^{\mathsf{T}}(k) \tilde{\mathbf{\Theta}}(k)$$
(15)

and evaluating its shift forward value V(k + 1) in view of (14) [19].

Unfortunately, the algorithm suffers from the drawback defined in Property P1.4 according to which the maximum rate of parametric convergence can be arbitrarily small and inacceptable from the practical point of view. In order to overcome this problem, in the next subsection we represent the adaptation algorithm with MRE that under the PE condition dramatically increases the rate of parametric convergence [18].

Scheme with MRE

We introduce an asymptotically stable positive transfer function of the form

$$L(z): = \prod_{i=1}^{M} \frac{1 - d_i}{z - d_i},$$
(16)

where $0 \le d_i \le 1$ are the constant design parameters, $M \in \mathcal{N}$. Premultiplying (9) by $\varphi(k)$ and then applying the operator L(z) we get the memory extended regression model

$$\mathbf{Y}(k) = \mathbf{\Omega}(k)\mathbf{\Theta} \tag{17}$$

with the memory extended output $\mathbf{Y}(k) = L(z)[\boldsymbol{\varphi}(k)y(k)]$ and the memory extended regressor matrix $\boldsymbol{\Omega}(k) = L(z)[\boldsymbol{\varphi}(k)\boldsymbol{\varphi}^{\mathsf{T}}(k)]$.

The regression model (17) allows us to design the adaptation algorithm with MRE

$$\hat{\mathbf{\theta}}(k+1) = \hat{\mathbf{\theta}}(k) + \gamma (\mathbf{I}_N + \gamma \mathbf{\Omega}(k))^{-1} \mathbf{E}(k), \qquad (18)$$

where $\mathbf{E}(k) = \mathbf{Y}(k) - \mathbf{\Omega}(k)\hat{\mathbf{\theta}}(k)$.

The algorithm (18) yields the parametric error model

$$\tilde{\boldsymbol{\theta}}(k+1) = (\mathbf{I}_N - \gamma(\mathbf{I}_N + \gamma \boldsymbol{\Omega}(k))^{-1} \boldsymbol{\Omega}(k)) \tilde{\boldsymbol{\theta}}(k)$$
(19)

which can be used for the proof of the following properties.

Proposition 2. In the closed-loop identification system, the algorithm (18) with φ defined in (10), y defined by (11), and L(z) given by (16) provides Properties P1.1, P1.2, P1.3 and:

- P2.1. If $\theta(k) \to 0$ as $k \to \infty$, then the rate of parametric convergence can be increased by increasing γ .

The proposition is proved via the Lyapunov function (15) and its shift forward value V(k + 1) in view of (19) [18].

Frequencies estimation

Replacing in (12) θ_i and c_i (i = 1, 2, ..., N) by their estimates θ_i generated by the algorithm of adaptation (13) or (18) and the estimates \hat{c}_i , respectively, we obtain the system of algebraic equations

$$\begin{cases} \hat{\theta}_{1} = \hat{c}_{1} + \hat{c}_{2} + \dots + \hat{c}_{N}, \\ \hat{\theta}_{2} = -(\hat{c}_{1}\hat{c}_{2} + \hat{c}_{1}\hat{c}_{3} + \dots + \hat{c}_{N-1}\hat{c}_{N}) \\ \vdots \\ \hat{\theta}_{N} = (-1)^{N+1}\hat{c}_{1}\hat{c}_{2} \dots \hat{c}_{N}. \end{cases}$$
(20)

Solving the latter, we get \hat{c}_i and taking into account Assumption 1 calculate the frequencies $\hat{\omega}_i(k)$ via the following equalities:

$$\hat{\omega}_i(k) = \frac{1}{T} \arccos\left(\hat{c}_i(k)\right). \tag{21}$$

Now, discuss the obtained result according to the properties defined by Propositions 1 and 2.

Remark 2.

- R2.1. As shown in the next section, the boundedness of $\varphi(k)$ can be provided via suitable design of input u(k);
- R2.2. According to the properties of discrete-time systems, for the adaptation algorithm with MRE (18) the rate of parametric convergence can be increased up to N steps;

- R2.3. The PE property of $\varphi(k)$ depends on the input signal u(k). As discussed in the next section, the certainty equivalent control law provides the PE condition for $x(t) \neq 0$ in the interval of excitation [k + 1, k + K] (see Definition 1) even if the reference signal is zero;
- R2.4. Transient solution of the system (20) always exists as a result of the Vieta's theorem, however can be complex. Therefore, in order to implement $\hat{\omega}_i(k)$ instead of (21), we use

$$\hat{\omega}_i(k) = \frac{1}{T} \arccos\left(Real\{\hat{c}_i(k)\}\right),\,$$

where $Real\{\cdot\}$ is the real part of a complex number.

In the next section, we demonstrate the application of the proposed identifier in the identification-based adaptive control of the plant (2).

Application: identification-based adaptive control

Problem statement

The problem is to design a control law that being applied to the plant (2) will ensure the boundedness of all the closed-loop signals in the system and:

O1. if $|x(k)| \ge x_0$ for all $k \ge N_0$ ($N_0 \in \mathcal{N}$), where $x_0 \in \mathbb{R}_+$ is a predefined threshold, then it will guarantee the limiting equality

$$\lim_{k \to \infty} \varepsilon(k) = \lim_{k \to \infty} (x_m(k) - x(k)) = 0$$
(22)

in which $x_m(k)$ is a bounded reference signal with known next step value $x_m(k+1)$;

O2. otherwise it will guarantee the limiting inequality

$$\lim_{k \to \infty} |\varepsilon(k)| \le \Delta, \tag{23}$$

where $\Delta \in \mathbb{R}_+$ is a maximum steady state error that is assumed to be reduced by changing design parameters.

Below, we represent the design of identificationbased adaptive controller for the plant (2). The key distinguishable property of this controller compared to the existing one from [16] consists in separate design of adaptive identifier tolerant to the state zero crossing and an adjustable controller. The problem solution is calculated in three steps:

- 1. Design of an observer of the TV parameter $\psi(k)$ required to recover some information about the parameter dynamics;
- 2. Design of an adjustable control law;
- 3. Recalculation of the estimates $\hat{\theta}$ given by the identifier obtained in the previous and its replacement in the control law designed.

TV parameter observer

In order to design a control law, it is needed to recover the information about the TV parameter dynamics. To this end, we, first, apply the IMP and represent the parameter $\psi(k)$ as the output of the exosystem given by the partial case of the canonical form recently presented in [16].

Lemma 2. The TV parameter can be represented as the output of the linear regression

$$\Psi(k) = \vartheta^{\mathsf{T}} \xi(k), \tag{24}$$

where $\xi(k) \in \mathbb{R}^{2N}$ is the state of the filter

$$\boldsymbol{\xi}(k+1) = \mathbf{G}\boldsymbol{\xi}(k) + \mathbf{I}\boldsymbol{\psi}(k) \tag{25}$$

with the initial condition $\xi(0) = [\psi(0), \psi(1), ..., \psi(2N-1)]$, the matrix

$$\mathbf{G} = \begin{bmatrix} \mathbf{O}_{(2N-1)\times 1} & \mathbf{I}_{(2N-1)} \\ 0 & \mathbf{O}_{1\times(2N-1)} \end{bmatrix}$$

m-dimensional vector $\mathbf{I} = [0, 0, ..., 0, 1]^{\mathsf{T}}$ selected so that the pair (**G**, **I**) is controllable, and the vector of unknown parameters $\boldsymbol{\vartheta} \in \mathbb{R}^{2N}$ given by

гл

$$\boldsymbol{\vartheta} = [\bar{\boldsymbol{\vartheta}}_1^{\mathsf{T}}, \bar{\boldsymbol{\vartheta}}_2^{\mathsf{T}}, \dots, \bar{\boldsymbol{\vartheta}}_N^{\mathsf{T}}]^{\mathsf{T}}$$
(26)

with

$$\bar{\boldsymbol{\vartheta}}_{1} = \begin{bmatrix} -1\\ \theta_{1} \end{bmatrix},$$

$$\bar{\boldsymbol{\vartheta}}_{j} = \begin{cases} \begin{bmatrix} j^{j-1} & j^{j-1} & j^{-j-1}\\ -C & +\sum & C & \theta_{2i} \\ N & i=1 & N-2i \\ \sum & C & \theta_{2i-1} \end{bmatrix} & \text{if } 2j \le N+1, \\ \begin{bmatrix} j^{-1} & N-2i+1 & j^{-j-1} \\ -C & +\sum & C & \theta_{2i} \\ N & i=1 & N-2i \\ \sum & C & \theta_{2i-1} \end{bmatrix} & \text{if } N+1 < 2j \le 2N, \end{cases}$$

$$(27)$$

 $j = 2, 3, ..., N, \theta_i$ are defined by the harmonics frequencies via (12).

The lemma is proved by defactorization of the discrete equation (8) shifted forward by 2*N* steps and introduction of the notations $\xi_i(k) = \psi(k - i + 1)(i = 1, 2, ..., 2N)$.

v are duplicated. Namely, it is seen from (26), (27) that

$$\boldsymbol{\vartheta} = \underbrace{[-1, \vartheta_2, \vartheta_3, \dots, \vartheta_{N+1}, \vartheta_N, \dots, \vartheta_2]^{\mathsf{T}}}_{2N \text{ elements}}.$$

This fact will allow us to simplify the recalculation from the estimates of θ_i given by (20) (dependent from the frequencies) into the estimates of ϑ used by the control law.

Remark 4. The function $\vartheta(\theta)$ is linear and can be represented in the compact form

$$\boldsymbol{\vartheta} = \mathbf{m} + \mathbf{M}\boldsymbol{\theta},\tag{28}$$

where

$$\mathbf{m} = \begin{bmatrix} \bar{\mathbf{m}}_1 \\ \bar{\mathbf{m}}_2 \\ \vdots \\ \bar{\mathbf{m}}_N \end{bmatrix} \in \mathbb{R}^{2N}, \mathbf{M} = \begin{bmatrix} \bar{\mathbf{M}}_1^T \\ \bar{\mathbf{M}}_2^T \\ \vdots \\ \bar{\mathbf{M}}_N^T \end{bmatrix} \in \mathbb{R}^{2N \times N}$$
(29)

are the vector and matrix given by

$$\bar{\mathbf{M}}_{j}^{\mathsf{T}} = \begin{cases} \begin{bmatrix} j-1 \\ -C \\ N \\ 0 \end{bmatrix} \text{ and} \\ \begin{bmatrix} j-1 & j-i-1 \\ \sum & C & \varsigma_{2i}^{\mathsf{T}} \\ j & j-i \\ \sum & C & \varsigma_{2i-1}^{\mathsf{T}} \\ \sum & C & \varsigma_{2i-1}^{\mathsf{T}} \\ \end{bmatrix} \text{ if } 2j \le N+1, \\ \begin{bmatrix} N-j+1 & j-i-1 \\ \sum & C & \varsigma_{2i}^{\mathsf{T}} \\ j & j-i \\ \sum & C & \varsigma_{2i}^{\mathsf{T}} \\ N-j+1 & j-i \\ \sum & C & \varsigma_{2i-1}^{\mathsf{T}} \\ \sum & C & \varsigma_{2i-1}^{\mathsf{T}} \\ N-j+1 & j-i \\ \sum & C & \varsigma_{2i-1}^{\mathsf{T}} \\ j & j \le N+1, \end{cases}$$
(30)

respectively, j = 1, 2, ..., N, $\varsigma_k = [0, ..., 0, 1, 0, ..., 0]^{\top}$ (k = 1, 2, ..., N) is the N-th dimensional coordinate vector.

Since the TV parameter $\psi(k)$ is unknown, the state ξ of the filter (25) is not measurable. However, it can be recovered via the observer recently proposed in [16].

Lemma 3. The observer

$$\hat{\xi}(k) = \zeta(k) + \mathbf{I} \frac{x(k-1)x(k)}{x^2(k-1) + \sigma^2(k-1)},$$
 (31)

$$\zeta(k+1) = \mathbf{G}\zeta(k) + \mathbf{GI} \frac{x(k-1)x(k)}{x^2(k-1) + \sigma^2(k-1)} - \mathbf{I} \frac{x(k)}{x^2(k) + \sigma^2(k)} u(k),$$
(32)

where $\hat{\xi}$ is the estimate of ξ ,

$$\sigma(k) = \begin{cases} 0 & \text{if } |x(k)| \ge x_0, \\ \sigma_0 & \text{otherwise,} \end{cases}$$
(33)

 x_0 , $\sigma_0 > 0$ are constant design parameters, provides the following properties:

1. For any initial conditions $\xi(0)$, $\zeta_j(0)$ the error $\epsilon(k) \triangleq \xi(k) - \hat{\xi}(k)$ satisfies the equality

$$\boldsymbol{\epsilon}(k) = \mathbf{G}^{k} \boldsymbol{\epsilon}(0) + (z\mathbf{I} - \mathbf{G})^{-1} \mathbf{I} \left[\frac{\sigma^{2}(k)}{x^{2}(k) + \sigma^{2}(k)} \boldsymbol{\psi}(k) \right]. \quad (34)$$

2. The norm $\|\epsilon(k)\|_{\infty}$ is bounded and can be reduced by reduction of the gain σ_0 and/or x_0 .

The equality (34) is proved by calculation of the next step value $\epsilon(k + 1) = \xi(k + 1) - \hat{\xi}(k + 1)$ in view of (25), (31), (32), and (2). The boundedness of $\|\epsilon(k)\|_{\infty}$ is followed from (34) and the boundedness of the function $\sigma^2(k)$

$$\overline{x^2(k) + \sigma^2(k)} \Psi(k)$$

Applying the result of Lemma 3, we represent the TV plant parameter in the form

$$\Psi(k) = \boldsymbol{\vartheta}^{\mathsf{T}} \hat{\boldsymbol{\xi}}(k) + \boldsymbol{\vartheta}^{\mathsf{T}} \boldsymbol{\epsilon}(k). \tag{35}$$

Now, using the parameterization (35), in the next section we propose the adaptive controller.

Control law

Replacing (35) in the plant equation (2), we obtain the expression suitable for controller design:

Научно-технический вестник информационных технологий, механики и оптики, 2025, том 25, № 2 Scientific and Technical Journal of Information Technologies, Mechanics and Optics, 2025, vol. 25, no 2

$$x(k+1) = \mathbf{\vartheta}^{\mathsf{T}} \hat{\boldsymbol{\xi}}(k) x(k) + u(k) + f(k), \qquad (36)$$

where $f(k) = \vartheta^{T} \epsilon(k) x(k)$ is considered as an additive disturbance satisfying the property defined by the following lemma.

Lemma 4. $f(k) \in \mathcal{L}_{\infty} \forall k \ge 0$.

Proof. Using the notion of \mathcal{L}_{∞} -norms of functions in view of (34), we have

$$||f(k)||_{\infty} - = \left\| \mathbf{\vartheta}^{\mathsf{T}} \left(\mathbf{G}^{k} \boldsymbol{\epsilon}(0) + (z\mathbf{I} - \mathbf{G})^{-1} \mathbf{I} \left[\frac{\sigma^{2}(k)}{x^{2}(k) + \sigma^{2}(k)} \boldsymbol{\psi}(k) \right] \right) \boldsymbol{x}(k) \right\|_{\infty} \leq \\ \leq ||\mathbf{\vartheta}||_{\infty} \left(||\mathbf{G}^{k} \boldsymbol{\epsilon}(0)||_{\infty} ||\boldsymbol{x}(k)||_{\infty} + \left\| \frac{\sigma^{2}(k)}{x^{2}(k) + \sigma^{2}(k)} \boldsymbol{\psi}(k) \right\|_{\infty} ||\boldsymbol{x}(k)||_{\infty} \right).$$

Here, we keep in mind that $||(z\mathbf{I} - \mathbf{G})^{-1}\mathbf{I}||_{\infty} = 1$ and

$$\left\| \frac{\sigma^2(k)}{x^2(k) + \sigma^2(k)} \psi(k) \right\|_{\infty} ||x(k)||_{\infty} \leq \frac{\sigma^2(k)}{\sqrt{x^2(k) + \sigma^2(k)}} ||\psi(k)||_{\infty}.$$

Since the matrix **G** is nilpotent and $\mathbf{G}^{2N} = \mathbf{O}_{2N \times 2N}$, while $||x(k)||_{\infty} < \infty$ for all k = 0, 1, ..., 2N, then

$$\|f(k)\|_{\infty} \le \frac{\sigma^{2}(k)}{\sqrt{x^{2}(k) + \sigma^{2}(k)}} \|\vartheta\|_{\infty} \|\psi(k)\|_{\infty} \quad \forall k \ge 2N.$$
(37)

As a result, $||f(k)||_{\infty}$ is bounded $\forall k \ge 0$. This completes the proof.

Using (36), we select the adjustable control law in the form

$$u(k) = -\hat{\boldsymbol{\vartheta}}^{\mathsf{T}}(k)\hat{\boldsymbol{\xi}}(k)x(k) - \lambda\varepsilon(k) + x_m(k+1), \qquad (38)$$

where $\lambda \in \mathbb{R}$ is a constant design parameter such that $|\lambda| < 1$, $\hat{\vartheta} \in \mathbb{R}^{2N}$ is the estimate of ϑ given by replacement of $\hat{\theta}_i$ (*i* = 1, 2, ..., *N*) instead of θ in (28):

$$\hat{\boldsymbol{\vartheta}} = \mathbf{m} + \mathbf{M}\hat{\boldsymbol{\theta}}.$$
 (39)

In the later, the elements of $\hat{\theta}$ are defined by the estimated frequencies of harmonics via (20) and generated by the adaptation algorithm (13) or (18).

Replacing (38) in (36) in view of (28) and (39), we obtain the error model

$$\varepsilon(k+1) = \lambda \varepsilon(k) + \tilde{\mathbf{\Theta}}^{\mathsf{T}}(k) \mathbf{M}^{\mathsf{T}} \hat{\boldsymbol{\xi}}(k) x(k) - f(k), \qquad (40)$$

where $\hat{\boldsymbol{\theta}} = \boldsymbol{\theta} - \hat{\boldsymbol{\theta}}$ is the identification error.

Now, using the error model (40) we are in position to formulate out main result.

Proposition 3. Under Assumption 1, the adaptive controller consisting of the adjustable control law (38), observer (31)–(33), identification algorithm (13) or (18), and the recalculation rule (29), (30), (39) being applied to the plant (2) ensures the following properties:

- P3.1 the boundedness of all the closed-loop signals in the system;
- P3.2 the objectives O1 and O2 according to the limits (22), (23);

- P3.3 $\hat{\mathbf{\theta}}(k) = \mathbf{\theta} - \hat{\mathbf{\theta}}(k)$ and $\omega_i - \hat{\omega}_i(k)$ tend to zero exponentially fast, and for the system closed by the identifier (18) the rate of the convergence $\hat{\mathbf{\vartheta}}(k) \rightarrow 0$.

Proof. Let us rewrite plant model (2) with the parameterization (24) (see Lemma 2) and recalculation formula (28) in the following form:

$$\overline{y} = \mathbf{\Theta}^{\mathsf{T}} \overline{\mathbf{\phi}},\tag{41}$$

where $\bar{y} = x(k) - u(k-1) - \mathbf{m}^{\mathsf{T}} \xi(k-1)x(k-1)$, $\bar{\mathbf{\phi}} = \mathbf{M}^{\mathsf{T}} \xi(k)x(k)$. Using these notations and the notation of f(k), the model (40) can be represented as

$$\varepsilon(k+1) = \lambda \varepsilon(k) - \bar{e}(k) - \hat{\mathbf{\theta}}^{\mathsf{T}}(k) \mathbf{M}^{\mathsf{T}} \boldsymbol{\epsilon}(k) \mathbf{x}(k), \qquad (42)$$

where $\bar{e} \triangleq \bar{y} - \hat{\theta}\bar{\varphi} = \tilde{\theta}^{\mathsf{T}}\bar{\varphi}$ is the identification error.

Using the arguments of Lemma 4 we can see that the signal $\hat{\theta}^{T} \mathbf{M} \epsilon x$ in (42) is bounded and satisfies the inequality

$$\|\hat{\boldsymbol{\theta}}^{\mathsf{T}}\mathbf{M}^{\mathsf{T}}\boldsymbol{\epsilon}\boldsymbol{x}(k)\|_{\infty} \leq \frac{\sigma^{2}(k)}{\sqrt{x^{2}(k) + \sigma^{2}(k)}} \|\hat{\boldsymbol{\theta}}^{\mathsf{T}}\mathbf{M}^{\mathsf{T}}\|_{\infty} \|\boldsymbol{\psi}(k)\|_{\infty}$$
$$\forall k \geq 2N, \tag{43}$$

where $\sigma(k)$ is given by (33).

For the regression (9), in accordance with Properties P1.1 and P1.2 of Propositions 1 and 2, both estimation schemes (13) and (18) ensure that $e(k) = y(k) - \hat{\theta}(k)\varphi(k) = \tilde{\theta}^{\top}(k)\varphi(k)$, $\hat{\theta}(k+1) - \hat{\theta}(k) \in \mathcal{L}_{\infty} \cap \mathcal{L}_2$ and e(k), $\hat{\theta}(k+1) - \hat{\theta}(k) \rightarrow 0$ as $k \rightarrow \infty$. Since the regressions (9) and (41) representing the same plant model have the same structure, then the estimators (13) and (18) additionally ensure that $\bar{e}(k) = \tilde{\theta}^{\top}(k)\bar{\varphi}(k) \in \mathcal{L}_{\infty} \cap \mathcal{L}_2$ and $\bar{e}(k) \rightarrow 0$ as $k \rightarrow \infty$.

Since according to Propositions 1 and 2, $\hat{\theta}(k) \in \mathcal{L}_{\infty}$ and, as shown above, the inequality (43) holds, then $\varepsilon(k) \in \mathcal{L}_{\infty}$. As a result, x(k), $u(k) \in \mathcal{L}_{\infty}$. This completes the proof of Property P3.1.

If $|x(k)| \ge x_0$ for all $k \ge N_0$ ($N_0 \in \mathcal{N}$), then $\sigma(k) = 0$, $\epsilon(k) = \xi(k) - \xi(k) \rightarrow 0$ as $k \rightarrow \infty$. Therefore, $\epsilon(k) \rightarrow 0$ as $k \rightarrow \infty$ according to (42). Otherwise, $\epsilon(k)$, $f(k) \in \mathcal{L}_{\infty}$ according to (23). As seen from (37), Δ can be reduced by decreasing $\sigma(\sigma_0)$. This proves Property P3.2.

To prove Property P3.3, we take into account the equivalence of the regressors $\varphi(k)$ and $\overline{\varphi}$ for $x(k) \neq 0$. Since the generator (25), (24) is not redundant due to the knowledge of the number of harmonics N, $\xi(k) \in \mathcal{PE}$. As a result, $\varphi, \overline{\varphi} \in \mathcal{PE}$ for $x(k) \neq 0$. Therefore, Property P3.3 is provided by the estimators (13) and (18) according to Propositions 1 and 2. This completes the proof of the proposition.

Now concluding the main result, we make the following remarks:

Remark 5.

- R5.1. As seen from the proof of Proposition 3, the problem objectives O1 and O2 do not require exact identification of the harmonics frequencies ω_i , however require convergence $\bar{e}(t) \rightarrow 0$. At the same time, since the exact number of harmonics is known, then the regressor $\varphi \in \mathcal{PE}$, and the parametric errors $\hat{\theta}(t)$ as well as the frequencies estimation errors converge to zero exponentially fast. As a result, the control error

 $\varepsilon(t)$ converge to zero according to objective O1 or to the residual set according to objective O2 exponentially fast;

— R5.2. As seen from the structure of the plant (2), if the state x(k) crosses zero, then the state $\xi(k)$ of the TV parameter model (25), (24) is no longer to be observable. In order to overcome this problem, the threshold x_0 is suitably involved into the observer. At the same time, in contrast to the method of direct adaptive control recently reported in [16], the proposed algorithm of frequencies estimation is independent from the value of x(k) and the property of observability.

Simulation results

Consider the plant (2) with the uncertain time-varying parameter $\psi(k) = a_1 \sin(\omega_1 kT + \varphi_1) + a_2 \sin(\omega_2 kT + \varphi_2)$, in which $a_1, a_2, \omega_1, \omega_2, \varphi_1$, and φ_2 are unknown constants set to the following values: $a_1 = 0.2$; $a_2 = 0.3$; $\omega_1 = 2$; $\omega_2 = 3$; $\varphi_1 = 3$; and $\varphi_2 = 4$. The interval of discretization *T* is set to 0.1 (sec).

The problem is twofold:

— calculate the estimates $\hat{\omega}_i(k)$ (i = 1, 2) such that $\lim_{k \to \infty} (\omega_i - \hat{\omega}_i(k)) = 0;$

- design a control law ensuring the boundedness of all the closed-loop signals and driving the tracking error $\varepsilon(k) = x_m(k) - x(k)$ to zero, where $x_m(k)$ is generated by the reference model $x_m(k+1) = 2 + \sin(2kT)$. In the simulation, we select the estimator with MRE

(18), for which we select the first-order filter $L(z) = \frac{0.3}{z - 0.7}$

For the TV parameter observer, we select

$$\mathbf{G} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \ \mathbf{I} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}$$

 $x_0 = 0.01$, and $\sigma_0 = 0.01$, while for the control law given by (38) we choose $\lambda = 0.1$.

All initial conditions in the scheme are set to zero.

The simulation results are presented in Fig. 1 and Fig. 2 and demonstrate the convergence of the frequencies estimates $\hat{\omega}_i$ (*i* = 1, 2) to the original values ω_i and, as a result, the convergence of the tracking error to zero. As shown by Fig. 2, the rate of estimation and, as a result, the rate of reference tracking can be increased by increasing the coefficient gamma.



Fig. 1. Transients in the system closed by the estimator with MRE (18) with $\gamma = 100$. Transients of: the state x(k) and the reference model output $x_m(k)$ (*a*); the tracking error $\varepsilon(k)$ (*b*); the control signal u(k) (*c*)



Fig. 2. Response frequency estimates $\hat{\omega}_1(a)$, $\hat{\omega}_2(b)$ and the tracking error $\varepsilon(k)(c)$ in the system closed by the estimator with MRE (18) for different γ

Conclusion

In the paper, two adaptive frequencies estimators for the first order linear discrete system with uncertain multisinusoidal TV parameter are designed. The first estimator is based on the gradient scheme of adaptation, while the second one uses the discrete version of the adaptation algorithm with MRE. Then, the indirect (identification-based) adaptive control laws using these estimators are proposed, and their properties are analyzed.

The estimation algorithms together with the control laws are supposed to be developed toward the following axes:

— solution to the problem for extended class of plants [16]
$$x_{i}(k+1) = \psi_{i}(k)x_{i}(k) + x_{i+1}(k),$$

$$x_{n}(k+1) = \psi_{n}(k)x_{n}(k) + u(k),$$

$$y(k) = x_{1}(k),$$

where $i = 1, 2, ..., n - 1, x_j \in \mathbb{R}$ (j = 1, 2, ..., n) are the elements of the state vector $\mathbf{x} = [x_1, x_2, ..., x_n]^T$ with the initial condition $x_j(0), y \in \mathbb{R}$ is the plant output, $u(k) \in \mathbb{R}$ is the control signal, $\psi_j(k)$ are the unknown parameters given by the functions $\psi_j(k) = \sum_{i=1}^{N} a_{j,i} \sin(\omega_{j,i}kT + \varphi_{j,i})$, with

References

- Francis B.A., Wonham W.M. The internal model principle for linear multivariable regulators. *Applied Mathematics and Optimization*, 1975, vol. 2, no. 2, pp. 170–194. https://doi.org/10.1007/bf01447855
- Nikiforov V., Gerasimov D. Adaptive regulation in systems with unknown parameters. In: adaptive regulation. *Lecture Notes in Control and Information Sciences*, 2022, vol. 491, pp. 223–265. https://doi.org/10.1007/978-3-030-96091-9_5
- Abidi K. Spatial periodic adaptive control approach for rotary systems in sampled time. *International Journal of Robust and Nonlinear Control*, 2014, vol. 24, no. 7, pp. 1177–1188. https://doi.org/10.1002/ rnc.2931
- Chulaevsky V. Almost Periodic Operators and Related Nonlinear Integrable Systems (Nonlinear Science Theory and Applications). Manchester University Press, 1989, 105 p.
- Yakubovich V.A., Starzhinskii V.M. Linear Differential Equations with Periodic Coefficients. Wiley, 1975, 839 p.
- Ruby L. Applications of the Mathieu equation. *American Journal of Physics*, 1996, vol. 64, no. 1, pp. 39–44. https://doi.org/10.1119/1.18290
- Ahn H.S., Chen Y. Time periodical adaptive friction compensation. Proc. of the IEEE International Conference on Robotics and Biomimetics, 2004, pp. 362–367. https://doi.org/10.1109/ ROBIO.2004.1521805
- Glower J.S. MRAC for systems with sinusoidal parameters. International Journal of Adaptive Control and Signal Processing, 1996, vol. 10, no. 1. pp. 85–92. https://doi.org/10.1002/(SICI)1099-1115(199601)10:1<85::AID-ACS388>3.0.CO;2-0
- Narendra K., Esfandiari K. Adaptive control of linear periodic systems using multiple models. *Proc. of the IEEE Conference on Decision and Control (CDC)*, 2018, pp. 589–594. https://doi. org/10.1109/CDC.2018.8619514
- Xu J.-X. A new periodic adaptive control approach for time-varying parameters with known periodicity. *IEEE Transactions on Automatic Control*, 2004, vol. 49, no. 4, pp. 579–583. https://doi.org/10.1109/ TAC.2004.825612
- Zhu S., Sun M.X. Robust adaptive repetitive control for a class of nonlinear periodically time-varying systems. *International Journal* of Control, 2022, vol. 95, no. 1, pp. 187–196. https://doi.org/10.1080 /00207179.2020.1786767
- Pang B., Jiang Z.-P., Mareels I. Reinforcement learning for adaptive optimal control of continuous-time linear periodic systems. *Automatica*, 2020, vol. 118, pp. 109035. https://doi.org/10.1016/j. automatica.2020.109035
- Yu M., Huang D. A switching periodic adaptive control approach for time-varying parameters with unknown periodicity. *International Journal of Adaptive Control and Signal Processing*, 2015, vol. 29, no. 12, pp. 1526–1538. https://doi.org/10.1002/acs.2560
- Kozachek O., Bobtsov A., Nikolaev N. Adaptive observer for a nonlinear system with partially unknown state matrix and delayed measurements. *IFAC-PapersOnLine*, 2023, vol. 56, no. 2, pp. 8702– 8707. https://doi.org/10.1016/j.ifacol.2023.10.051
- Gerasimov D., Popov A., Hien N.D., Nikiforov V. Adaptive control of LTV systems with uncertain periodic coefficients. *IFAC-PapersOnLine*, 2023, vol. 56, no. 2, pp. 9185–9190. https://doi. org/10.1016/j.ifacol.2023.10.160
- Gerasimov D., Hien N.D., Nikiforov V.O. Direct adaptive control of LTV discrete-time systems with uncertain periodic coefficients. *Proc.*

- unknown constants $a_{j,i}$, $\varphi_{j,i}$ and $\omega_{j,i}$. The maximum numbers of harmonics N_i are assumed known.
 - solution to the problem for continuous time systems of the form

$$\dot{x} = \psi(t)x + u$$

where $\psi(t) = \sum_{i=1}^{N} a_i \sin(\omega_i kT + \varphi_i)$ is the uncertain TV

parameter with unknown constant a, ω and φ .

Литература

- Francis B.A., Wonham W.M. The internal model principle for linear multivariable regulators // Applied Mathematics and Optimization. 1975. V. 2. N 2. P. 170–194. https://doi.org/10.1007/bf01447855
- Nikiforov V., Gerasimov D. Adaptive regulation in systems with unknown parameters. In: adaptive regulation // Lecture Notes in Control and Information Sciences. 2022. V. 491. P. 223–265. https:// doi.org/10.1007/978-3-030-96091-9_5
- Abidi K. Spatial periodic adaptive control approach for rotary systems in sampled time // International Journal of Robust and Nonlinear Control. 2014. V. 24. N 7. P. 1177–1188. https://doi.org/10.1002/ rnc.2931
- Chulaevsky V. Almost Periodic Operators and Related Nonlinear Integrable Systems (Nonlinear Science Theory and Applications). Manchester University Press, 1989. 105 p.
- 5. Yakubovich V.A., Starzhinskii V.M. Linear Differential Equations with Periodic Coefficients. Wiley, 1975. 839 p.
- Ruby L. Applications of the Mathieu equation // American Journal of Physics. 1996. V. 64. N 1. P. 39–44. https://doi.org/10.1119/1.18290
- Ahn H.S., Chen Y. Time periodical adaptive friction compensation // Proc. of the IEEE International Conference on Robotics and Biomimetics. 2004. P. 362–367. https://doi.org/10.1109/ ROBIO.2004.1521805
- Glower J.S. MRAC for systems with sinusoidal parameters // International Journal of Adaptive Control and Signal Processing. 1996. V. 10. N 1. P. 85–92. https://doi.org/10.1002/(SICI)1099-1115(199601)10:1<85::AID-ACS388>3.0.CO;2-0
- Narendra K., Esfandiari K. Adaptive control of linear periodic systems using multiple models // Proc. of the IEEE Conference on Decision and Control (CDC). 2018. P. 589–594. https://doi. org/10.1109/CDC.2018.8619514
- Xu J.-X. A new periodic adaptive control approach for time-varying parameters with known periodicity // IEEE Transactions on Automatic Control. 2004. V. 49. N 4. P. 579–583. https://doi.org/10.1109/ TAC.2004.825612
- Zhu S., Sun M.X. Robust adaptive repetitive control for a class of nonlinear periodically time-varying systems // International Journal of Control. 2022. V. 95. N 1. P. 187–196. https://doi.org/10.1080/00 207179.2020.1786767
- Pang B., Jiang Z.-P., Mareels I. Reinforcement learning for adaptive optimal control of continuous-time linear periodic systems // Automatica. 2020. V. 118. P. 109035. https://doi.org/10.1016/j. automatica.2020.109035
- Yu M., Huang D. A switching periodic adaptive control approach for time-varying parameters with unknown periodicity // International Journal of Adaptive Control and Signal Processing. 2015. V. 29. N 12. P. 1526–1538. https://doi.org/10.1002/acs.2560
- Kozachek O., Bobtsov A., Nikolaev N. Adaptive observer for a nonlinear system with partially unknown state matrix and delayed measurements // IFAC-PapersOnLine. 2023. V. 56. N 2. P. 8702– 8707. https://doi.org/10.1016/j.ifacol.2023.10.051
- Gerasimov D., Popov A., Hien N.D., Nikiforov V. Adaptive control of LTV systems with uncertain periodic coefficients // IFAC-PapersOnLine. 2023. V. 56. N 2. P. 9185–9190. https://doi. org/10.1016/j.ifacol.2023.10.160
- Gerasimov D., Hien N.D., Nikiforov V.O. Direct adaptive control of LTV discrete-time systems with uncertain periodic coefficients // Proc. of the IEEE 63rd Conference on Decision and Control (CDC).

of the IEEE 63rd Conference on Decision and Control (CDC), 2024, pp. 4303–4308. https://doi.org/10.1109/CDC56724.2024.10886757

- Gerasimov D.N., Belyaev M.E., Nikiforov V.O. Performance improvement of discrete MRAC by dynamic and memory regressor extension. *Proc. of the 18th European Control Conference (ECC)*, 2019, pp. 2950–2956. https://doi.org/10.23919/ecc.2019.8795874
- Gerasimov D.N., Belyaev M.E., Nikiforov V.O. Improvement of transient performance in MRAC by memory regressor extension. *European Journal of Control*, 2021, vol. 59. pp. 264–273. https://doi. org/10.1016/j.ejcon.2020.10.002
- Goodwin G., Sin K. Adaptive Filtering Prediction and Control. Prentice-Hall, 1984, 540 p.
- 20. Tao G. Adaptive Control Design and Analysis. John Wiley & Sons, 2003, 640 p.

Authors

Dang Hien Ngo — PhD Student, ITMO University, Saint Petersburg, 197101, Russian Federation, sc 58069032500, https://orcid.org/0000-0001-6673-921X, danghien719@gmail.com

Dmitry N. Gerasimov — PhD, Associate Professor, ITMO University, Saint Petersburg, 197101, Russian Federation, sc 36637147000, https:// orcid.org/0000-0001-8306-4138, dngerasimov@itmo.ru

Received 10.01.2025 Approved after reviewing 04.02.2025 Accepted 20.03.2025



2024. P. 4303-4308. https://doi.org/10.1109/ CDC56724.2024.10886757

- Gerasimov D.N., Belyaev M.E., Nikiforov V.O. Performance improvement of discrete MRAC by dynamic and memory regressor extension // Proc. of the 18th European Control Conference (ECC). 2019. P. 2950–2956. https://doi.org/10.23919/ecc.2019.8795874
- Gerasimov D.N., Belyaev M.E., Nikiforov V.O. Improvement of transient performance in MRAC by memory regressor extension // European Journal of Control. 2021. V. 59. P. 264–273. https://doi. org/10.1016/j.ejcon.2020.10.002
- Goodwin G., Sin K. Adaptive Filtering Prediction and Control. Prentice-Hall, 1984. 540 p.
- Tao G. Adaptive Control Design and Analysis. John Wiley & Sons, 2003. 640 p.

Авторы

Нго Данг Хиен — аспирант, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, sc 58069032500, https://orcid.org/0000-0001-6673-921X, danghien719@gmail.com

Герасимов Дмитрий Николаевич — кандидат технических наук, доцент, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, вс 36637147000, https://orcid.org/0000-0001-8306-4138, dngerasimov@itmo.ru

Статья поступила в редакцию 10.01.2025 Одобрена после рецензирования 04.02.2025 Принята к печати 20.03.2025

Работа доступна по лицензии Creative Commons «Attribution-NonCommercial» **I/İTMO**

НАУЧНО-ТЕХНИЧЕСКИЙ ВЕСТНИК ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ, МЕХАНИКИ И ОПТИКИ март–апрель 2025 Том 25 № 2 http://ntv.ifmo.ru/ SCIENTIFIC AND TECHNICAL JOURNAL OF INFORMATION TECHNOLOGIES, MECHANICS AND OPTICS March–April 2025 Vol. 25 No 2 http://ntv.ifmo.ru/en/ ISSN 2226-1494 (print) ISSN 2500-0373 (online)

ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ, МЕХАНИКИ И ОПТИКИ

КОМПЬЮТЕРНЫЕ СИСТЕМЫ И ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ COMPUTER SCIENCE

doi: 10.17586/2226-1494-2025-25-2-253-260 УДК 004.032.26

Разработка файловой системы для хранения данных интеллектуальной системы видеонаблюдения

Алексей Николаевич Субботин¹, Наталия Александровна Жукова²

¹ Санкт-Петербургский государственный электротехнический университет «ЛЭТИ» им. В.И. Ульянова (Ленина), Санкт-Петербург, 197022, Российская Федерация

¹ ГУП «Петербургский метрополитен», Санкт-Петербург, 190013, Российская Федерация

² Санкт-Петербургский федеральный исследовательский центр Российской академии наук (СПб ФИЦ РАН), Санкт-Петербург, 199178, Российская Федерация

¹ alesu1543@gmail.com^{\Box}, https://orcid.org/0000-0002-4823-6288

² nazhukova@mail.ru, https://orcid.org/0000-0001-5877-4461

Аннотация

Введение. Рассмотрена проблема создания файловой системы с характеристиками, отличными от универсальных, для хранения данных интеллектуальных систем видеонаблюдения. Доступ к файловой системе является определяющим фактором, от которого зависит быстродействие всей системы. Скорость выполнения операций с данными определяется не только быстротой работы шины данных и наличием современного процессора, но и драйвером доступа к жесткому диску, который может ограничивать возможности системы выполнять основные функции: наблюдение, анализ изображений, определение образов и событий. Существует потребность в использовании более производительного сервера, что требует дополнительных материальных затрат, в разработке специализированной файловой системы для повышения скорости записи и чтения с жесткого диска. Использование специализированной файловой системы, ориентированной на решение одной или ограниченного числа задач, может значительно повысить скорость работы системы при использовании серверов с одинаковыми техническими характеристиками. В интеллектуальных системах видеонаблюдения применение специализированной файловой системы может обеспечить повышение скорости обработки изображений и точности определения объектов в видеопотоке благодаря повышению скорости записи и чтения данных с диска. Анализ существующих файловых систем показал, что имеющиеся решения не позволяют обеспечить требуемую скорость работы с данными в интеллектуальных системах видеонаблюдения при наличии ограничений на количество и мощность используемых технических средств. Предложена специализированная файловая система для хранения данных систем интеллектуального видеонаблюдения. Метод. Разработана файловая система с описанием, ориентированная на решение задачи хранения данных в системах интеллектуального наблюдения. Используемая база данных обладает функциями чтения, поиска, записи и обновления структурированных данных, размещенных в ее таблицах. База данных оптимизирована для работы с данными интеллектуальных систем видеонаблюдения, имеет ограниченную размерность столбцов, задаваемую в соответствии с размещаемыми в ней сущностями. Особенность база данных состоит в том, что она постоянно находится в оперативной памяти, синхронизация данных с жестким диском выполняется через заданный интервал времени. Примененная база данных, подобно Redis, работает значительно быстрее традиционных. Разработанный драйвер действует напрямую с жестким диском, не использует функции операционной системы, что повышает скорость работы с данными. Основные результаты. Сравнение скорости записи и чтения данных при использовании разработанного и существующих универсальных драйверов показало, что применение нового драйвера позволяет повысить скорость записи и чтения на 43,4 % относительно New Technology File System (NTFS). Обсуждение. В рамках проведенного исследования выполнена разработка файловой системы для интеллектуальных систем видеонаблюдения. Отмечено, что подобные специализированные файловые системы могут разрабатываться для применения в других областях, где требуется повысить скорость (снизить время) записи и чтения данных с диска.

Ключевые слова

файловая система, интеллектуальная система видеонаблюдения, хранение данных, взаимодействие с объектной базой данных, снижение времени доступа к данным, база данных

© Субботин А.Н., Жукова Н.А., 2025

Благодарности

Исследование выполнено при финансовой поддержке государственного бюджета, номер проекта № FFZF-2025-0019.

Ссылка для цитирования: Субботин А.Н., Жукова Н.А. Разработка файловой системы для хранения данных интеллектуальной системы видеонаблюдения // Научно-технический вестник информационных технологий, механики и оптики. 2025. Т. 25, № 2. С. 253–260. doi: 10.17586/2226-1494-2025-25-2-253-260

Development of a file system for storing data of an intelligent video surveillance system

Alexey N. Subbotin¹[∞], Nataly A. Zhukova²

¹ Saint Petersburg Electrotechnical University "LETI", Saint Petersburg, 197022, Russian Federation

¹ State Unitary Enterprise "Petersburg Metropolitan", Saint Petersburg, 190013, Russian Federation

² St. Petersburg Federal Research Center of the Russian Academy of Sciences, Saint Petersburg, 199178, Russian Federation

¹ alesu1543@gmail.com^{\Box}, https://orcid.org/0000-0002-4823-6288

² nazhukova@mail.ru, https://orcid.org/0000-0001-5877-4461

Abstract

The article considers the problem of creating a file system with characteristics different from universal ones for storing data of intelligent video surveillance systems. Access to the file system is a determining factor on which the performance of the entire system depends. A fast data bus and a modern processor do not always determine the speed of data operations, but also the hard disk access driver which, accordingly, can limit the system ability to perform basic functions: surveillance, image analysis, detection of images and events. It is necessary to select a more productive server which is expensive, or use a specialized driver to increase the speed of writing and reading on the hard disk. The use of a specialized file system focused on solving one or a limited number of problems can significantly increase the speed of systems in cases where the server is used with the same technical characteristics. In intelligent video surveillance systems, the use of a specialized file system can provide an increase in the speed of image processing and the accuracy of object detection in the video stream, due to the increased speed of reading and writing from the disk. An analysis of existing file systems has shown that the existing solutions do not provide the required speed of working with data in intelligent video surveillance systems when using technical means with the same computing characteristics. In this article, the authors propose a specialized file system for storing data in intelligent video surveillance systems. A file system has been developed that is focused on solving one problem: storing data in intelligent surveillance systems. The developed driver increases the speed of accessing the data on the hard drive. The new file system for storing data in an intelligent video surveillance system works together with a database for one, separate task. A comparison of the speed of writing and reading data using the developed driver and using existing universal drivers made. As a result of the comparison, it has been established that the use of the new driver has increased the speed of writing and reading by 43.4 % relative to NTFS file system. As part of the study, a file system for intelligent video surveillance systems was developed, but similar specialized file systems can be developed for use in other areas where it is necessary to increase the speed (reduce the time) of writing and reading data from the file system.

Keywords

file system creation, intelligent video surveillance system, data storage, interaction with object database, reduction of data access time, database, file system

Acknowledgements

This work was supported by the state budget, project No. FFZF-2025-0019.

For citation: Subbotin A.N., Zhukova N.A. Development of a file system for storing data of an intelligent video surveillance system. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 2025, vol. 25, no. 2, pp. 253–260 (in Russian). doi: 10.17586/2226-1494-2025-25-2-253-260

Введение

Хранение последовательности файлов, как единого потока данных, приводит к значительным временным затратам при перезаписи информации и добавлении новых данных в ранее созданные файлы. Исходя из этого, все современные файловые системы поддерживают кластеризацию, которая предусматривает условное разделение на фрагменты дискового пространства для быстрого добавления новых данных в файлы. Первые сектора жесткого диска заняты таблицей с информацией о файлах. Информация об одном файле занимает 1 КБ в таблице New Technology File System (NTFS), где содержится дата и время файла, занимаемые кластеры, длинное имя и другая служебная информация. Изначально драйвер для работы с файловой системой был отдельной программой, поставляемой с устройством многократной записи. Но в последствии корпорацией IBM был создан универсальный драйвер FAT для всех жестких дисков и интегрирован в ядро операционной системы DOS (Disk Operating System). Такой подход применяется и в ядре Linux, где интегрированы файловые системы: Ext2, Ext3, Ext4, JFS, ReiserFS, XFS, Btrfs. Практически все современные операционные системы: Windows, FreeBSD, macOS, Linux, OpenSolaris, QNX и др. считают драйвер файловой системы частью ядра. Такой подход необходим для чтения служебных файлов операционной системы после передачи управления из сектора загрузчика в сектор операционной системы. После команды от загрузчика

система проверяет базовые порты записи/чтения (Input/ Output, I/O) по спецификации процессора x86, загружает драйвера файловой системы, читает свои файлы, переопределяет порты І/О и проверяет работоспособность оборудования [1, 2]. Далее проверяется допустимость загрузки по техническим характеристикам: частота процессора, размер оперативной памяти и жесткого диска. Файловая система имеет большое значение на этапе загрузки операционной системы. Однако нестандартные файловые системы распространены и в наше время, поскольку современные файловые системы перегружены дополнительными функциями и сильно уступают простому бинарному чтению в стеке с ХХХ по ҮҮҮ адреса. Не всегда необходимо журналирование и разбиение файлов на кластеры, а таблица разметки диска может быть реализована в самом драйвере (программе для доступа к данным в одном разделе). Раздел с любой файловой системой, созданной Linux, Windows, FreeBSD можно просто удалить, а данные читать своей программой с адреса сепарации до конечного адреса неразмеченного пространства. По такому принципу работают программы Norton Recovery Files¹, которые восстанавливают удаленные файлы из корзины Windows.

В представленной работе показаны пути повышения скорости чтения и записи данных в интеллектуальной системе видеонаблюдения (ИСВН) за счет использования нестандартной файловой системы. Это требует решения таких задач, как разработка файловой системы для хранения данных, драйвера доступа к файловой системе, который за счет непосредственной работы с жестким диском позволит повысить скорость работы. Проведена экспериментальная оценка скорости доступа к данным.

Рассмотрение аналогов

Исследования, направленные на создание файловой системы в Computer Science, ведутся с 60-х годов XX века, когда появилась первая операционная система GM-HAA и Multics-MIT. Так, многие трояны и шифровальщики по данным Лаборатории Касперского² используют неразмеченное дисковое пространство и создают свои скрытые разделы для хранения файлов. Но данный подход может быть применен и на пользу. Например, брендовые компьютеры (Acer, Asus, Dell, Apple и пр.) применяют скрытый файловый раздел в начале диска для восстановления оригинальной версии операционной системы со всеми драйверами, настройками производителя и фирменными программами [3-5]. Раздел не виден операционной системе, но он может быть прочитан специальной программой с адреса дискового пространства XXX до YYY. Однако вирусы научились читать этот восстановочный раздел и дописывать туда свой код и подменять файлы. После восстановления изначальная копия операционной системы уже содержит свежие вирусы.

Одним из примеров систем, в которых используются специализированные файловые системы, являются системы, разрабатываемые для технологических встраиваемых компьютеров, где сильно ограничены вычислительные ресурсы (Raspberry Pi, ASUS Tinker Board S, ClockworkPi, Arduino Mega 2560 и пр.). Отсутствие мощного процессора и современного чипсета не позволяет быстро передавать данные между жестким диском и регистрами процессора, что необходимо для выполнения таких функций, как просмотр видео. В камерах видеонаблюдения Axis³ есть встроенное программное обеспечение P-Iris, где тоже используется специальная файловая система для повышения скорости работы с данными через процессоры: ARTPEC-3 и ARTPEC-B.

В настоящее время специализированные файловые системы используются в программных системах VMware, где применяется специализированная кластерная файловая система VMFS для хранения данных виртуальных машин. Широко распространена файловая система для хранения данных на оптических дисках CDfs, файловая система Cramfs, обеспечивающая возможность сжатия данных при записи в операционной системе GNU/Linux. Как сервисная функция используется менеджер логических томов и файловая система ZFS от Sun Microsystems без фрагментации с изменением томов как у NTFS.

Анализ широкого круга существующих ИСВН, показал, что в них не используются специализированные файловые системы. В процессе анализа рассматривались такие популярные системы видеонаблюдения как Видеоинтеллект, ISS, МТС Облачное видеонаблюдение, Ситроникс, Бевард, Ростелеком, Максима и многие др., предоставляющие широкие возможности по распознаванию образов с использованием различных профилей для определения номеров машин на автостоянке и других профильных объектов, например, детского садика, школы и т. д. Все средства предоставляют возможности для динамического анализа видеоряда и позволяют выполнять аналитику предиктивного характера. В имеющихся ИСВН (например, от производителя IVS) обеспечивается сохранение статистической информации на видеосервер, для доступа к данным используются драйвера по умолчанию, поддерживаемые операционной системой NTFS. Это приводит к замедлению обращений к диску почти в два раза. До настоящего времени специализированные файловые системы и драйвера для ИСВН не создавались, поскольку объемы обрабатываемых данных были меньше и возможностей имеющихся технических средств было достаточно для их обработки. Создание специализированной файловой системы позволит обеспечить повышение скорости (снижения времени) доступа к данным.

Определение проблемы

ИСВН характеризуются высокой интенсивностью чтения и записи с дискового хранилища, которое может быть представлено как озеро данных или как диско-

¹ [Электронный ресурс]. https://www.norton.com/ (дата обращения: 22.03.2025).

² [Электронный ресурс]. https://www.kaspersky.ru/ (дата обращения: 22.03.2025).

³ [Электронный ресурс]. https://www.axis.com (дата обращения: 22.03.2025).

вое пространство на сервере в облачной платформе. Данные от ИСВН содержат как статистическую информацию, так и видеоряд, по которому можно выявлять объекты и события и их классифицировать. К статистической информации относится: дата и время работы ИСВН, наблюдаемые объекты (цеха, столовые, коридорные и другие помещения общего назначения), место расположения видеокамер, их характеристики, погодные (туман, освещенность, время дня и ночи и пр.) и температурные условия. Хранимая информация используется операторами ИСВН на устройствах визуализации событий и просмотра хроники для определения чрезвычайных ситуаций на объектах (метрополитен, стадион, аэропорт, магазин, детский садик, школа и др.) в зависимости от профиля деятельности учреждения и поставленных задач перед отделом безопасности (мелкое хулиганство, воровство, террористическая угроза и пр.). В таких системах доступ к данным требует много времени (скорость доступа к данным варьируется от 10 до 24 кадров/с с фрагментами видео и статистической информацией в зависимости от настроек системы).

Современные видеосерверы позволяют выполнять значительное количество операций в секунду, но для повышения их производительности требуется изменение технических характеристик и больших затрат. Разработка специализированной программы (драйвера) для работы с данными для определенной задачи, в частности, интеллектуального видеонаблюдения может позволить значительно повысить скорость доступа к данным на жестком диске (рис. 1). При разработке специализированной файловой системы необходимо точно определить размерностью файлов, которые будут храниться. Файловая система должна быть создана как логическое продолжение базы данных для ИСВН [6, 7].

Специализированный драйвер должен разрабатываться как часть системы, взаимодействующей со своей базой данных в виде объектов, которые считываются драйвером с жесткого диска. При этом дальнейшая работа с данными ИСВН осуществляется в оперативной памяти видеосервера.



Рис. 1. Место файловой системы на видеосервере в интеллектуальной системе видеонаблюдения.

IoT — Internet of Things

Fig. 1. The place of the file system on the video server in the intelligent video surveillance system

Постановка задачи

Современные файловые системы, такие как FAT32, NTFS, EXT4 и многие другие наделены огромным количеством функций и большой универсальностью, позволяющей решать любые задачи. Однако если отказаться от универсальности, а разработать драйвер (программу) файловой системы под одну задачу, то скорость доступа к данным может возрасти значительно. Файловая система, созданная под одну задачу несравнимо быстрее, чем универсальный драйвер. Стандартный (универсальный) драйвер файловой системы проходит следующие шаги: WinAPI, контроллер, файловая таблица, кластеризация, адресация и пр. Данные шаги значительно увеличивают время выполнения главной функции I/O в ИСВН. Создав свой драйвер для ИСВН, можно значительно повысить скорость (снизить время) доступа к данным на жестком диске без изменения технических характеристик видеосервера и сэкономить технические ресурсы на обработку данных с видеокамер [8-10].

Таким образом, для обеспечения высокой скорости чтения и записи данных в ИСВН требуется решить следующие задачи:

- создать программу для работы с жестким диском (драйвер файловой системы);
- разработать базу данных для хранения информации ИСВН;
- выполнить интеграцию драйвера с ИСВН через объекты и методы базы данных;
- обеспечить взаимную совместимость трех компонентов: ИСВН, базы данных, файловых систем;
- разработать программы для оценки скорости под разные операционные системы.

Типовая задача разработки специализированного драйвера для ИСНВ может быть сформулирована следующим образом. Необходимо хранить статистическую информацию, серию кадров с камер видеонаблюдения (10-24 кадров/с с разрешением 4k и размерностью 3840 × 2160 пикселов и глубиной dpi: 450 с цветностью sRGB), видеофрагменты (25-35 МБ в формате MPEG/AVI и длиной 5–7 с), информацию о найденных объектах и событиях (одна строка с 31 ячейкой: дата, время, погодные условия, освещенность, количество объектов, точность определения в процентах, предполагаемое событие, характеристики видеокамеры, время года, время суток, день недели и информацию об объекте: цех, проходная, столовая, раздевалка и т. д.). Технические характеристики видеосервера: 2x Intel Xeon Gold gen3, 32 ГБ ОЗУ, 15х LAN 1 Гбит/с, 250GB SSD 2.5" SATA, CXД — HDD 10TB Ent 7.2k SATA, IPMI 2.0, подключение 2 мониторов (1x HDMI, 1x DVI/ HDMI), видеокарта nVidia GT 2Gb, операционная система Windows 11 IoT Enterprise High End¹ с поддержкой популярных программ: ITV/AXXON, Лини, AXIS, Macroscop, Milestone, TRASSIR. При использовании универсального драйвера ИСВН с рассмотренной конфигурацией позволяет выполнять сбор и обработку данных не более, чем с 12 видеокамер в непрерывном

¹ [Электронный pecypc]. https://www.videomax.ru/ production/videomax_ip/ (дата обращения: 22.03.2025).

Научно-технический вестник информационных технологий, механики и оптики, 2025, том 25, № 2 Scientific and Technical Journal of Information Technologies, Mechanics and Optics, 2025, vol. 25, no 2

режиме. Разрабатываемый драйвер должен обеспечивать запись и параллельное чтение в постоянном режиме 100 Мбит/с при использовании одной видеокамеры. За счет создания нового специализированного драйвера должна быть обеспечена возможность одновременного подключения в режиме записи данных не менее, чем 20 видеокамер в непрерывном режиме.

Описание разработанной файловой системы

Для создания специализированной файловой системы предполагается использовать базу данных SubBase, которая разработана непосредственно для хранения информации ИСВН. База данных SubBase значительно быстрее аналогов и работает по принципу Redis¹, когда вся информация находится в оперативной памяти и периодически изменения сбрасываются на жесткий диск в MongoDB или любую другую документ-ориентированную системы управления базами данных (СУБД).

База данных ИСВН SubBase имеет ячеистую структуру и хранит информацию в виде таблиц, в которых размещаются данные определенного назначения в зависимости от их типа (дата, строка, текст, бинарное значение). Имеются таблицы в SubBase, которые хранят бинарные данные в формате BSON: серии кадров (JPEG/PNG), видеофрагменты (AVI/MPEG4), аудиопотоки в момент детекции объектов и событий, которые были на их основе определены. База данных (рис. 2) представлена в графическом виде в адресном пространстве жесткого диска, который имеет начальный и конечный адреса (условно: 00000001 и 00FFFFFF). Свободное дисковое пространство резервируется от адреса 00000XXX до адреса 00000YYY с емкостью SSS. Отличительной особенностью разработанной базы данных SubBase является жесткая и стабильная структура, состоящая из таблиц фиксированной длины, количеством N в двух копиях. Ячейки жестко ограничены в каждой таблице, имеют последовательность при записи данных, что позволяет достичь высокой скорости данных I/O в ИСВН.

ИСВН может работать с базой данных любого типа (SQL/NoSQL), но быстрее всего она работает с профильной базой SubBase с оптимизациями на уровне жесткого диска (MyISAM/InnoDB).

Обычно файловая система требует неразмеченное пространство на жестком диске [11–13], созданное утилитами для создания и управления разделами диска (Fdisk, GNU Parted, GParted, GNOME Disks, KDE Partition Manager и пр.). Однако предлагаемая файловая система SubSysVideo не требует дополнительных действий от разработчика. Удаление раздела и создание скрытого раздела не обязательно. Достаточно создать файл с условным названием «fileSubSysVideo.sys» в скрытом виде, только для чтения, и зафиксировать без возможности записи и чтения через WinAPI в NTFS. Стандартный драйвер Windows не будет получать доступа к файлу с размерностью SSS и не станет дефрагментироваться, что очень важно. Это необходимо,





Рис. 2. Графическое представление базы данных SubBase *Fig.* 2. Graphical representation of the SubBase database

чтобы начальный и конечный адреса файла на жестком диске оставались неизменными. Далее запускается программа на языке программирования С++, скомпилированная на GPP (GNU) и читает побайтово дисковое пространство файла от X с размерностью SSS. Размер файла должен равняться двум копиям развернутой базы данных SubBase-v.254s в 4 ГБ. Доступ к данным ИСВН [14-16] получает через объект, которым является вся база данных SubBase. Объект управления SubBase подобъект SubAction, а для работы с таблицами и строками — подобъект SubTable. Подобъект базы данных новый объект, созданный с использованием метода «CreateObj();» главного объекта SubBase. У подобъекта SubAction имеются методы: Open, CopyBack, Create, Story и многие другие, которые позволяют копировать базу данных на жесткий диск из оперативной памяти. Методы: Insert, Update, Delete подобъекта SubTable (который создается через методы объекта SubAction) работают непосредственно с записями таблицы и обеспечивают: добавление новых записей, обновление существующих записей, удаление записей из одной таблицы. Используется размерность файловой системы 8 ГБ (4 ГБ + 4 ГБ) для основной базы данных, ко-





Fig. 3. Scheme of access to data of the intelligent video surveillance system using the developed file system

торая сохраняется на жесткий диск один раз в час, и резервной копии, которая копируется один раз в сутки. Вследствие создания резервных копий базы данных SubBase обеспечивается надежность работы, за счет разработки нового драйвера файловой системы под условным названием SubSysVideo-1.9.17fs достигается высокая скорость работы.

На рис. 3 схематично изображено предлагаемое решение, позволяющее решить проблему низкой скорости доступа к файловой системе применительно к одной задаче ИСВН. На схеме ИСВН изображена по центру и работает с базой данных.

Экспериментальное исследование разработанной файловой системы

В рамках настоящей работы разработано приложение для MacOS Sequoia v.15, которое позволяет измерить скорость чтения и записи I/O данных в разных файловых системах. В ходе эксперимента проводилось копирование одного объекта (размером 4 ГБ) из оперативной памяти на жесткий диск, затем чтение с жесткого диска (обратная загрузка одного объекта в оперативную память). Для этого жесткий диск SSD размером 256 ГБ был разделен на два раздела: 154 ГБ под операционную систему и по 24 ГБ под каждую файловую систему. Чтение и запись объекта из RAM в раздел 24 ГБ выполнено последовательно через каждые 1,75 мин и с начала раздела, чтобы избежать несогласованности нитей данных.

Интерфейс приложения для измерения времени доступа (рис. 4) компилируется в кроссплатформенной среде Lazarus IDE под все популярные операционные системы (Windows 64k, Linux, QNX, MacOS, BSD). Использованы стандартные библиотеки для работы с файловой системой из индикативной среды разработки Lazarus. В инсталляторе IDE всегда присутствует модуль StrUtils, который предоставляет методы: Assign(subfs), Rewrite(subfs, pin), Write(subfs, str), Close(subfs) для работы с типом данных «file of integer». На рис. 4 представлен перечень файловых систем с возможностью отметить все настройки, которые позволяют указать размер объекта (4 ГБ), интервал между замерами (по умолчанию — 1,75 мин), отметить возможность записывать с начала раздела; состояние (название текущей файловой системы и время доступа). Правее от блока состояния показаны кнопки со стрелками для просмотра статистики по всем файловым системам, а по центру — возврат на первый пункт NTFS. Допустим сброс настроек в первоначальное значение (по кнопке «Сбросить»). Внизу размещены кнопки управления: «Старт» и «Стоп» для процесса проверки.

Проведены измерения скорости записи и чтения данных в пяти host операционных системах: Windows 11 (техническая версия сборки: 10.0.22631.4169), MacOS Sequia 15, FreeBSD 14.1, QNX Neutrino 7.1 и Debian 12 Linux 6.7.6. При сравнении рассматривались самые популярные файловые системы: NTFS, FAT32, exFAT, EXT, HFS+, APFS и созданная в настоящей работе — SubSysVideo-1.9.17fs. После проведения измерений (таблица) выяснилось, что предложенная файловая





Fig. 4. Interface of the application for measuring data access time under MacOS (*a*) and Windows (*b*)

система работает быстрее на 43,4 % относительно NTFS.

Для воспроизведения эксперимента необходимо установить Python v.3.12.7pre под Mac OS X¹ и использовать приложение² для записи и чтения одного объекта из оперативной памяти в начало любого раздела. Также необходимо подготовить отдельные разделы по 24 ГБ, отформатированные в разных файловых системах. Для проведения измерений применялся компьютер Мас с техническими характеристиками: Apple MacBook Air (Z12700036), 13.3", (WQXGA), 256 ГБ SSD (450-550 ME/c), 7-core GPU M1, Mac OS Sequia, Silver. Результаты, полученные в таблице и в эксперименте, показали, что больше всего дополнительных функций находятся в файловой системе NTFS при работе из FreeBSD 14.1, что обусловлено второстепенностью драйвера NTFS под BSD систему. NTFS обладает огромным количеством дополнительных функций от главного разработчика Microsoft, относительно других файловых систем: FAT32, exFAT, EXT, HFS+, APFS и др. Быстрее всего работает файловая система ЕХТ, поскольку обладает минимумом функций. Отметим, что в

¹ [Электронный pecypc]. https://www.python.org/downloads/ macos/ (дата обращения: 22.03.2025).

² [Электронный ресурс]. https://github.com/alex1543/ practPythonServ/tree/main/mac (дата обращения: 22.03.2025).

Название	Операционная система										
файловой системы	Windows 11	MacOS Sequia 15	FreeBSD 14.1	QNX Neutrino 7.1	Debian 12 Linux 6.7.6	Увеличение по сравнению c SubSysVideo-1.9.17fs, %					
NTFS	14,7	15,6	15,8	15,7	15,5	43,4					
FAT32	11,9	12,8	13,1	12,9	12,7	30,9					
exFAT	12,6	13,7	13,9	13,8	13,3	35,1					
EXT	9,7	10,3	11,2	10,7	10,1	15,7					
HFS+	13,8	14,4	14,5	14,5	13,9	38,6					
APFS	12,7	13,1	13,7	13,4	12,8	33,4					
SubSysVideo-1.9.17fs	8,4	8,7	9,1	8,9	8,6						

Таблица. Время доступа к данным на запись и чтение (один объект в озере данных — 4 1	ΓБ), с
<i>Table.</i> Read and write access time (one object in the data lake – 4 GB), s	

настоящей работе удалось ускорить функции записи и чтения в SubSysVideo-1.9.17fs на 15,7 % относительно самой быстрой файловой системы ЕХТ и на 43,4 % относительно самой медленной файловой системы NTFS от Microsoft.

Заключение

Разработана файловая система, ориентированная на решение одной задачи — хранение данных интеллектуальной системы видеонаблюдения. Выполненный анализ позволил выявить, что время доступа к данным на жестком диске может быть значительно сокращено за счет исключения универсальных функций, реализуемых современными файловыми системами. Предложена файловая система для хранения данных интеллектуальной системы видеонаблюдения и база данных для хранения ее информации, разработан драйвер, который работает напрямую с жестким диском, не использует функции операционной системы, что повышает скорость работы с данными.

Литература

- Avhad A.R., Gangad P.S., Kharote S.M., Muntode S.S., Sanap M.D. Blockchain based secure file transfer system with password protection // International Journal of Advanced Research in Science, Communication and Technology (IJARSCT). 2024. V. 4. N 2. P. 299– 303. https://doi.org/10.48175/ijarsct-19651
- Panjuta D., Jabbarzadeh J. C# File System // Learning C# Through Small Projects. Springer, 2024. P. 167–198. https://doi. org/10.1007/978-3-031-51914-7_6
- Needhi J., Prasath R., Vikram K.K., Vishnu G. Performance optimization of voice-assisted file management systems // International Journal of Engineering and Computer Science. 2024. V. 13. N 7. P. 26250–26256. https://doi.org/10.18535/ijecs/v13i07.4854
- Cho K., Bahn H. A lightweight file system design for unikernel // Applied Sciences. 2024. V. 14. N 8. P. 3342. https://doi.org/10.3390/ app14083342
- Gui J., Wang Y., Shuai W. Improving reading performance by file prefetching mechanism in distributed cache systems // Concurrency and Computation: Practice and Experience. 2024. V. 6. N 22. P. e8215 https://doi.org/10.1002/cpe.8215
- Yuliana M., Hidayah N., Sudarsono A. Implementation of Web-Based file sharing security system // MOTIVECTION: Journal of Mechanical, Electrical and Industrial Engineering. 2024. V. 6. N 1. P. 41–52. https:// doi.org/10.46574/motivection.v6i1.314

Проведенные экспериментальные исследования показали, что скорость доступа к данным повысилась на 43,4 % относительно файловой системы NTFS, которая оказалась самой медленной файловой системой из-за использования большого количества сервисных функций.

Разработка специализированных файловых систем, предназначенных для решения одной или нескольких задач, может быть востребована и в других областях деятельности. Например, такие файловые системы могут использоваться для хранения данных цифровых двойников, систем статистической обработки и визуализации информации, где необходимо обеспечить высокую скорость доступа к данным жесткого диска. Специализированные файловые хранилища могут быть разработаны не только для доступа к данным, размещенным на жестких дисках, но и на других носителях: flash-накопителях, оптических дисках, лентах и прочих устройствах хранения данных.

References

- Avhad A.R., Gangad P.S., Kharote S.M., Muntode S.S., Sanap M.D. Blockchain based secure file transfer system with password protection. *International Journal of Advanced Research in Science*, *Communication and Technology (IJARSCT)*, 2024, vol. 4, no. 2, pp. 299–303. https://doi.org/10.48175/ijarsct-19651
- Panjuta D., Jabbarzadeh J. C# File System. *Learning C# Through Small Projects*. Springer, 2024, pp. 167–198. https://doi.org/10.1007/978-3-031-51914-7_6
- Needhi J., Prasath R., Vikram K.K., Vishnu G. Performance optimization of voice-assisted file management systems. International *Journal of Engineering and Computer Science*, 2024, vol. 13, no. 7, pp. 26250–26256. https://doi.org/10.18535/ijecs/v13i07.4854
- Cho K., Bahn H. A lightweight file system design for unikernel. *Applied Sciences*, 2024, vol. 14, no. 8, pp. 3342. https://doi. org/10.3390/app14083342
- Gui J., Wang Y., Shuai W. Improving reading performance by file prefetching mechanism in distributed cache systems. *Concurrency and Computation: Practice and Experience*, 2024, vol. 36, no. 22, pp. e8215 https://doi.org/10.1002/cpe.8215
- Yuliana M., Hidayah N., Sudarsono A. Implementation of Web-Based file sharing security system. *MOTIVECTION: Journal of Mechanical, Electrical and Industrial Engineering*, 2024, vol. 6, no. 1, pp. 41–52. https://doi.org/10.46574/motivection.v6i1.314

- Man T., Osipov V.Yu., Zhukova N., Subbotin A., Ignatov D. Neural networks for intelligent multilevel control of artificial and natural objects based on data fusion: A survey // Information Fusion. 2024. V. 110. P. 102427. https://doi.org/10.1016/j.inffus.2024.102427
- Man T., Vodyaho A., Zhukova N., Subbotin A., Shichkina Y. Urban intelligent assistant on the example of the escalator passenger safety management at the subway stations // Scientific Reports. 2023. V. 13. N 1. P. 15914. https://doi.org/10.1038/s41598-023-42535-x
- Osipov V., Zhukova N., Subbotin A., Glebovskiy P., Evnevich E. Intelligent escalator passenger safety management // Scientific Reports. 2022. V. 12. N 1. P. 5506. https://doi.org/10.1038/s41598-022-09498-x
- Cowan D.D., Stepien T.M., Ierusalimschy R., Lucena C.J.P. Application integration: Constructing composite applications from interactive components // Software: Practice and Experience. 1993. V. 23. N 3. P. 255–275. https://doi.org/10.1002/spe.4380230304
- Sudharsan S., Sakthi Anand A., Shanmugaraj K., Palani Samy K.C. Deep learning-based intelligent video surveillance system for realtime motion detection // International Scientific Journal of Engineering and Management. 2024. V. 3. N 4. P. 1–10. https://doi. org/10.55041/ISJEM01492
- Li J., Zheng Z., Li Y., Ma R., Xia S. Multitask deep learning for Edge Intelligence Video Surveillance system // Proc. of the IEEE 18th International Conference on Industrial Informatics (INDIN). 2020. P. 579–584. https://doi.org/10.1109/INDIN45582.2020.9442166
- Жукова Н.А., Субботин А.Н. Алгоритм динамического распределения обработки изображений в облачных системах интеллектуального видеонаблюдения // Информационно-управляющие системы. 2024. № 6 (133). С. 15–26. https://doi. org/10.31799/1684-8853-2024-6-15-26
- Bhatia J., Patel T., Trivedi H., Majmudar V. Htv dynamic load balancing algorithm for virtual machine instances in cloud // Proc. of the International Symposium on Cloud and Services Computing. 2012. P. 15–20. https://doi.org/10.1109/ISCOS.2012.25
- Rahm E., Do H.H. Data cleaning: problems and current approaches // Bulletin of the IEEE Computer Society Technical Committee on Data Engineering. 2000. V. 23. N 4. P. 3–13.
- Saecker M., Markl V. Big data analytics on modern hardware architectures: a technology survey // Lecture Notes in Business Information Processing. 2013. V. 138, P. 125–149. https://doi. org/10.1007/978-3-642-36318-4_6

Авторы

Субботин Алексей Николаевич — аспирант, Санкт-Петербургский государственный электротехнический университет «ЛЭТИ» им. В.И. Ульянова (Ленина), Санкт-Петербург, 197022, Российская Федерация; ведущий инженер-программист, ГУП «Петербургский метрополитен», Санкт-Петербург, 190013, Российская Федерация, 57226847043, https://orcid.org/0000-0002-4823-6288, alesu1543@ gmail.com

Жукова Наталия Александровна — доктор технических наук, профессор, старший научный сотрудник, Санкт-Петербургский федеральный исследовательский центр Российской академии наук (СПб ФИЦ РАН), Санкт-Петербург, 199178, Российская Федерация; 56406142300, https://orcid.org/0000-0001-5877-4461, nazhukova@ mail.ru

Статья поступила в редакцию 06.10.2024 Одобрена после рецензирования 15.01.2025 Принята к печати 25.03.2025

- Man T., Osipov V., Zhukova N., Subbotin A., Ignatov D. Neural networks for intelligent multilevel control of artificial and natural objects based on data fusion: A survey. *Information Fusion*, 2024, vol. 110, pp. 102427. https://doi.org/10.1016/j.inffus.2024.102427
- Man T., Vodyaho A., Zhukova N., Subbotin A., Shichkina Y. Urban intelligent assistant on the example of the escalator passenger safety management at the subway stations. *Scientific Reports*. 2023. vol. 13, no. 1, pp. 15914. https://doi.org/10.1038/s41598-023-42535-x
- Osipov V., Zhukova N., Subbotin A., Glebovskiy P., Evnevich E. Intelligent escalator passenger safety management. *Scientific Reports*, 2022, vol. 12, no. 1, pp. 5506. https://doi.org/10.1038/s41598-022-09498-x
- Cowan D.D., Stepien T.M., Ierusalimschy R., Lucena C.J.P. Application integration: Constructing composite applications from interactive components. *Software: Practice and Experience*, 1993, vol. 23, no. 3, pp. 255–275. https://doi.org/10.1002/spe.4380230304
- Sudharsan S., Sakthi Anand A., Shanmugaraj K., Palani Samy K.C. Deep learning-based intelligent video surveillance system for realtime motion detection. *International Scientific Journal of Engineering and Management*, 2024, vol. 3, no. 4, pp. 1–10. https://doi. org/10.55041/ISJEM01492
- Li J., Zheng Z., Li Y., Ma R., Xia S. Multitask deep learning for Edge Intelligence Video Surveillance system. *Proc. of the IEEE 18th International Conference on Industrial Informatics (INDIN)*, 2020, pp. 579–584. https://doi.org/10.1109/INDIN45582.2020.9442166
- Zhukova N.A., Subbotin A.N. Dynamic distribution algorithm for image processing in cloud-based intelligent video surveillance systems. *Information and Control Systems*, 2024, no. 6 (133). pp. 15– 26. (in Russian). https://doi.org/10.31799/1684-8853-2024-6-15-26
- Bhatia J., Patel T., Trivedi H., Majmudar V. Htv dynamic load balancing algorithm for virtual machine instances in cloud. *Proc. of the International Symposium on Cloud and Services Computing*, 2012, pp. 15–20. https://doi.org/10.1109/ISCOS.2012.25
- Rahm E., Do H.H. Data cleaning: problems and current approaches. Bulletin of the IEEE Computer Society Technical Committee on Data Engineering, 2000, vol. 23 no. 4, pp. 3–13.
- Saecker M., Markl V. Big data analytics on modern hardware architectures: a technology survey. *Lecture Notes in Business Information Processing*, 2013, vol. 138, pp. 125–149. https://doi. org/10.1007/978-3-642-36318-4

Authors

Alexey N. Subbotin — PhD Student, Saint Petersburg Electrotechnical University "LETI", Saint Petersburg, 197022, Russian Federation; Leading Engineer, Software Developer, State Unitary Enterprise "Petersburg Metropolitan", Saint Petersburg, 190013, Russian Federation, SC 57226847043, https://orcid.org/0000-0002-4823-6288, alesu1543@ gmail.com

Nataly A. Zhukova — D.Sc., Professor, Senior Researcher, St. Petersburg Federal Research Center of the Russian Academy of Sciences, Saint Petersburg, 199178, Russian Federation; <u>55</u> 56406142300, https://orcid. org/0000-0001-5877-4461, nazhukova@mail.ru

Received 06.10.2024 Approved after reviewing 15.01.2025 Accepted 25.03.2025



Работа доступна по лицензии Creative Commons «Attribution-NonCommercial» **VİTMO**

НАУЧНО-ТЕХНИЧЕСКИЙ ВЕСТНИК ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ, МЕХАНИКИ И ОПТИКИ март–апрель 2025 Том 25 № 2 http://ntv.ifmo.ru/ SCIENTIFIC AND TECHNICAL JOURNAL OF INFORMATION TECHNOLOGIES, MECHANICS AND OPTICS March–April 2025 Vol. 25 № 2 http://ntv.ifmo.ru/en/ ISSN 2226-1494 (print) ISSN 2500-0373 (online)

ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ, МЕХАНИКИ И ОПТИКІ

doi: 10.17586/2226-1494-2025-25-2-261-272 УДК 004.932

Кластеризация пространственных данных с неявно выраженной полигональной структурой на основе топологических подходов Сергей Владимирович Еремеев⊠

Муромский институт (филиал) ФГБОУ ВО «Владимирский государственный университет имени Александра Григорьевича и Николая Григорьевича Столетовых», Муром, 602264, Российская Федерация sv-eremeev@yandex.ru[⊠], https://orcid.org/0000-0001-8482-1479

Аннотация

Введение. Кластеризация является одним из фундаментальных подходов для интеллектуального анализа данных, который в сфере геоинформатики и обработки изображений используется для поиска знаний и скрытых закономерностей пространственной информации. При автоматической векторизации объектов на спутниковых снимках из-за несовершенства этих технологий на линейных и полигональных объектах появляются пропущенные элементы, что препятствует полноценному анализу и визуализации данных. Представлен новый метод кластеризации геометрических примитивов с неявно выраженной полигональной структурой с возможностью устранения неполных данных в векторных моделях. Метод. Предлагаемый метод основан на итерационном формировании пространственных структур при растяжении исходных объектов линейного типа. В отличие от многих подходов кластеризации элементы группируются в кластеры не по принципу ближайшего евклидова расстояния, а за счет определения ближайшего пересечения между отрезками. Такой подход позволяет корректно разделить рядом стоящие линейные объекты в разные кластеры. Для формирующихся пространственных структур на каждой итерации в зависимости от коэффициента растяжения вычисляются их топологические особенности, что обеспечивает возможность обнаруживать и проводить фильтрацию неявно выраженных полигональных структур. Основные результаты. Разработанный метод апробирован для кластеризации линейных геометрических примитивов на векторных моделях городской инфраструктуры. Проведено сравнение работы метода с аналогами: k-средних, Density-Based Spatial Clastering of Application with Noise (DBSCAN) и агломеративной кластеризацией. Исследования показали, что с использованием метрики оценки качества кластеризации в виде индексов инерции и Жаккара предложенный метод имеет преимущество в случаях корректного разделения близко расположенных кластеров. Обсуждение. Полученные результаты могут быть использованы в геоинформационных системах для актуализации векторных данных после применения технологий автоматической векторизации объектов на спутниковых снимках.

Ключевые слова

кластеризация пространственной информации, топологический анализ данных, гомотетия, геоинформационные системы, оценка качества кластеризации

Благодарности

Исследование выполнено при финансовой поддержке Российского научного фонда в рамках научного проекта № 23-21-10064.

Автор выражает благодарность магистру в области информационных систем Егай Марии Владимировне за помощь в подготовке исходных данных и проведении экспериментов.

Ссылка для цитирования: Еремеев С.В. Кластеризация пространственных данных с неявно выраженной полигональной структурой на основе топологических подходов // Научно-технический вестник информационных технологий, механики и оптики. 2025. Т. 25, № 2. С. 261–272. doi: 10.17586/2226-1494-2025-25-2-261-272

© Еремеев С.В., 2025

Научно-технический вестник информационных технологий, механики и оптики, 2025, том 25, № 2 Scientific and Technical Journal of Information Technologies, Mechanics and Optics, 2025, vol. 25, no 2

Clustering of spatial data with implicit polygonal structure based on topological approaches

Sergey V. Eremeev⊠

Murom Institute (Branch) of Vladimir State University, Murom, 602264, Russian Federation sv-eremeev@yandex.ru[⊠], https://orcid.org/0000-0001-8482-1479

Abstract

Clustering is one of the fundamental approaches for data mining, which in the field of geoinformatics and image processing is used to search for knowledge and hidden patterns of spatial information. During automatic vectorization of objects on satellite images due to imperfections of these technologies, missing elements appear on linear and polygonal objects, which prevent full-fledged data analysis and visualization. The paper considers the problem of clustering geometric primitives with implicit polygonal structure with the possibility of eliminating incomplete data in vector models. The proposed method is based on the iterative formation of spatial structures by stretching the original linear objects. Unlike many clustering approaches, elements are grouped into clusters not by the principle of nearest Euclidean distance, but by determining the nearest intersection between segments. This approach allows us correctly dividing adjacent objects into different clusters. For the spatial structures formed at each iteration, their topological features are calculated depending on the stretching coefficient, which makes it possible to detect and filter implicit polygonal structures. The developed method is tested for clustering of linear geometric primitives on vector models of urban infrastructure. The performance of the method is compared with its competitors: *k*-means, DBSCAN, and agglomerative clustering. The research has shown that using the clustering quality assessment metric in the form of inertia and Jaccard indices, the proposed method has an advantage due to the correct separation of closely located clusters.

Keywords

spatial information clustering, topological data analysis, homothety, geographic information systems, clustering quality assessment

Acknowledgements

This study was supported by the Russian Science Foundation, project no. 23-21-10064.

The author expresses gratitude to the Master of Information Systems, Maria Vladimirovna Egay, for her assistance in preparing the initial data and conducting the experiments.

For citation: Eremeev S.V. Clustering of spatial data with implicit polygonal structure based on topological approaches. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 2025, vol. 25, no. 2, pp. 261–272 (in Russian). doi: 10.17586/2226-1494-2025-25-2-261-272

Введение

Кластеризация является одним из фундаментальных подходов для интеллектуального анализа данных, под которым понимается группировка объектов по схожим признакам. В сфере геоинформатики и обработки изображений кластеризация используется для поиска знаний и скрытых закономерностей пространственной информации.

Сложность обрабатываемых данных с каждым годом возрастает [1, 2]. Это обуславливает потребность в разработке новых идей и компьютерных методов для извлечения ценной информации о пространственных объектах из быстрорастущего объема цифровых данных [3].

В настоящее время наблюдается активный рост научных работ, в которых предлагаются решения для автоматической векторизации объектов на спутниковых снимках с последующим их распределением по разным слоям. Однако существующие технологии векторизации несовершенны. Это подтверждает выполненный систематический обзор в работе [4] для выделения на изображении линейных отрезков, состоящих из двух соединенных между собой точек. В работе [4] обозначены следующие основные проблемы, которые препятствуют обнаружению линейных и полигональных объектов из растрового изображения: нестабильность конечных точек отрезка из-за шума, теней и освещенности изображения; нехватка наборов данных для машинного обучения; перекрытие объектов другими объектами. В результате векторизации на линейных и полигональных объектах образуются пропущенные элементы, что затрудняет применение полноценного анализа и визуализации пространственных данных. Таким образом, на некоторых участках получается набор векторных объектов с неполными данными. Актуальной является задача разработки метода кластеризации линейных отрезков, которые представляют неявно выраженную пространственную структуру, и дальнейшего устранения пропущенных данных в векторных моделях.

Обзор существующих подходов

В качестве исходных данных рассмотрим наборы из линейных отрезков, которые потенциально образуют полигональные объекты. Пример таких данных показан на рис. 1. Заметим, что визуально отрезки образуют контуры трех зданий и могут быть представлены как полигоны.

Сформированные наборы данных получаются в результате работы различных технологий автоматического преобразования информации в векторный формат из спутниковых снимков или облака точек после лидарной съемки. В [5] для выделения отрезков на крышах зданий из облака точек применена спектральная кластеризация прямолинейных сегментов. В работе [6] из облака точек построены отрезки, формирующие полигональный объект, на основе семантической маркировки и кластеризации с использованием плотности



Puc. 1. Векторный набор данных из линейных отрезков с неявно выраженной полигональной структурой *Fig. 1.* Vector dataset of linear segments with implicit polygonal

structure

точек. Из точечных данных в [7] предложено выделять ключевые точки и относительно них формировать линейные отрезки. Также по опорным точкам и исходным спутниковым снимкам в работе [8] построены прямоугольные контуры зданий, но отметим, в данной работе опорные точки ставит пользователь вручную, что является дорогостоящим процессом при обработке больших объемов пространственной информации.

Результаты с достаточно высокой степенью точности по обнаружению отрезков на изображении представлены в работе [9], в которой предлагается включение сверточной нейронной сети в классический детектор Line Segment Detector (LSD).

Существует множество методов для кластеризации. Современный обзор методов кластеризации больших данных представлен в [10]. Наиболее широко используются следующие из них: *k*-средних, Density-Based Spatial Clastering of Application with Noise (DBSCAN), агломеративная кластеризация. Рассмотрим особенности этих методов и применим их для кластеризации векторных данных.

Метод *k*-средних разбивает исходное множество на *k* кластеров так, что внутри группы суммарное квадратичное отклонение объектов от центра кластера минимизируется. Полноценный обзор существующих разновидностей метода *k*-средних и существующие проблемы его применения показаны в [11]. Из-за использования евклидова расстояния в качестве метрики сходства одной из основных проблем является его ограничение при обнаружении других форм кластеров, что может повлиять на анализ пространственных объектов. Кроме того, для больших объемов пространственных данных определить оптимальное количество кластеров является достаточно сложной задачей.

Метод DBSCAN относится к подходам кластеризации на основе плотности, где объекты, расположенные в области с высокой плотностью, формируют отдельные кластеры, отделенные между собой областями с низкой плотностью [12].

Агломеративная иерархическая кластеризация начинает формировать отдельные кластеры для каждой точки исходных данных и объединяет кластеры с наименьшим расстоянием между ними, чтобы создать более крупные кластеры. Этот процесс повторяется до тех пор, пока не будет сформирован один большой кластер, содержащий все данные.

Для группировки векторных объектов в потенциальные полигоны рассмотрим следующие подходы. При кластеризации линейных отрезков ведутся разработки для более выгодного их объединения [13]. Кроме того, в [14] рассматривается кластеризация отрезков в многомерном пространстве, что позволяет использовать результаты в статистическом анализе данных.

В работе [15] применено понятие связности для группировки объектов, что позволяет сделать разбиение на кластеры более оптимально. Похожая тенденция для кластеризации с использованием пространственных отношений наблюдается в [16], где для группировки объектов дополнительно применяется их атрибутивная информация. Кроме того, связи между объектами в виде топологических отношений в сочетании с агломеративной кластеризацией рассмотрены в работе [17].

Топологические отношения и топология в целом играют ключевую роль в анализе пространственных данных. Предлагаемый в настоящей работе метод также будет опираться на топологические подходы. Пространственные объекты, не связанные между собой, образуют сложную структуру, которая формируется по определенным и не известным заранее правилам. Это выражается в объединении пространственных данных в специальные формы кластеров, не зависящие от евклидова расстояния между исходными объектами. Существует множество ситуаций, когда рядом стоящие объекты относятся к разным кластерам. Большинство методов кластеризации не позволяют корректно разделить объекты на кластеры для таких исходных данных, потому что в одну группу объединяются ближайшие по расстоянию объекты. Это ограничивает обнаружение скрытых связей в пространственных структурах данных. Для исследования из множества пространственных данных выделим отдельный класс объектов линейного типа, которые неявным образом представляют собой полигональные объекты.

Целью настоящей работы является улучшение обнаружения неявно выраженных полигональных структур. Метод обеспечивает повышение эффективности кластеризации объектов линейного типа и выявление скрытых полигональных структур на основе доступной пространственной информации при соответствующей оценке качества полученных решений. В отличие от известных подходов, объекты объединяются в кластеры без вычисления ближайшего евклидова расстояния между ними. Вместо этого предлагается использовать принцип растяжения исходных линейных отрезков до их ближайшего пересечения между собой с последующим формированием структур полигонального типа.

Изложен принцип объединения линейных отрезков в кластеры, который основывается не на ближайшем расстоянии между отрезками, а на ближайшем пересечении между ними при их растяжении. Показан способ формирования пространственных структур, каждая из которых представляет собой кластер, образованный в результате пересечения отрезков при их растяжении. Определяются топологические свойства образованных структур, которые дают возможность обнаруживать скрытые объекты полигонального типа.

Методы и материалы

Основные принципы кластеризации на основе топологических подходов. Топологические подходы при анализе данных используют инструменты топологии для изучения формы и связей между данными. При кластеризации данные абстрагируются в виде графа. Это позволяет находить шаблоны и структуры, такие как компоненты связности, циклы и другие, которые не могут быть выявлены с помощью статистических методов.

Графы могут быть использованы для представления пространственных объектов и их отношений на географических картах. Топология в этом контексте изучает свойства и структуру геометрических объектов в виде точек, линий и полигонов, а также определяет их связность и отношения.

Среди современных топологических подходов выделим топологический анализ данных. В последнее время он получил широкое распространение в разных областях, включая анализ пространственных данных. Этот подход рассматривает данные как множество точек в метрическом пространстве. Точки могут быть связаны между собой. Для этого вокруг каждой точки строится шар с постепенным увеличением радиуса. При пересечении шаров точки соединяются между собой отрезками, образуя топологические структуры. После этого анализируются диаграммы персистентности, которые показывают, какая структура сохраняется в течение определенного промежутка пространства [18, 19].

Описание предлагаемого метода. Рассмотрим общую схему разрабатываемого метода кластеризации в сочетании с предварительной обработкой пространственной информации (рис. 2).

Исходными данными для предварительного этапа (этап 1) являются спутниковые снимки в видимом диапазоне в формате RGB (Red, Green, Blue). С использованием существующих подходов, наиболее популярные из которых преобразование Хафа и метод LSD, на снимках происходит выделение линейных отрезков. В силу различных причин, включая природные особенности и несовершенство подходов, в процессе формирования отрезков появляются ошибки, т. е. некоторые отрезки имеют пропуски, несмотря на то, что визуально их объединение представляет полигональную структуру. Заключительным этапом предварительной обработки является векторизация данных в отдельный слой геоинформационной системы. При этом отрезки хранятся в векторном виде в шейп-файлах (.shp).

Векторные данные поступают на вход для обработки (этап 2), который включает три основных шага: формирование пространственных структур (кластеров), вычисление признаков структур и их анализ для обнаружения полигональных объектов. Рассмотрим каждый из шагов более подробно.

На первом шаге (формирование пространственных структур) исходные данные представляют множество линейных отрезков $L = \{l_1, l_2, ..., l_{|L|}\}$ на плоскости, каждый элемент которого $l \in L$ состоит из двух концевых точек (x_1, y_1) и (x_2, y_2) .



Рис. 2. Общая схема предлагаемого метода кластеризации объектов линейного типа с учетом предварительной обработки спутниковых снимков



Для всех отрезков $l \in L$ введем растяжение в k раз. Для этого зададим гомотетию $H_{O,k}$ с центром в точке $O(x_0, y_0)$ и коэффициентом растяжения k. Координаты точки O для отрезка $l \in L$ вычисляются по формуле: $x_0 = \frac{x_1 + x_2}{2}, y_0 = \frac{y_1 + y_2}{2}$. Определим гомотетию в виде:

$$H_{O,k}:\begin{cases} x_i' = k(x_i - x_0) + x_0 \\ y_i' = k(y_i - y_0) + y_0 \end{cases}$$

где (x_i, y_i) и (x_i', y_i') — координаты концевых точек отрезка до и после растяжения $(i = 1, 2); (x_0, y_0)$ — точка середины исходного отрезка; $k \in [1, +\infty]$ — коэффициент растяжения.

Применяя гомотетию $H_{O,k}$ к каждому отрезку из L с постепенным возрастанием коэффициента растяжения k, получим обновленные отрезки с увеличенной длиной. При этом отрезки начнут пересекаться, образовывать более сложные структуры: компоненты связности и циклы. При пересечении пространственных элементов выполним их группировку. При появлении циклов будут формироваться полигональные структуры, представляющие собой кластеры, образованные на основе исходных отрезков из L.

Для анализа данных при создании топологических структур построим соответствующие диаграммы их жизненного цикла, которые позволят узнать на каком шаге итерационного процесса были сформированы эти структуры. Итерационный процесс применения гомотетии $H_{O,k}$ заканчивается тогда, когда будут построены все потенциальные полигональные объекты.

На рис. 3 показан пример итерационного процесса, где исходное множество отрезков состоит из 12 элементов (рис. 3, a). На начальном шаге при k = 1 имеем 12 отдельных компонент связности (рис. 3, b) в виде коротких штрихов. При постепенном растяжении каждого отрезка $l \in L$ в k раз происходит их пересечение, как, например, элементов l и 3 (рис. 3, c). В точке пересечения отрезки больше не растягиваются. Элементы 1 и 3 образуют полилинию, состоящую из двух отрезков. Будем при этом говорить, что одна компонента поглощает другую, т. е. в этой ситуации элемент *1* расширяется за счет присоединения элемента 3. У образованной полилинии сохраняется нумерация от поглощаемой компоненты связности. В данном случае приоритет при поглощении имеет элемент с наименьшим номером. В качестве приоритета можно также установить длину отрезка. Таким образом, элемент 3 прекращает свое существование при k = 1,2. Это показано на рис. 3, d, где все элементы имеют одинаковую длину штрихов, кроме элемента 3.

При k = 1,4 появляется первый замкнутый объект (рис. 3, *e*) и остается до конца итерационного процесса. Этому замкнутому объекту соответствует компонента связности 1 (полигон 1), так как при объединении с ней элементы 2, 3 и 4 прекратили свое существование, что



Puc. 3. Итерационный процесс формирования компонент связности и замкнутых объектов с отображением их топологических свойств на основе линейного растяжения отрезков

Fig. 3. Iterative process of formation of connectivity components and closed objects with representation of their topological properties on the basis of linear stretching of segments



Puc. 3. Продолжение *Fig. 3.* Continuation

наглядно показано на рис. 3, f в виде коротких штрихов. Эти штрихи у элементов 2, 3 и 4 увеличиваться дальше не будут. Информация о том, что появился замкнутый объект (рис. 3, g), показана на диаграмме, на которой продемонстрирован период существования созданных полигонов в зависимости от коэффициента растяжения.

Работа следующих шагов итерационного процесса, включая итоговый результат, показана на рис. 3, *h*-*m* с отображением текущих топологических характеристик элементов на диаграммах. Особое внимание обратим на рис. 3, *m*. Он показывает, при каком коэффициенте растяжения появились полигоны 1–3 при пересечении соответствующих отрезков. Самым последним появился полигон 3, ему соответствует компонента связности с номером 7. Заметим, что на рис. 3, *l* компоненты связности с номерами 1, 5 и 7 имеют самые длинные штрихи. Это означает, что их не поглощали другие компоненты связности.

Диаграммы имеют практический потенциал. Они способствуют более детальному изучению эволюции пространственных объектов, а также помогают выявить паттерны и тенденции в их развитии, что облегчает интерпретацию данных.

На втором шаге (вычисление топологических особенностей) происходит расчет топологических свойств сформированных структур. Данный процесс является итерационным и при каждом увеличении коэффициента растяжения происходит изменение топологических особенностей структур. Объединение двух компонент связности осуществляется по принципу ближайшего



Fig. 3. End

пересечения их концевых отрезков. При таком подходе в кластер группируются не сами исходные отрезки, а их измененное представление после растяжения. Кластером в этом случае является полилиния, включающая несколько отрезков, в том числе и концевые. Именно концевые отрезки продолжают растяжение до тех пор, пока не пересекутся с другими полилиниями. Процесс останавливается, если концевые отрезки не могут достичь пересечения с другими отрезками, когда они параллельны или требуется слишком большое значение коэффициента растяжения. Всю топологическую информацию о пространственных структурах хранят диаграммы, отражающие период существования компонент связности и полигонов относительно коэффициента растяжения.

На третьем шаге (выделение полигональных структур) происходит анализ топологических свойств. Например, можно по этим свойствам отфильтровать только те полигоны, которые появились на начальных значениях коэффициента растяжения. Это говорит о том, что неявно выраженные полигональные структуры имеют небольшие пропуски и с помощью растяжения исходные отрезки объединяются в кластеры, дополняя геометрическую форму полигона недостающими данными.

Предложенный метод имеет ограничения в случае недостижимости замкнутых объектов. Потому его использование целесообразно для таких исходных векторных данных, которые имеют неявно выраженную полигональную структуру.

Результаты экспериментальных исследований

Критерии для оценки качества кластеризации. Для оценки качества кластеризации множества линейных отрезков с неявно выраженной структурой были использованы следующие метрики: индексы Жаккара и инерции.

Индекс Жаккара — классическая мера сходства между двумя наборами данных, которая применяется в различных областях, включая геоинформатику [20]. Для двух сравниваемых полигональных объектов *A* и *B* индекс Жаккара вычислим как отношение площади их пересечения к площади объединения:

$$J(A, B) = \frac{S(A) \cap S(B)}{S(A) \cup S(B)},\tag{1}$$

где S(A) и S(B) — площади объектов A и B.

Индекс инерции является одним из наиболее распространенных подходов для оценки качества кластеризации и используется во многих практических задачах, в том числе, и для оптимизации [21]. Он предназначен для измерения суммарного внутрикластерного разброса данных в кластерах. Индекс вычисляет сумму квадратов расстояний между каждым объектом и центроидом своего кластера:

InertiaIndex =
$$\sum_{k=1}^{K} \sum_{y \in P_k} dist(y, c_k)^2$$
, (2)

где K — количество кластеров; $P_1, P_2, ..., P_K$ — непересекающиеся кластеры, состоящие из объектов $y \in P_k$ (k = 1, 2, ..., K); c_k — центроид кластера P_k ; $dist(y, c_k)^2$ — квадрат евклидова расстояния между объектом $y \in P_k$ и центроидом c_k из его класса.

При минимизации индекса инерции объекты внутри каждого кластера находятся близко друг к другу, а кластеры разделены между собой.

Проведение экспериментов и сравнение с аналогами. Проведение экспериментов базируется на схеме рис. 2. Первый этап является предварительным и предназначен для подготовки векторных данных из спут-



Puc. 4. Исходные изображения с разной плотностью расположения зданий *Fig. 4.* Input images with different building densities

никовых снимков. Примеры исходных изображений показаны на рис. 4. Были отобраны снимки, которые содержат здания с разной плотностью застройки, включая очень близко расположенные.

На основе преобразования Хафа были выделены отрезки, которые описывают контур зданий, при этом часть отрезков были отредактированы вручную. Применена геоинформационная система QGIS¹, с помощью которой для каждого исходного снимка был создан векторный слой из отрезков. В результате выбраны для исследования работы метода только те векторные данные, которые представляют собой неявно выраженные контуры зданий. Исходные векторные данные (табл. 1) доступны по ссылке².

Для проведения экспериментов на основе предлагаемого метода были применены основные шаги. Постепенное растяжение исходных отрезков позволяет формировать пространственные структуры различной формы, соответствующие кластерам. Анализ топологических свойств дает возможность обнаружить циклы, чтобы выявить замкнутые контуры.

В качестве аналогов были взяты методы кластеризации: *k*-средних, DBSCAN и метод агломеративной кластеризации. Все эксперименты проводились на основе разработанного модуля в QGIS. Так как эти методы работают с точечными данными, то в качестве исходных данных для них использованы концевые точки отрезков. Для улучшения визуализации для каждого кластера построены объемные контуры зданий (табл. 1). Анализ результатов показал, что методы *k*-средних, DBSCAN и метод агломеративной кластеризации имеют ограничения на выявление сложных кластеров с нестандартной формой или структурой.

С другой стороны, предлагаемый метод позволяет сгруппировать линейные отрезки в полигональные структуры с достаточно высокой точностью (табл. 1), но при определенных условиях. По сравнению с аналогами предлагаемый метод имеет преимущество в тех случаях, когда потенциальные полигональные объекты с неполными данными расположены рядом друг с другом. Если предполагаемые кластеры имеют значительное расстояние между собой, то, как показывает эксперимент 8, методы *k*-средних и DBSCAN имеют идентичный результат с предложенным методом.

Выполним сравнение ожидаемых кластеров с их предсказанными положениями, размерами и количеством. Кластер считается правильно установленным, если он покрывает площадь ожидаемой области свыше 75 %, но при этом не заходит на зону других кластеров более чем на 25 %. Для определения процента пересечения полигонов был использован индекс Жаккара (1). В случае DBSCAN кластером считается только тот объект, который содержит минимум три точки. Результаты приведены в табл. 2 в виде двух чисел.

Из табл. 2 видно, что методы *k*-средних и агломеративная кластеризация часто останавливаются слишком

¹ Электронный ресурс] Режим доступа: https://www.qgis. org/ (дата обращения: 05.12.2024).

² [Электронный ресурс] Режим доступа: https://github.com/ creems84/Howole-Geo/blob/main/VectorData.zip (дата обращения: 05.12.2024).

Номер эксперимента	Векторные данные	k-means	DBSCAN	Агломеративная кластеризация	Предложенный метод		
1							
2							
3			A DA A A A A A A A A A A A A A A A A A				
4	$\frac{\left[\left[\frac{1}{2}\right]^{2}}{\left[\frac{1}{2}\right]^{2}}\right] = \frac{1}{\left[\left[\frac{1}{2}\right]^{2}}$						
5							
6			000				
7	û LTÚ ú Úr ú						
8		000	4 N	000			
9			000				
10							

Таблица 1. Исходные наборы векторных данных и визуализация результатов кластеризации различными методами *Table 1.* Initial vector datasets and visualization of clustering results by different methods

рано, либо неконтролируемо разрастаются, из-за чего в лучшем случае форма объекта становится некорректной, как в эксперименте 1, а в худшем — в один кластер объединяются два или более зданий. Особенно эта проблема проявляется при двух рядом стоящих объектах, как в эксперименте 6.

Метод DBSCAN сталкивается с такой же проблемой, но, из-за отсутствия количества кластеров, она усугубляется тем, что для длинных зданий метод останавливается слишком рано, из-за чего появляется слишком большое число кластеров малого объема. Отметим, что предлагаемый метод, исправляет указанные недостатки. Для двух рядом стоящих зданий, если между ними есть хоть малейшее расстояние, то результат стремится к ожидаемому. Только в эксперименте 4, где начальные отрезки накладывались друг на друга, кластеры «слиплись». Исключением является эксперимент 1, где из-за слишком большого количества точек появились лишние кластеры, но при этом ожидаемый исход определился с высокой точностью.

Численные результаты экспериментов, полученные с помощью индекса инерции (2), приведены в табл. 3.

Таблица 2. Сравнение геометрических характеристик результатов кластеризации различными методами на основе индекса Жаккара

	0	Количество кластеров (правильно установленные/ошибочные)							
Номер эксперимента	Ожидаемое количество кластеров	k-means	DBSCAN	Агломеративная кластеризация	Предложенный метод				
1	7	3/6	0/10	3/6	7/8				
2	3	1/2	0/10	0/3	3/0				
3	5	3/2	0/5	3/2	5/0				
4	7	2/5	0/5	1/5	3/4				
5	3	0/3	0/2	0/3	3/0				
6	7	5/2	0/3	5/1	7/0				
7	7	3/4	4/2	4/2	7/0				
8	3	3/0	0/5	3/0	3/0				
9	7	5/2	5/1	1/6	7/0				
10	6	4/2	0/5	1/5	6/0				
Итого	55	29/28	9/48	21/33	51/12				

Table 2. Comparison of geometric characteristics of clustering results by different methods based on the Jaccard index

Примечание. Наилучшие значения показателей выделены жирным шрифтом.

Таблица 3. Индекс инерции для численной оценки качества кластеризац	(ии
Table 3. Inertia index for numerical estimation of clustering quality	

Номер эксперимента	k-means	DBSCAN	Агломеративная кластеризация	Предложенный метод
1	12 892 364	10 934 472	12 872 668	17 602 038
2	3 180 503	3 190 121	3 186 581	3 084 090
3	2 107 060	1 768 042	1 939 976	1 626 280
4	590 958	591 609	522 601	476 504
5	734 823	785 081	759 291	711 394
6	38 730 687	46 023 242	36 596 174	46 316 400
7	1 103 283	1 166 625	1 181 062	1 152 520
8	579 360	613 335	518 255	640 936
9	10 381 677	10 221 419	10 566 786	10 447 165
10	10 522 955	10 636 685	10 120 163	10 233 920
Среднее	8 082 367	8 593 063	7 826 355	9 229 124

Примечание. Наилучшие значения показателей выделены жирным шрифтом.

По индексу инерции метод *k*-средних показал лучшие результаты в одном эксперименте, DBSCAN — в двух, метод агломеративной кластеризации — в трех и предложенный метод — в четырех. По среднему значению индекса инерции среди всех экспериментов лучший результат у метода агломеративной кластеризации. На среднее значение для предложенного метода сильное влияние оказал результат в эксперименте 1, который резко отличается от аналогов. Однако в остальных случаях наблюдается положительная тенденция.

Обсуждение

Рассмотрим ситуации, когда предлагаемый метод стоит использовать для кластеризации линейных отрезков и выявления скрытых полигональных структур,

а в каких случаях делать это менее целесообразно. На рис. 5 продемонстрированы два примера. Светлые линии на рис. 5, *a*, *c* показывают истинные объекты, а жирные — выделенные отрезки.

В первом случае (рис. 5, b) метод корректно создает два кластера из растянутых отрезков, что отражено в двух замкнутых объектах несмотря на то, что боковые стенки зданий находятся близко друг к другу. Во втором случае (рис. 5, d) показан некорректный результат работы метода. Вместо двух объектов метод все отрезки сгруппировал в один кластер с учетом их растяжения. Метод не обнаружил диагональную часть здания в левом верхнем углу и достроил объект как пересечение вертикальной и горизонтальной линий. Кроме этого, метод не учел отсутствие боковых стенок у зданий и в итоге был сформирован один объект.



Рис. 5. Примеры корректного (*a*, *b*) и ошибочного (*c*, *d*) обнаружений полигональных структур

Fig. 5. Examples of correct (a, b) and erroneous (c, d) detection of polygonal structures

Таким образом, метод дает положительный результат, если неявно выраженные структуры содержат хотя бы один отрезок на каждой из своих сторон.

Для развития разработанного метода в дальнейшем планируется использовать предсказательные модели для определения более оптимального направления при растяжении исходных отрезков.

Заключение

В работе предложен метод для кластеризации линейных отрезков, которые образуют неявно выраженные полигональные структуры. Такие структуры появляются на векторных картах в геоинформационных системах при использовании технологий автоматической векторизации пространственных объектов на спутниковых снимках, так как эти технологии несовершенны.

Литература

- Воробьев А.В., Воробьева Г.Р. Геоинформационная система динамической пространственной кластеризации распределенных источников данных // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. 2023. № 64. С. 61–73. https://doi.org/10.17223/19988605/64/7
- Еремеев С.В., Абакумов А.В., Андрианов Д.Е., Богоявленский И.В., Никонов Р.А. Обнаружение многолетних бугров пучения с использованием декомпозиции цифровых моделей рельефа по топологическим признакам // Современные проблемы дистанционного зондирования Земли из космоса. 2023. Т. 20. № 6. С. 129–143. https://doi.org/10.21046/2070-7401-2023-20-6-129-143
- Еремеев С.В., Абакумов А.В., Андрианов Д.Е., Ширабакина Т.А. Метод векторизации спутниковых снимков на основе их разложения по топологическим особенностям // Информатика и автоматизация. 2023. Т. 22. № 1. С. 110–145. https://doi.org/10.15622/ ia.22.1.5
- Lin X., Zhou Y., Liu Y., Zhu C. A comprehensive review of image line segment detection and description: taxonomies, comparisons, and challenges // IEEE Transactions on Pattern Analysis and Machine Intelligence. 2024. V. 46. N 12. P. 8074–8093. https://doi.org/10.1109/ tpami.2024.3400881
- Zhang C., He Y., Fraser C.S. Spectral clustering of straight-line segments for roof plane extraction from airborne LiDAR point clouds // IEEE Geoscience and Remote Sensing Letters. 2018. V. 15. N 2. P. 267–271. https://doi.org/10.1109/LGRS.2017.2785380
- Zang D., Wang J., Zhang X., Yu J. Semantic extraction of roof contour lines from airborne lidar building point clouds based on multidirectional equal-width banding // IEEE Journal of Selected Topics in Applied

В основе метода лежит применение топологических подходов и идей для анализа данных. Показано, что гомотетия позволяет при постепенном растяжении линейных отрезков обнаружить и проанализировать топологические структуры из компонент связности и полигонов. Приведены результаты исследований, в которых на векторных моделях продемонстрирована работоспособность метода, что подтверждает численная оценка качества кластеризации с использованием индексов Жаккара и инерции. Предложенный метод позволяет устранить неполные данные и получить полигональные структуры, что очень необходимо для полноценного анализа и визуализации пространственной информации. Полученные результаты могут быть использованы для улучшения обнаружения неявно выраженных полигональных структур при автоматическом обновлении векторных карт по спутниковым снимкам. Эксперименты в сравнении с аналогами, такими как методы k-средних, DBSCAN и агломеративная кластеризация, показали, что предложенный метод позволяет обнаружить формы кластеров, отличающиеся от результатов существующих подходов в силу их ограничения на поиск наиболее близких точек в метрическом пространстве.

Преимуществом метода является корректное разделение близко расположенных кластеров. В перспективе метод может быть внедрен в технологии векторизации пространственных данных, а также использован при анализе сложной информации городской среды. Программная реализация метода предназначена для обработки данных в геоинформационной системе QGIS и вместе с тестовыми наборами данных доступна по ссылке¹.

¹ Электронный ресурс]. Режим доступа: https://github. com/creems84/Howole-Geo.git (дата обращения: 05.12.2024).

References

- Vorobev A.V., Vorobeva G.R. Geoinformation system for dynamic spatial clustering of distributed data sources. *Tomsk State University Journal of Control and Computer Science*, 2023, no. 64, pp. 61–73. (in Russian). https://doi.org/10.17223/19988605/64/7
- Eremeev S.V., Abakumov A.V., Andrianov D.E., Bogoyavlensky I.V., Nikonov R.A. Detection of perennial heaving mounds in digital elevation model images using decomposition by topological features. *Sovremennye problemy distantsionnogo zondirovaniya Zemli iz* kosmosa, 2023, vol. 20, no. 6, pp. 129–143. (in Russian). https://doi. org/10.21046/2070-7401-2023-20-6-129-143
- Eremeev S.V., Abakumov A.V., Andrianov D.E., Shirabakina T.A. Vectorization method of satellite images based on their decomposition by topological features. *Informatics and Automation*, 2023, vol. 22, no. 1, pp. 110–145. (in Russian). https://doi.org/10.15622/ia.22.1.5
- Lin X., Zhou Y., Liu Y., Zhu C. A comprehensive review of image line segment detection and description: taxonomies, comparisons, and challenges. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2024, vol. 46, no. 12, pp. 8074–8093. https://doi. org/10.1109/tpami.2024.3400881
- Zhang C., He Y., Fraser C.S. Spectral clustering of straight-line segments for roof plane extraction from airborne LiDAR point clouds. *IEEE Geoscience and Remote Sensing Letters*, 2018, vol. 15, no. 2, pp. 267–271. https://doi.org/10.1109/LGRS.2017.2785380
- Zang D., Wang J., Zhang X., Yu J. Semantic extraction of roof contour lines from airborne lidar building point clouds based on multidirectional equal-width banding. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2024, vol. 17, pp. 16316–16328. https://doi.org/10.1109/jstars.2024.3422973

Earth Observations and Remote Sensing. 2024. V. 17. P. 16316–16328. https://doi.org/10.1109/jstars.2024.3422973

- Xin X., Huang W., Zhong S., Zhang M., Liu Z. Xie Z. Accurate and complete line segment extraction for large-scale point clouds // International Journal of Applied Earth Observation and Geoinformation. 2024. V. 128. P. 103728. https://doi.org/10.1016/j. jag.2024.103728
- Miroshnichenko S.Yu., Titov V.S., Dremov E.N., Mosin S.A. Hough transform application to digitize rectangular spatial objects on aerospace imagery // SPIIRAS Proceedings. 2018. N 5 (60). P. 189– 215. https://doi.org/10.15622/sp.60.7
- Teplyakov L., Erlygin L., Shvets E. LSDNet: Trainable modification of LSD algorithm for real-time line segment detection // IEEE Access. 2022. V. 10. P. 45256–45265. https://doi.org/10.1109/ ACCESS.2022.3169177
- Shili H. Clustering in big data analytics: a systematic review and comparative analysis (review article). Scientific and Technical Journal of Information Technologies, Mechanics and Optics. 2023. V. 23. N 5. P. 967–979. https://doi.org/10.17586/2226-1494-2023-23-5-967-979
- Ikotun A.M., Ezugwu A.E., Abualigah L., Abuhaija B., Heming J. K-means clustering algorithms: a comprehensive review, variants analysis, and advances in the era of big data // Information Sciences. 2023. V. 622. P. 178–210. https://doi.org/10.1016/j.ins.2022.11.139
- Ester M., Kriegel H.-P., Sander J., Xu X. A density-based algorithm for discovering clusters in large spatial databases with noise // Proc. of the 2nd International Conference on Knowledge Discovery and Data Mining (KDD-96). 1996. P. 226–231.
- Jang J.H., Hong K.S. Fast line segment grouping method for finding globally more favorable line segments // Pattern Recognition. 2002.
 V. 35. N 10. P. 2235–2247. https://doi.org/10.1016/S0031-3203(01)00175-3
- Gao J., Langberg M., Schulman L.J. Clustering lines in highdimensional space: Classification of incomplete data // ACM Transactions on Algorithms. 2010. V. 7. N 1. P. 8. https://doi. org/10.1145/1868237.1868246
- Айдагулов Р.Р., Главацкий С.Т., Михалёв А.В. Модели кластеризации // Фундаментальная и прикладная математика. 2020. Т. 23. № 2. С. 17–36.
- Kang Y., Wu K., Gao S., Ng I., Rao J., Ye S., Zhang F., Fei T. STICC: a multivariate spatial clustering method for repeated geographic pattern discovery with consideration of spatial contiguity // International Journal of Geographical Information Science. 2022. V. 36. N 8. P. 1518–1549. https://doi.org/10.1080/13658816.2022.2053980
- Alomari H.W., Al-Badarneh A.F., Al-Alaj A., Khamaiseh S.Y. Enhanced approach for agglomerative clustering using topological relations // IEEE Access. 2023. V. 11. P. 21945–21967. https://doi. org/10.1109/access.2023.3252374
- Edelsbrunner H., Letscher H., Zomorodian A. Topological persistence and simplification // Discrete & Computational Geometry. 2002. V. 28. N 4. P. 511–533. https://doi.org/10.1007/s00454-002-2885-2
- Абакумов А.В., Еремеев С.В. Оценка возможности применения метода декомпозиции изображений по топологическим признакам для уменьшения энтропии при их сжатии // Научнотехнический вестник информационных технологий, механики и оптики. 2023. Т. 23. № 6. С. 1152–1161. https://doi: 10.17586/2226-1494-2023-23-6-1152-1161
- Xu F., Beard K. A unifying framework for analysis of spatial-temporal event sequence similarity and its applications // ISPRS International Journal of Geo-Information. 2021. V. 10. N 9. P. 594. https://doi. org/10.3390/ijgi10090594
- Rykov A., De Amorim R.C., Makarenkov V., Mirkin B. Inertia-based indices to determine the number of clusters in K-Means: an experimental evaluation // IEEE Access. 2024. V. 12. P. 11761–11773. https://doi.org/10.1109/ACCESS.2024.3350791

Автор

Еремеев Сергей Владимирович — кандидат технических наук, доцент, доцент, Муромский институт (филиал) ΦГБОУ ВО «Владимирский государственный университет имени Александра Григорьевича и Николая Григорьевича Столетовых», Муром, 602264, Российская Федерация, SC 56673892800, https://orcid.org/0000-0001-8482-1479, sv-eremeev@yandex.ru

- Xin X., Huang W., Zhong S., Zhang M., Liu Z. Xie Z. Accurate and complete line segment extraction for large-scale point clouds. *International Journal of Applied Earth Observation and Geoinformation*, 2024, vol. 128, pp. 103728. https://doi.org/10.1016/j. jag.2024.103728
- Miroshnichenko S.Yu., Titov V.S., Dremov E.N., Mosin S.A. Hough transform application to digitize rectangular spatial objects on aerospace imagery. *SPIIRAS Proceedings*, 2018, no. 5 (60), pp. 189– 215. https://doi.org/10.15622/sp.60.7
- Teplyakov L., Erlygin L., Shvets E. LSDNet: Trainable modification of LSD algorithm for real-time line segment detection. *IEEE Access*, 2022, vol. 10, pp. 45256–45265. https://doi.org/10.1109/ ACCESS.2022.3169177
- Shili H. Clustering in big data analytics: a systematic review and comparative analysis (review article). *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 2023, vol. 23, no. 5, pp. 967–979. https://doi.org/10.17586/2226-1494-2023-23-5-967-979
- Ikotun A.M., Ezugwu A.E., Abualigah L., Abuhaija B., Heming J. K-means clustering algorithms: a comprehensive review, variants analysis, and advances in the era of big data. *Information Sciences*, 2023, vol. 622, pp. 178–210. https://doi.org/10.1016/j.ins.2022.11.139
- Ester M., Kriegel H.-P., Sander J., Xu X. A density-based algorithm for discovering clusters in large spatial databases with noise. *Proc. of the 2nd International Conference on Knowledge Discovery and Data Mining (KDD-96)*, 1996, pp. 226–231.
- Jang J.H., Hong K.S. Fast line segment grouping method for finding globally more favorable line segments. *Pattern Recognition*, 2002, vol. 35, no. 10, pp. 2235–2247. https://doi.org/10.1016/S0031-3203(01)00175-3
- Gao J., Langberg M., Schulman L.J. Clustering lines in highdimensional space: Classification of incomplete data. ACM Transactions on Algorithms, 2010, vol. 7, no. 1, pp. 8. https://doi. org/10.1145/1868237.1868246
- Aidagulov R.R., Glavatsky S.T., Mikhalev A.V. Clustering models. Journal of Mathematical Sciences, 2022, vol. 262, no. 5. pp. 603–616. (in Russian). https://doi.org/10.1007/s10958-022-05841-9
- Kang Y., Wu K., Gao S., Ng I., Rao J., Ye S., Zhang F., Fei T. STICC: a multivariate spatial clustering method for repeated geographic pattern discovery with consideration of spatial contiguity. *International Journal of Geographical Information Science*, 2022, vol. 36, no. 8, pp. 1518–1549. https://doi.org/10.1080/13658816.20 22.2053980
- Alomari H.W., Al-Badarneh A.F., Al-Alaj A., Khamaiseh S.Y. Enhanced approach for agglomerative clustering using topological relations. *IEEE Access*, 2023, vol. 11, pp. 21945–21967. https://doi. org/10.1109/access.2023.3252374
- Edelsbrunner H., Letscher H., Zomorodian A. Topological persistence and simplification. *Discrete & Computational Geometry*, 2002, vol. 28, no. 4, pp. 511–533. https://doi.org/10.1007/s00454-002-2885-2
- Abakumov A.V., Eremeev S.V. Assessing the possibility of using the method of image decomposition based on topological features to reduce entropy during their compression. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 2023, vol. 23, no. 6, pp. 1152–1161. (in Russian). https://doi. org/10.17586/2226-1494-2023-23-6-1152-1161
- Xu F., Beard K. A unifying framework for analysis of spatial-temporal event sequence similarity and its applications. *ISPRS International Journal of Geo-Information*, 2021, vol. 10, no. 9, pp. 594. https://doi. org/10.3390/ijgi10090594
- Rykov A., De Amorim R.C., Makarenkov V., Mirkin B. Inertia-based indices to determine the number of clusters in K-Means: an experimental evaluation. *IEEE Access*, 2024, vol. 12, pp. 11761– 11773. https://doi.org/10.1109/ACCESS.2024.3350791

Author

Sergey V. Eremeev — PhD, Associate Professor, Associate Professor, Murom Institute (Branch) of Vladimir State University, Murom, 602264, Russian Federation, Sc 56673892800, https://orcid.org/0000-0001-8482-1479, sv-eremeev@yandex.ru

Received 02.11.2024 Approved after reviewing 25.12.2024 Accepted 30.03.2025 **VİTMO**

НАУЧНО-ТЕХНИЧЕСКИЙ ВЕСТНИК ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ, МЕХАНИКИ И ОПТИКИ март–апрель 2025 Том 25 № 2 http://ntv.ifmo.ru/ SCIENTIFIC AND TECHNICAL JOURNAL OF INFORMATION TECHNOLOGIES, MECHANICS AND OPTICS March–April 2025 Vol. 25 № 2 http://ntv.ifmo.ru/en/ ISSN 2226-1494 (print) ISSN 2500-0373 (online)

doi: 10.17586/2226-1494-2025-25-2-273-285 УДК 004.93

Структурный анализ изображения и адаптивный резонанс в искусственных нейронных сетях (обзорная статья) Вадим Ростиславович Луцив^{1⊠}, Мария Анатольевна Михалькова², Валерия Олеговна Ячная³

1.3 Санкт-Петербургский государственный университет аэрокосмического приборостроения, Санкт-Петербург, 190000, Российская Федерация

^{2,3} Институт физиологии имени И.П. Павлова РАН, Санкт-Петербург, 199034, Российская Федерация

¹ vluciv@mail.ru^{\equiv}, https://orcid.org/0000-0003-1160-3758

² mikhalkova maria@infran.ru, https://orcid.org/0000-0002-7186-5744

³ yachnaya_valeria@infran.ru, https://orcid.org/0000-0002-7239-2096

Аннотация

Введение. Наблюдаемый мир иерархически структурирован. Механизм взаимодействия встречных информационных потоков в иерархически организованной системе был назван «адаптивным резонансом» и успешно промоделирован полвека назад в искусственной нейронной сети, анализирующей обонятельные стимулы, а затем — при распознавании изображений. В дальнейшем полезный потенциал совместного применения в нейронных сетях принципов структурного анализа и адаптивного резонанса долго не использовался. В последние годы эти принципы были применены в капсульных сетях и позволили превзойти результаты, достигнутые другими моделями нейронных сетей. Это показало необходимость систематизировать использованные пути их реализации. Метод. Выполнен анализ научно-технических работ, опубликованных за последние 50 лет по применению принципов структурного анализа и адаптивного резонанса при обработке изображений в искусственных нейронных сетях. Основные результаты. Сравнительный анализ подтвердил эффективность применения этих принципов при решении задач автоматической обработки изображений и выявил пути их наиболее эффективной реализации в искусственных нейронных сетях. Обсуждение. В связи с успехами, достигнутыми сверточными сетями при распознавании изображений, их разработчики не применяли связанные с особенностями организации наблюдаемого мира принципы структурного анализа и адаптивного резонанса. Однако использование данных принципов позволяет более эффективно решать задачи обработки изображений, и дальнейшие исследования в области искусственных нейронных сетей целесообразно проводить в этом направлении.

Ключевые слова

искусственные нейронные сети, структурный анализ, адаптивный резонанс, иерархическая организация, сверточные сети, капсульные сети, трансформеры

Ссылка для цитирования: Луцив В.Р., Михалькова М.А., Ячная В.О. Структурный анализ изображения и адаптивный резонанс в искусственных нейронных сетях (обзорная статья) // Научно-технический вестник информационных технологий, механики и оптики. 2025. Т. 25, № 2. С. 273–285. doi: 10.17586/2226-1494-2025-25-2-273-285

Image structural analysis and adaptive resonance in artificial neural networks (Review paper)

Vadim R. Lutsiv¹, Maria A. Mikhalkova², Valeria O. Yachnaya³

^{1,3} Saint Petersburg State University of Aerospace Instrumentation, Saint Petersburg, 190000, Russian Federation ^{2,3} Pavlov Institute of Physiology of the Russian Academy of Sciences, Saint Petersburg, 199034, Russian Federation

¹ vluciv@mail.ru^{\equiv}, https://orcid.org/0000-0003-1160-3758

² mikhalkova_maria@infran.ru, https://orcid.org/0000-0002-7186-5744

³ yachnaya valeria@infran.ru, https://orcid.org/0000-0002-7239-2096

© Луцив В.Р., Михалькова М.А., Ячная В.О., 2025

Abstract

The observable world is hierarchically structured. The interaction technique of counter motion of information flows in the hierarchically organized systems was named "the adaptive resonance" and successfully modeled in artificial neural network of olfactory stimuli analysis and then applied for image recognition. The usefulness of application of the principles of hierarchical structural analysis and adaptive resonance was then forgotten for a long time. Recently, these principles were applied again in the capsule neural networks that outperformed the best modern models of other neural networks. This shows the necessity of systematizing the ways of practical implementation of these principles. The experience of application of structural analysis and adaptive resonance in the tasks of image recognition by artificial neural networks was inspected through the scientific and technical materials published last half-century. The comparative analysis carried out confirmed the efficiency of application of these principles in automatic image processing. The methods of most efficient realization of structural analysis and adaptive resonance in artificial neural networks were also determined. While the successful results of image recognition were reached by convolutional neural networks, their developers consigned to oblivion the principles of structural analysis and adaptive resonance following from the organization peculiarities of observable environment. However, the come-back to application of these principles would result in additional success in solving the tasks of image processing; thus, the further investigation of artificial neural networks is worth to be carried out in this area.

Keywords

artificial neural networks, structural analysis, adaptive resonance, hierarchical organization, convolutional networks, capsule networks, transformers

For citation: Lutsiv V.R., Mikhalkova M.A., Yachnaya V.O. Image structural analysis and adaptive resonance in artificial neural networks (Review paper). *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 2025, vol. 25, no. 2, pp. 273–285 (in Russian). doi: 10.17586/2226-1494-2025-25-2-273-285

Введение

Организация окружающего мира неоднородна. В нем можно выделить объекты и подобъекты, отличающиеся по определенным свойствам друг от друга.

Как живые зрительные системы, так и имитирующие их работу алгоритмы компьютерного зрения учитывают системное взаимодействие информации, извлеченной из анализируемых изображений на разных иерархических уровнях ее обработки. Структурированная информация более низкого иерархического уровня анализируется на более высоком уровне как единое (не разделенное на части) целое, — и качество этого анализа также соответствует более высокому уровню. В свою очередь, результаты, сформированные при анализе более высокого уровня, передаваемые «вниз по иерархии», позволяют скорректировать в нужную сторону методы и результаты анализа информации на более низких иерархических уровнях. Частный вариант такого двунаправленного анализа был назван — «адаптивным резонансом» [1-3]. Многие разработчики, продуктивно реализуя этот принцип, адаптивным резонансом его не называют, однако этот термин по своему смыслу хорошо соответствует взаимодействию в иерархической системе встречных (восходящих и нисходящих) потоков информации, приводящему к получению нового ее качества. Исходя из этого, такое взаимодействие в настоящей работе будет называться адаптивным резонансом. Это позволит также упорядочить терминологию, применяемую разными авторами, и сделать более объективным сравнительный анализ их достижений.

Структурный анализ информации в искусственных нейронных сетях: первые шаги к применению

Современный «бум» в исследовании искусственных нейронных сетей начался с момента опубликования в работе [4] простого и эффективного алгоритма обучения многослойных перцептронов методом обратного распространения ошибки. Однако в скором времени стало очевидным, что представление изображения как единого целого не позволяет многослойным перцептронам распознавать образы со сложными локальными изменениями формы, например изображения рукописных символов. В связи с этим японский исследователь Ямада в многослойном перцептроне вместо полносвязных слоев нейронов (в которых каждый нейрон последующего слоя соединен с каждым нейроном предыдущего) применил локальные связи нейронов между соседними слоями: каждый нейрон последующего слоя был связан только с локальной группой соседних нейронов предыдущего [5]. Такой подход позволил независимо анализировать локальные детали изображения и повысить надежность распознавания рукописных символов с 98,3 % до 98,8 %. Отметим, что данное решение можно считать шагом к структурному анализу, хотя пока еще в явном виде отдельные структурные элементы изображения не выделялись.

Структурные методы распознавания имеют существенное преимущество перед методами, анализирующими изображение как единое целое: отдельные элементы изображения могут не иметь строго фиксированного положения, но при этом локально пространственно смещаться друг относительно друга при условии, что величина смещения находится в допустимых пределах. Такая возможность была реализована К. Фукусимой в работе [6] в сети «Неокогнитрон». Отметим, что выделяемые сетью «Неокогнитрон» структурные элементы могли в различных анализируемых изображениях немного изменять свою форму. В то же время в этой сети не были реализованы ни целенаправленное выделение структурных элементов заданных типов (локально устойчиво выделялись элементы, не соответствующие какому-то заданному классу объектов), ни описание в явном виде их взаимного пространственного положения.

С момента разработки сети AlexNet [7] в 2011 году прошло уже почти полтора десятилетия. За это время было создано огромное количество модификаций сверточных сетей, однако большинство из них не реализуют структурный анализ изображения в явном виде: структурные элементы осмысленной формы не выделяются, и их взаимное положение в явном виде не описывается, поэтому в настоящей работе они не будут рассмотрены.

В работе [8] К. Фукусимой наряду с неявным использованием структурного анализа был применен метод адаптивного резонанса. Однако он был не первым исследователем, применившим адаптивный резонанс в искусственных нейронных сетях.

Первое применение адаптивного резонанса в искусственных нейронных сетях

Термин «адаптивный резонанс» был впервые применен С. Гроссбергом в работе [1] при анализе процессов обработки информации (в частности, обонятельных стимулов) в нейронах коры головного мозга. В работе [2] был представлен доклад о создании искусственной нейронной сети ART-1, реализующей при обработке информации принципы адаптивного резонанса, описанные в [1]. В [3] представлены результаты разработки более известной и совершенной искусственной нейронной сети ART-2, ориентированной на обработку не только бинарных, но и аналоговых сигналов. Стоит отметить, что, формально, в этих нейронных сетях усилия разработчиков направлены на автоматическое формирование «эталонных описаний» (адаптивных фильтров), объединяющих входные образы в группы, характеризующиеся сходными признаками (на адаптивную категоризацию образов). Однако, фактически, эта настройка выполняется путем пропускания входных сигналов («снизу вверх») через адаптивные фильтры (выполняется распознавание входных образов), а параметры фильтров модифицируются под управлением «ожиданий», передаваемых «сверху вниз» в результате выполненного распознавания входных сигналов, т. е. происходит распознавание сигналов с применением принципа адаптивного резонанса.

В отличие от результатов, полученных в работе [2], в [8] применен принцип адаптивного резонанса в явном виде для распознавания изображений. В работе [8] такой принцип назван «селективным вниманием», хотя автор знаком с методом адаптивного резонанса (упоминает работу [2] в списке литературы по нейронным сетям). С применением адаптивного резонанса К. Фукусиме в работе [8] удалось при выполнении распознавания автоматически разделять наложенные друг на друга изображения, подавлять искажающий шум и восстанавливать утерянные детали. На рис. 1 представлена вычислительная схема, иллюстрирующая направления «прямого» прохождения информации о входном изображении U_{c0} (при выполнении его поэтапной фильтрации и распознавания) и «обратного» прохождения сигналов, сформированных на более высоких иерархических уровнях анализа изображения и корректирующих обработку информации на более низких иерархических уровнях. Схема дана в том виде, как она была представлена в работе [8], не существенные для обсуждения детали нами опущены. Рис. 2 [8] иллюстрирует постепенную коррекцию анализируемого изображения под действием операций адаптивного резонанса, выполняемых согласно вычислительной схеме (рис. 1).

Нейронные сети, впервые в явном виде применившие методы структурного анализа и адаптивного резонанса

Дополнительные возможности для структурного анализа появляются, если в отличие от сверточных сетей Фукусимы [6, 8] и Хинтона [7] нейронной сетью анализируются целенаправленно выделенные структурные элементы заданных типов и их взаимные пространственные положения. Именно такие нейронные сети были разработаны в 80-х годах прошлого века. В основу этих разработок была положена нейронная сеть Хопфилда–Танка [9]. В 90-х годах в Государственном оптическом институте им. С.И. Вавилова в Санкт-Петербурге сеть Хопфилда–Танка использовалась сначала для установления взаимного соответствия тексе-



Puc. 1. Схема прохождения сигналов «прямого» и «обратного» распространений в нейронной сети [8].
 U_{s1}, U_{s2}, U_{s3} — матрицы (слои) нейронов S-типа, реализующие фильтры для узнавания локальных деталей изображения;
 U_{c0}, U_{c1}, U_{c2}, U_{c3} — матрицы (слои) нейронов «подвыборки»; W_{c0}, W_{c1}, W_{c2}, W_{s1}, W_{s2}, W_{s3} — матрицы нейронов, корректирующие распознавание локальных деталей на основе результатов анализа более высоких иерархических уровней;
 P — матрица пикселов входного изображения; x — сигнал «нет отклика»

Fig. 1. Diagram of forward and backward signal propagation in the Fukushima's neural network [8].

U_{s1}, U_{s2}, U_{s3} — stand for the neuron matrices (layers) of S-type implementing the filters for local image details recognition;
 U_{c0}, U_{c1}, U_{c2}, U_{c3} — stand for the neuron matrices of pooling layers; W_{c0}, W_{c1}, W_{c2}, W_{s1}, W_{s2}, W_{s3} — stand for the neuron matrices correcting the local details recognition processes based on the analysis results reached at the higher hierarchical levels;
 P — matrix of input image pixels; x — "no response" signal

Научно-технический вестник информационных технологий, механики и оптики, 2025, том 25, № 2 Scientific and Technical Journal of Information Technologies, Mechanics and Optics, 2025, vol. 25, no 2

	0	1	2	3	4	5	• 6	7	8	9	10	11	12	1 3	14	15
U _{c0}	2	2	24	24	2	4	2	2	2	2	2	24	2	2	4	
W _{c0}	4	4	4	4	4	4		2	2	2	2	2	2	I		

Рис. 2. Процесс коррекции пикселов анализируемого изображения U_{c0} под управлением «гипотезы» W_{c0} об ожидаемых их значениях, итеративно улучшаемой на более высоких иерархических уровнях анализа [8].

1-15 — номера итераций обработки изображения; «▼» — моменты переключения внимания

Fig. 2. Correction of the pixels of input image U_{c0} using the hypothesis W_{c0} about their expected values that is iteratively enhanced at the higher hierarchical levels of analysis [8] (the sequential digits 1–15 stand for the image processing iteration numbers, " ∇ " shows the attention switching moments)

лов в стереопарах [10, 11], а затем для структурного сопоставления контурных структурных описаний распознаваемого изображения и эталона [12–15].

Двумерный решающий слой нейронной сети Хопфилда-Танка представляет собой динамическую матрицу соответствия, строкам которой соответствуют структурные элементы первого объекта, а столбцам структурные элементы второго. Если параметры *i*-го структурного элемента первого объекта совпадают с заданной точностью с параметрами *j*-го элемента второго, в матрице соответствия становится активным нейрон на пересечении і-й строки и ј-го столбца, причем активность нейрона пропорциональна структурному сходству соответствующих ему сопоставленных элементов. В качестве сопоставляемых структурных элементов сеть использовала результаты аппроксимации контуров, выделенных на двух сопоставляемых изображениях, отрезками прямых линий и углами между ними (как показано в [15], такие структурные описания наиболее устойчивы к естественным изменениям наблюдаемых сцен).

Отметим, что нейронная сеть Хопфилда–Танка в наиболее полной мере реализовала преимущества структурного анализа. Использовались наиболее устойчивые структурные элементы целенаправленно выделенных типов, и при их сопоставлении учитывалось их взаимное пространственное положение. Действительно, результаты компьютерного моделирования [12] показали, что выполняемое сетью структурное сопоставление оказалось очень устойчивым к взаимному геометрическому преобразованию сопоставляемых изображений, шуму и искажению параметров формы структурных элементов.

Однако сложность представленной в [12] нейронной сети (количество активно действующих связей между нейронами) было приблизительно пропорционально четвертой степени числа структурных элементов. Эта проблема была частично решена построением иерархической нейронной сети Хопфилда–Танка для сопоставления двухуровневых иерархических структурных описаний [13, 14]. Сопоставляемые изображения делились на локальные регионы. На нижнем иерархическом уровне сопоставлялись контурные структурные описания отдельных регионов, а на верхнем уровне сопоставляемыми структурными элементами были сами локальные регионы. Кроме того, в данной иерархической сети был реализован принцип адаптивного резонанса. В отличие от сетей Гроссберга и Фукусимы, где с использованием этого принципа настраивались весовые коэффициенты суммирования пикселов, в данном случае адаптивный резонанс применялся другим способом: корректировался состав групп структурных элементов, входящих в локальные регионы, сформированные на изображении.

Применение в [13, 14] иерархического структурного описания и адаптивного резонанса в двухуровневой сети Хопфилда-Танка позволило работать с увеличенным числом контурных структурных элементов, однако, в целом, трудность увеличения их количества преодолена не была. Проблема была решена путем перехода от использования нейронной сети к применению при структурном сопоставлении метода оптимизированного обхода дерева решений [16]. В реализующем дерево решений алгоритме [17] также использовалось сопоставление иерархических контурных структурных описаний с применением принципа адаптивного резонанса (пример коррекции этим методом контурных структурных описаний, показан на рис. 3), а в [18] авторы успешно расширили этот алгоритм для сопоставления изображений трехмерных сцен. В исследовании [17] (расширенное его описание затем было приведено в [15]) была достигнута возможность сопоставления структурных описаний чрезвычайно сильно измененных изображений, как это методом их мозаичного наложения проиллюстрировано на рис. 4. Это продемонстрировало возможности методов структурного анализа и адаптивного резонанса, однако примененный подход, основанный на обходе дерева решений, более не был связан с нейронными сетями.

Первая сверточная сеть, выполняющая иерархический структурный анализ изображения

Как было отмечено, сети Фукусимы и Хинтона и многие другие сверточные сети выделяют локальные элементы изображения, но в явном виде его структурного анализа не выполняют. При развитии сверточных сетей Росс Гиршик с коллегами представил на конференции CVPR-2015 искусственную сеть, в явном виде реализующую структурный анализ изображения [19]. В качестве основы разработки был использован алгоритм обнаружения объектов с раздельным анализом их частей [20], который занял второе после сверточных сетей место в соревнованиях компьютерных программ



Рис. 3. Коррекция сопоставляемых структурных описаний с применением принципа адаптивного резонанса. Пары сопоставляемых контурных структурных описаний: до (*a*, *b*) и после (*c*, *d*) коррекции. Цифрами 1, 2 и 3 помечены группы структурных элементов, соответствующие друг другу в исходной и исправленной паре описаний

Fig. 3. Correction of matched structural descriptions based on application of adaptive resonance: a pair of initial structural descriptions (*a*) and (*b*); the same descriptions corrected based on adaptive resonance (*c*) and (*d*). The digits 1, 2 and 3 mark the groups of structural elements corresponding to each other in the initial and in the corrected pair of descriptions



Рис. 4. Структурное сопоставление снимков: радиолокационного (*a*) и выполненного в видимом оптическом диапазоне (*c*); мозаичное наложение (регистрация) сопоставленных снимков (*b*) в общей системе координат

Fig. 4. Structural matching of the radar image (*a*) and optical image (*c*) of the same terrain; the matching result (*b*) is shown by registration of these images in mosaic fashion in a common coordinate system

распознавания изображений 2011 года¹. Утверждается также, что использованная в [19] технология может быть применена и для реализации других неитеративных алгоритмов обнаружения объектов с раздельным анализом частей, например, как в работе [21]. Однако, если решение принимается итеративными алгоритмами, такими как [22], их реализация методами, предложенными в [19], требует дальнейшего переосмысления.

В работе [19] для удобства сравнения с алгоритмом-прототипом [20] технология сверточных сетей использована только для формирования векторов признаков. Конечная их классификация, аналогично [20], выполнена с помощью машины опорных векторов. В продолжение этого исследования в работе [23] сформирована единая сверточная сеть, выполняющая как формирование векторов признаков вышеописанным способом, так и их классификацию. Как утверждают авторы [19], их решение на основе технологии сверточных сетей позволило существенно (с 33,4 % до 42 % на основе базы изображений PASCAL VOC 2010) улучшить среднюю точность обнаружения объектов по сравнению с результатами прототипа [20]. В ряде случаев дальнейшее улучшение точности обнаружения было получено с полным применением технологии нейронных сетей [23] (для формирования векторов признаков и их классификации). Однако результаты обнаружения объектов алгоритмами [19, 23] оказались значительно хуже средней точности обнаружения (53,7 %), достигнутой сетью R-CNN [24]. С тех пор было предложено еще много других более эффективных сверточных сетей, поэтому алгоритм обнаружения объектов с раздельным анализом частей и его реализации нейронными сетями теперь упоминаются нечасто. Отметим, что хотя в [19] впервые за долгое время и был реализован в нейронной сети иерархический структурный анализ изображения, о преимуществах применения адаптивного резонанса разработчики искусственных нейронных сетей еще долгое время не вспоминали.

Капсульные сети

Наконец, во время победного шествия сверточных нейронных сетей около десяти лет назад в своем блоге²

¹[Электронный ресурс]. http://www.image-net.org (дата обращения: 29.06.2023).

²Электронный pecypc]. https://www.reddit.com/user/ geoffhinton/ (дата обращения: 06.08.2024).

Джеффри Хинтон отметил, что применение операции подвыборки (pooling) было большой ошибкой: оно уничтожает информацию о том, где находятся выделенные сверточной сетью детали изображения. Вместо этого он предложил кодировать информацию о пространственном положении выделенных деталей относительно объекта в целом и реализовать такой пространственный анализ в специальных «капсулах», являющихся расширенной версией обычного искусственного нейрона, применяемого в сверточных сетях [25]. Особенности пространственного положения и начертания деталей изображения описываются в капсулах специальным вектором, длина которого также кодирует информацию о вероятности присутствия на изображении в данном положении деталей заданных классов.

В то же время в работе [25] матрицы, задающие положения деталей, должны кем-то задаваться извне. Остается также не до конца решенной проблема присутствия в изображении нескольких подлежащих распознаванию объектов одновременно. Эти проблемы решаются Джеффри Хинтоном, Сарой Сэбоур и Николасом Фросстом в более поздней работе [26], использующей динамическую маршрутизацию информации между капсулами, которая реализует ничто иное как принцип адаптивного резонанса (правда, данный термин в [26] не использован). Отметим, что через 30 лет после опубликования работы Фукусимы [8] разработчики нейронных сетей снова вернулись к указанной тематике. Причем в работе [26] результаты применения капсульной сетью иерархического структурного анализа и адаптивного резонанса иллюстрируются, как и у Фукусимы (рис. 2), разделением наложенных друг на друга изображений рукописных цифр (цветные цифры на рис. 5 [26]). Правда, теперь это решение получено на качественно более высоком уровне: метод адаптивного резонанса применяется не просто для регулирования коэффициентов передачи пикселов локальных деталей изображения объекта, а для принятия решения, соответ-



Рис. 5. Примеры раздельного распознавания сетью наложенных цифр: L — истинные цифры; R — цифры, реконструированные сетью; P и «*» — ошибки, обусловленные неправильной классификацией или реконструкцией, или предъявлением сети неизвестных ей символов

Fig. 5. Examples of separate recognition of superposed digits: L — the true digits; R — the ones reconstructed by network; P and "*" stand for the errors caused by incorrect classification and reconstruction or by unfamiliar input symbols respectively ствуют ли эти детали его структурному описанию. При этом структурное описание уже приобрело формально-логическую форму (описывается в [26] в виде аффинного преобразования координат деталей в систему координат всего объекта), хотя эти детали пока все еще не имеют устойчивого смыслового значения и в скрытом виде выделяются в изображении капсульной сетью.

Вычислительная схема, реализующая динамическую маршрутизацию (адаптивный резонанс) в капсульной сети [26] представлена на рис. 6. Вектор \mathbf{u}_i кодирует вероятность обнаружения і-ой капсулой предыдущего иерархического уровня соответствующего ей (і-го) структурного элемента в позиции, соответствующей направлению этого вектора; W_{іі} — матрица пересчета вектора описания і-го найденного структурного элемента в вектор **ŭ**_{ili} в системе координат всего объекта, включающего всего М таких структурных элементов. Результатом взвешенного суммирования векторов **й**_{*iii*}, описывающих наличие и положение структурных элементов, найденных капсулами нижнего уровня, является вектор s_i, описывающий прогнозируемое положение и вероятность обнаружения объекта *j*-го класса. Результатом нормализации вектора s_i , аналогичной вычислению функции сигмоиды в сверточных сетях, является выходной вектор v_i j-ой капсулы высшего иерархического уровня. Весовой коэффициент c_{ii}, вычисленный пропорционально скалярному произведению векторов $\mathbf{\tilde{u}}_{i|i}$ и \mathbf{s}_i , показывает соответствие пространственного положения найденного *i*-го структурного элемента пространственному положению объекта как единого целого. При некорректном пространственном положении найденного элемента соответствующий ему весовой коэффициент с_{іі} уменьшается, уменьшая таким образом влияние этого элемента на принятие решение об обнаружении объекта *j*-го класса.

Авторы капсульной сети утверждают [26], что по сравнению со сверточными сетями традиционной архитектуры они получили практически одинаковую точность распознавания изображений на базе данных MNIST (99,3 % против 99,2 %), но существенно улучшили точность на базе данных AFFNIST (79 % против 66 %). Это свидетельствует о более эффективной работе их сети с геометрически преобразованными изображениями, содержащимися в базе AFFNIST, и подтверждает эффективность явного применения в капсулах иерархического структурного анализа, поддержанного применением адаптивного резонанса.

Подобно тому, как победа сверточной сети [7] в соревновании ImageNet воодушевила исследователей к разработке целой плеяды сверточных сетей, вслед за опубликованием в [25, 26] идеи капсульных сетей появляется заметное количество работ, развивающих идею применения капсул. Отметим некоторые из таких работ.

В работе [27] были разрешен ряд трудностей, характерных для предыдущей капсульной сети [26], но главной особенностью новой сети стало изменение пути реализации динамической маршрутизации. Она выполняется в [27] с применением метода ожидания-максимизации, что улучшило возможность различения похожих объектов. Ошибка классификации уменьшилась



Рис. 6. Реализация динамической маршрутизации (адаптивного резонанса) *j*-ой капсулой высшего иерархического уровня при обнаружении объекта, соответствующего *j*-го класса

Fig. 6. Realization of dynamic routing (adaptive resonance) by the *j*-th capsule detecting an object of *j*-th type in the higher hierarchical level

по сравнению с предыдущей капсульной сетью [26] (1,4 % против 2,7 %) и значительно сильнее уменьшилась по сравнению со сверточными сетями новейшей архитектуры (для последних ошибка составляла 5,2 %). Сравнение производилось на основе изображений из базы данных SmallNORB [28].

Важной при анализе изображений, включающих не один, а несколько объектов, является необходимость принятия решения, к какому из объектов принадлежат найденные в изображении составляющие их части. Эта проблема решается в иерархическом капсульном автоэнкодере (stacked capsule autoencoder) [29], который в отличие от ранее разработанной капсульной сети [26] обучается «без учителя». «Иерархическая» структура автоэнкодера имеет два уровня. Нижний уровень отвечает за обнаружение и определение параметров пространственного положения отдельных частей, составляющих присутствующие в изображении объекты. Верхний уровень определяет отдельные объекты, состоящие из обнаруженных частей, и уточняет пространственное положение последних. Нижний уровень автоэнкодера самообучается формированию эталонных описаний частей, составляющих объекты, и определению пространственных положений этих частей, максимизируя вероятность правильного восстановления входного изображения из выученных эталонных описателей их формы и параметров их пространственного положения. Аналогичным образом обучаются «без учителя» и капсулы верхнего иерархического уровня автоэнкодера. При обучении максимизируется правдоподобие наличия и параметров положения частей, обнаруженных на нижнем уровне и соответствующих каждому из объектов, существующих в изображении. Максимизируется также правдоподобие соответствия композиции пикселов входного изображения наличию в нем объектов, состоящих из обнаруженных частей.

«Изюминкой» верхнего иерархического уровня автоэнкодера [29] является применение в нем трансформера (set transformer) [30], выполняющего назначение «хаотично» найденных в изображении частей предположительно существующим в нем объектам. Трансформер [30] распределяет обнаруженные части по соответствующим им объектам, выполняя их «адаптивную кластеризацию».

По аналогии с разработанной капсульной сетью [26], иерархический капсульный автоэнкодер [29] применяет при анализе изображения принцип адаптивного резонанса. Однако теперь в [29] не капсулы, относящиеся к частям, предсказывают наличие и положение объекта соответствующего класса, а сами объекты «голосуют» за существование относящихся к ним частей с соответствующими пространственными положениями. Резюмируя рассмотрение капсульной сети [29], отметим, что структурный анализ все еще выполняется в ней в не полностью явном виде. Выделяемые на изображении структурные элементы (части объектов) не имеют устойчивого смыслового значения и в скрытом виде выделяются капсульной сетью в качестве часто встречаемых (характерных) элементов изображений обучающего набора. Например, в [29] в качестве таких структурных элементов упоминаются не имеющие осмысленной формы локальные элементы контуров рукописных символов.

Также в работе [29] отмечено, что авторам удалось при классификации изображений из базы данных MNIST немного превзойти точность наилучших решений [31] распознавания изображений при использовании классификаторов, обученных «без учителя» (99 % против 98,7 %).

Еще одним вариантом реализации капсульных сетей является применение «канонических капсул» (canonical capsules). В работе [32] с их помощью выполняется разделение облака точек в трехмерном пространстве на группы, выполняющие роль структурных элементов, однако в дальнейшем для структурного анализа рассматриваемой трехмерной сцены эти группы никак не используются.

Со времени публикации работы [26] появилось большое количество исследований по капсульным сетям. Обратимся теперь к трансформерам, которые в некоторых приложениях выполняют более полноценный, чем капсульные сети Хинтона структурный анализ.

Трансформеры

Появление трансформеров обычно связывают с опубликованием работы [33], в заголовке которой уже сформулирована основная идея — применение при анализе данных «механизма внимания». Нейронная сеть «Трансформер» [33] была первоначально разработана для выполнения анализа и преобразования текстов, которые четко разделены на устойчивые смысловые единицы — слова, для которых однозначно определены позиции в анализируемой фразе. Предложенный в [33] механизм внимания позволял определять результаты анализа фразы смысловыми связями между составляющими ее словами, что можно считать вариантом иерархического структурного анализа текстовых данных. Более того, в декодирующей части трансформера результаты анализа более высокого уровня (сформированная уже часть преобразованной фразы) используются с применением синтезированного при обучении трансформера механизма внимания для выбора последующих частей фразы, что можно интерпретировать как применение принципа адаптивного резонанса.

Результаты, достигнутые трансформером при преобразовании текстов, убедительно демонстрируют преимущества применения иерархического структурного анализа и адаптивного резонанса. Однако попытка использования трансформеров для обработки изображений (Image transformer [34]), далека от применения полноценного структурного анализа. В качестве структурных элементов, которым в классическом трансформере [33] соответствовали слова, в трансформере изображений [34] использованы отдельные пикселы, упорядоченные в порядке растрового сканирования изображения, и в кодах их пространственных позиций использован этот же порядок. Кроме того, совокупность пикселов поделена на локальные блоки, внутри которых и применяется локально механизм внимания (в отличие от глобального его применения в трансформере [33]). В связи с локальным характером анализа изображения результаты его обработки также имели достаточно локальный характер: обрабатывались изображения очень малого размера (всего 32 × 32 пиксела) в задачах повышения их пространственного разрешения или синтеза их отсутствующих фрагментов. При локальном характере анализа изображения речь о классификации его содержания как единого целого идти не могла, а использование пикселов в качестве объектов, обрабатываемых механизмом внимания, делало очень локальной область его применения (вычислительные затраты, связанные с применением этого механизма, растут пропорционально квадрату числа объектов внимания).

Перечисленные проблемы трансформера изображений [34] были достаточно успешно решены Алексеем Досовицким и его коллегами в работе [35]. В отличие от [34] они использовали в качестве структурных элементов изображения (токенов) не отдельные пикселы, а локальные его фрагменты размерами 16 × 16 пикселов. С одной стороны, это наполнило используемые структурные элементы более осмысленным содержанием. С другой стороны, увеличение размеров структурных элементов позволило уменьшить их количество и при приемлемых вычислительных затратах сделать область применения механизма внимания не локальной, а глобальной, включив в нее всю площадь изображения. Таким образом, в [35] были более полноценно, чем в традиционных сверточных сетях, использованы преимущества применения структурного анализа, но при обработке изображений использовался только энкодер трансформера, а результаты кодирования классифицировались многослойным перцептроном, поэтому применение адаптивного резонанса в видео-трансформере [35] обнаружить не удается.

Подобно трансформеру изображений [34] в видеотрансформере [35] авторы остановились на «одномерном» кодировании информации о пространственном положении структурных элементов: фрагменты, используемые в качестве токенов, упорядочены в виде одномерной цепочки, полученной растровым сканированием изображения. Точность классификации изображений, достигаемая в [35], оказывается сопоставимой с точностью, достигаемой большой моделью сверточной сети ResNet-152. Предположительно, это может быть связано с тем, что и в трансформере [35], и в сверточных сетях локально анализируемые фрагменты (структурные элементы) изображений выбираются случайным образом в сканирующем изображение окне. При сравнении с результатами классификации сетью ResNet-152, предобученной на малой базе данных ImageNet (ImageNet1k, содержавшей 1,3 млн изображений), точность классификации видео-трансформером, предобученным на изображениях из значительно большей базы данных ImageNet21k (14 млн изображений), проигрывала (87,54 % против 85,3 %). Только при использовании предобучения на очень большой базе данных JFT-300М (33 млн изображений) видео-трансформеру, удалось получить точность классификации 87,6 % и «победить», наконец, сеть ResNet-152. Вероятно, на большей обучающей выборке применяемому в [35] механизму внимания удалось зафиксировать дополнительно и взаимосвязи таких «случайно выбираемых» структурных элементов при одномерном кодировании их пространственных положений.

GLOM

Проблема автоматического выбора структурных элементов, наконец, более рационально решается в вычислительной архитектуре GLOM (от английского agglomerate — собирать вместе), предложенной Д. Хинтоном в [36]. Хинтон предложил иерархически организованную искусственную нейронную сеть, на каждом уровне которой самоорганизуются «островки активности», соответствующие областям изображения, локально отличающимся по каким-либо признакам от окружения. Действительно, такой выбор структурных элементов, представляется более осмысленным, чем вырезание из изображения случайно выбранных фрагментов. Проиллюстрируем предложенную Хинтоном технологию формирования структурных элементов разного иерархического уровня (рис. 7).

В нейронных слоях более высокого иерархического уровня объединяются, согласно [36], пространственно соседние структурные элементы предыдущего уровня, обладающие сходством по каким-либо признакам (сходство показано на рис. 7 сходством цветов и направлений векторов выходных параметров нейронов). Таким об-



Рис. 7. Формирование островков активности, нейронов со сходными выходными параметрами [38]. На рисунке справа видно, как формируются однородные области, соответствующие граням параллелепипеда. *l* — порядковый номер одного из случайно выбранных иерархических уровней системы

Fig. 7. Formation of activity islands composed of neurons with similar output parameters [38]; it can be seen at the right how the homogeneous areas corresponding to faces of parallelepiped are formed.

'l' denotes a sequential number of one of randomly chosen hierarchical levels of system

разом в изображении постепенно из локальных частей собираются более крупные структуры. Одновременно с этим нисходящие (от более высоких иерархических уровней нейронной сети к более низким) информационные связи способствуют обособлению и увеличению однородности параметров нейронов локальных областей активности (более мелких структурных элементов) в слоях более низких уровней иерархии. Следовательно, в нейронной сети Хинтона реализуется принцип адаптивного резонанса.

Можно пояснить предложенный Хинтоном итеративный процесс формирования островков активности формулой:

$$\mathbf{O}_{l,j}(t) = \sqrt[3]{w_{l+1,j}\mathbf{O}_{l+1,j}(t)w_{l-1,j}\mathbf{O}_{l-1,j}(t)\sum_{i\in\Omega_j}\mathbf{O}_{l,i}(t-1)a_{l,i}(t)},$$

где $\mathbf{O}_{l,j}(t)$ — вектор выходных параметров *j*-го нейрона в слое *l* в момент времени *t*; $w_{l-1,j}$ и $w_{l+1,j}$ — весовые коэффициенты связей с одноименными (*j*-ми) нейронами предыдущего и последующего слоев; $a_{l,i}$ — весовые коэффициенты усреднения векторов выходных параметров нейронов в окрестности $\Omega_{j,j}$ -го нейрона *l*-го слоя. В разных частях работы [36] усреднение по окрестности Ω_{j} упоминается как «гауссова выборка» (gaussian sampling), либо как суммирование с весовыми коэффициентами «внимания» $a_{l,i}$, пропорциональными экспоненте скалярного произведения «взвешиваемых» векторов.

В рамках текущего исследования не удалось найти информацию о результатах реального воплощения архитектуры GLOM в виде программно реализованной модели. В [36] архитектура GLOM представлена в виде идеи. Практическая реализуемость всех представленных в GLOM этапов иерархической обработки изображений с применением известных сегодня технологий, основанных на искусственных нейронных сетях, не вызывает сомнений, поэтому предложенные в [36] идеи эффективной реализации в нейронной сети методов иерархического структурного анализа и адаптивного резонанса представляются интересными и перспективными. Единственное, что вызывает сомнение в [36], это формирование островков активности (пространственно обособленных структурных элементов) путем локального гауссова усреднения параметров выходов соседствующих в слое нейронов. Двумерная гауссова фильтрация выходных сигналов нейронов слоя может приводить к локальному размытию, а не к локализации однородных островков. Хинтон утверждает в [36], что для формирования островков активности с четкими границами, состоящих из нейронов, характеризующихся сходными выходными векторами, требуется также специальная организация взаимодействия нейронов между соседними слоями, которое он уподобляет регуляризации, но четкого обоснования эффективности такой регуляризации он не приводит. Однако известны эффективные способы формирования хорошо очерченных островков активности, например, с помощью кластеризации в самоорганизующихся картах Кохонена [37], что подтверждает реализуемость идей, представленных в [36]. В книге авторов настоящей работы [38] также приводятся соображения, как организовать в этом случае передачу информации о сформированных кластерах в нейронный слой следующего иерархического уровня, однако работоспособность этих идей еще требует подтверждения путем их воплощения в программно реализованных моделях.

Обсуждение

В приведенной таблице приведены итоги исследований, достигнутые в результате использования принципов структурного анализа и адаптивного резонанса при анализе изображений в искусственных нейронных сетях.

Полученные результаты (таблица) подтверждают спонтанность возврата исследователей к применению принципов структурного анализа и адаптивного резонанса: они «неожиданно» с разными названиями реализуются в новых нейронных архитектурах, при настройке которых используются «разношерстные» базы данных и разнородные оценки качества обучения, выполняемые в сравнении с разнообразными нейронными сетями-аналогами. Следует, однако, отметить, что в капсульных нейронных сетях, конструируемых в последние годы под руководством Джеффри Хинтона, систематически делаются попытки использовать преимущества применения этих принципов при обработке изображений. Эти преимущества оказались особенно заметными, когда решались задачи, связанные с анализом изображений, подвергающихся геометрическим преобразованиям, поскольку именно в капсульных сетях наиболее гибко используется информация о взаимном пространственном положении фрагментов изображения, что, как раз, и определяет преимущества применения структурного анализа по сравнению с методами, рассматривающими изображение как единое целое.

Отметим также, что, к сожалению, в нейронных сетях, реализованных уже к настоящему времени в виде программных моделей, полноценное выделение структурных элементов в виде объектов осмысленной формы не производилось, поэтому выполняемый в них иерархический структурный анализ с применением адаптивного резонанса нельзя считать полноценным. Ближе всего к решению этой проблемы подошел Джеффри Хинтон в своей нейронной вычислительной архитектуре GLOM [36], которая пока представлена только в виде общей концепции. Идеи, реализованные в GLOM, представляются перспективными, поэтому можно предположить, что дальнейшие успехи в области автоматического анализа изображений могут быть достигнуты при их реальном практическом воплощении в программных моделях новых нейронных сетей.

Выполнение дальнейших исследований в направлении использования иерархического структурного анализа и адаптивного резонанса представляется тем более перспективным, что применение этих принципов в «классическом» трансформере [33] позволило уже

Таблица. Сравнение результатов анализа изображений нейронными сетями, применяющими методы структурного анализа и адаптивного резонанса, с результатами сетей-аналогов

Table. Comparison of the image analysis results in the neural networks applying the methods of structural analysis and adaptive resonance with the results of networks-analogs

Новая нейронная сеть	Результаты новой сети	Сеть-аналог, с которой выпол- нено сравнение	Результаты сети-аналога
Нейронная сеть [5] с локаль- ными связями	Точность классификации 98,8 % на выборке изображе- ний, сформированной в [5]	Многослойный перцептрон с полносвязными слоями ней- ронов	Точность классификации 98,3 % на той же выборке изображений, сформирован- ной в [5]
Сеть [19] с раздельным ана- лизом частей изображения	Точность обнаружения на базе PASCAL VOC 2010 42 %	R-CNN [24]	Точность обнаружения на базе PASCAL VOC 2010 53,7 %
Капсульная сеть [26]	Точность классификации на базе AFFNIST 79 %	Сверточные сети новейшей архитектуры (state of the art)	Точность классификации на базе AFFNIST 66 %
Капсульная сеть [27], при- меняющая метод ожидания- максимизации	Ошибка классификации 1,4 % на базе SmallNORB	Капсульная сеть [26]	Ошибка классификации 2,7 % на базе SmallNORB
Капсульная сеть [27], при- меняющая метод ожидания- максимизации	Ошибка классификации 1,4 % на базе SmallNORB	Сверточные сети новейшей архитектуры (state of the art)	Ошибка классификации 5,2 % на базе SmallNORB
Иерархический капсульный автоэнкодер [29]	Точность классификации на базе MNIST 99 %	Нейронная сеть [31], выпол- няющая ассоциативную кла- стеризацию	Точность классификации на базе MNIST 98,7 %
Видео-трансформер [35]	Точность классификации 85,3 %, предобучение на большой базе ImageNet21k (14 млн изображений)	ResNet-152	Точность классификации 87,54 %, предобучение на ма- лой базе ImageNet1k (1,3 млн изображений)
Видео-трансформер [35]	Точность классификации 87,6 %, предобучение на очень большой базе JFT- 300M (33 млн изображений)	ResNet-152	Точность классификации 87,54 %, предобучение на ма- лой базе ImageNet1k (1,3 млн изображений)

достичь хороших результатов при обработке текстов. Рассмотренные результаты применения структурного анализа и адаптивного резонанса при сопоставлении изображений в контурном структурном классификаторе (рис. 3 и рис. 4) также подтверждают перспективность их использования. Заметим, что, к сожалению, этот структурный классификатор был реализован в практически значимой форме пока не в виде нейронных сетей, а другим способом [17] — на основе технологии оптимизированного обхода дерева решений.

Заключение

Иерархическая организация наблюдаемого мира требует применения иерархических алгоритмов его

Литература

- Grossberg S. Adaptive pattern classification and universal recoding: II. Feedback, expectation, olfaction, illusions // Biological Cybernetics. 1976. V. 23. N 4. P. 187–202. https://doi.org/10.1007/ BF00340335
- Carpenter G.A., Grossberg S. Category learning and adaptive pattern recognition: a neural network model // Proc. of the 3rd Army Conference on Applied Mathematics and Computing. 1986. P. 37–56.
- Carpenter G.A., Grossberg S. ART 2: Self-organization of stable category recognition codes for analog input patterns // Applied Optics. 1987. V. 26. N 12. P. 4919–4930. https://doi.org/10.1364/ AO.26.004919
- Rumelhart D.E. Learning Internal Representations by Error Propagation. Institute for Cognitive Science, 1985. 34 p.
- Yamada K., Kami H., Tsukumo J., Temma T. Handwritten numeral recognition by multilayered neural network with improved learning algorithm // Proc. of the International 1989 Joint Conference on Neural Networks. 1989. P. 259–266. https://doi.org/10.1109/ IJCNN.1989.118708
- Fukushima K. Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position // Biological Cybernetics. 1980. V. 36. N 4. P. 193–202. https://doi.org/10.1007/bf00344251
- Krizhevsky A., Sutskever I., Hinton G.E. ImageNet classification with deep convolutional neural networks // Proc. of the 26th International Conference on Neural Information Processing Systems (NIPS'12). 2012. V. 1. P. 1097–1105.
- Fukushima K. A neural network model for selective attention in visual pattern recognition // Biological Cybernetics. 1986. V. 55. N 1. P. 5–15. https://doi.org/10.1007/BF00363973
- Hopfield J.J., Tank D.W. Neural computation of decisions in optimization problems // Biological Cybernetics. 1985. V. 52. N 3. P. 141–152. https://doi.org/10.1007/bf00339943
- Lutsiv V.R., Novikova T.A. On the use of neurocomputer for stereoimage processing // Pattern Recognition and Image Analysis // Advances in Mathematical Theory and Applications. 1992. V. 2. N 4. P. 441–444.
- Данилов Е.П., Луцив В.Р., Новикова Т.А., Малышев И.А. Разработка методов и алгоритмов реализации интеллектуальных вычислительных устройств с использованием нейронных сетей в интересах задач анализа и сопоставления изображений. Отчет о работе 22201-007-94, Шифр «Нейросеть». СПб.: Государственный оптический институт им. С.И. Вавилова, С.-Петербург, 1994. С. 39–50.
- Dolinov D.S., Zherebko A.K., Lutsiv V.R., Novikova T.A. Using artificial neural networks in image processing problems // Journal of Optical Technology. 1997. V. 64. N 2. P. 112–118.
- Lutsiv V.R., Malyshev I.A., Pepelka V.A. Automatic fusion of multiple-sensor and multiple-season images // Proceedings of SPIE. 2001. V. 4380. P. 174–183. https://doi.org/10.1117/12.436990
- 14. Lutsiv V.R., Malyshev I.A., Pepelka V.A., Potapov A.S. The Target independent algorithms for description and structural matching of

анализа, одной из основ которого является метод адаптивного резонанса. В представленном исследовании продемонстрировано что, создатели искусственных нейронных сетей прибегали к применению принципов структурного анализа и адаптивного резонанса эпизодически, на десятилетия забывая о них и вновь возвращаясь к их реализации, что целенаправленного анализа преимуществ и методов применения этих принципов в нейронных сетях не проводилось. В настоящей работе сделана попытка заполнить этот пробел и показать наиболее перспективные направления применения рассмотренных принципов при разработке алгоритмов обработки изображений в искусственных нейронных сетях на примерах капсульных нейронных сетей и вычислительной архитектуры GLOM.

References

- Grossberg S. Adaptive pattern classification and universal recoding: II. Feedback, expectation, olfaction, illusions. *Biological Cybernetics*, 1976, vol. 23, no. 4, pp. 187–202. https://doi.org/10.1007/ BF00340335
- Carpenter G.A., Grossberg S. Category learning and adaptive pattern recognition: a neural network mode. *Proc. of the 3rd Army Conference* on Applied Mathematics and Computing, 1986, pp. 37–56.
- Carpenter G.A., Grossberg S. ART 2: Self-organization of stable category recognition codes for analog input patterns. *Applied Optics*, 1987, vol. 26, no. 12, pp. 4919–4930. https://doi.org/10.1364/ AO.26.004919
- 4. Rumelhart D.E. *Learning Internal Representations by Error Propagation.* Institute for Cognitive Science, 1985, 34 p.
- Yamada K., Kami H., Tsukumo J., Temma T. Handwritten numeral recognition by multilayered neural network with improved learning algorithm. *Proc. of the International 1989 Joint Conference on Neural Networks*, 1989, pp. 259–266. https://doi.org/10.1109/ IJCNN.1989.118708
- Fukushima K. Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biological Cybernetics*, 1980, vol. 36, no. 4, pp. 193–202. https://doi. org/10.1007/bf00344251
- Krizhevsky A., Sutskever I., Hinton G.E. ImageNet classification with deep convolutional neural networks. *Proc. of the 26th International Conference on Neural Information Processing Systems (NIPS'12)*, 2012, vol. 1, pp. 1097–1105.
- Fukushima K. A neural network model for selective attention in visual pattern recognition. *Biological Cybernetics*, 1986, vol. 55, no. 1, pp. 5–15. https://doi.org/10.1007/BF00363973
- Hopfield J.J., Tank D.W. Neural computation of decisions in optimization problems. *Biological Cybernetics*, 1985, vol. 52, no. 3, pp. 141–152. https://doi.org/10.1007/bf00339943
- Lutsiv V.R., Novikova T.A. On the use of neurocomputer for stereoimage processing. Pattern Recognition and Image Analysis. *Advances in Mathematical Theory and Applications*, 1992, vol. 2, no. 4, pp. 441–444.
- Danilov E.P., Lutciv V.R., Novikova T.A., Malyshev I.A. Development of methods and algorithms for the implementation of intelligent computing devices using neural networks for the purposes of image analysis and comparison problems. Statement of work no. 22201-007-94, Code "Neural network". St. Petersburg, Open Joint Stock Company «S. I. Vavilov State Optical Institute», 1994, pp. 39– 50. (in Russian)
- Dolinov D.S., Zherebko A.K., Lutsiv V.R., Novikova T.A. Using artificial neural networks in image processing problems. *Journal of Optical Technology*, 1997, vol. 64, no. 2, pp. 112–118.
- Lutsiv V.R., Malyshev I.A., Pepelka V.A. Automatic fusion of multiple-sensor and multiple-season images. *Proceedings of SPIE*, 2001, vol. 4380, pp. 174–183. https://doi.org/10.1117/12.436990
- Lutsiv V.R., Malyshev I.A., Pepelka V.A., Potapov A.S. The Target independent algorithms for description and structural matching of

aerospace photographs // Proceedings of SPIE. 2002. V. 4741. P. 351-362. https://doi.org/10.1117/12.478732

- Луцив В. Автоматический анализ изображений: Объектно-независимый структурный подход. Саарбрюккен: Ламберт Академик Паблишинг, 2011. 308 с.
- Слэйгл Д.Р. Искусственный интеллект: подход на основе эвристического программирования. М.: Мир, 1973. 319 с.
- Lutsiv V.R., Malyshev I.A., Potapov A.S. Hierarchical structural matching algorithms for registration of aerospace images // Proceedings of SPIE. 2003. V. 5238. P. 164–175. https://doi. org/10.1117/12.511770
- Ponomarev S., Lutsiv V., Malyshev I. Automatic structural matching of 3D image data // Proceedings of SPIE. 2015. V. 9649. P. 96490M. https://doi.org/10.1117/12.2194312
- Girshick R., Iandola F., Darrell T., Malik J. Deformable part models are convolutional neural networks // Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2015. P. 437–446. https://doi.org/10.1109/cvpr.2015.7298641
- Felzenszwalb P.F., Girshick R.B., McAllester D., Ramanan D. Object detection with discriminatively trained part-based models // IEEE Transactions on Pattern Analysis and Machine Intelligence. 2010. V. 32. N 9. P. 1627–1645. https://doi.org/10.1109/tpami.2009.167
- Yang Y., Ramanan D. Articulated human detection with flexible mixtures of parts // IEEE Transactions on Pattern Analysis and Machine Intelligence. 2013. V. 35. N 12. P. 2878–2890. https://doi. org/10.1109/tpami.2012.261
- Wang Y., Tran D., Liao Z. Learning hierarchical poselets for human parsing // CVPR 2011. P. 1705–1712. https://doi.org/10.1109/ cvpr.2011.5995519
- Wan L., Eigen D., Fergus R. End-to-end integration of a convolutional network, deformable parts model and non-maximum suppression // Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2015. P. 851–859. https://doi.org/10.1109/ cvpr.2015.7298686
- Girshick R., Donahue J., Darrell T., Malik J. Rich feature hierarchies for accurate object detection and semantic segmentation // Proc. of the IEEE Conference on Computer Vision and Pattern Recognition. 2014. P. 580–587. https://doi.org/10.1109/CVPR.2014.81
- Hinton G.E., Krizhevsky A., Wang S.D. Transforming autoencoders // Lecture Notes in Computer Science. 2011. V. 6791. P. 44– 51. https://doi.org/10.1007/978-3-642-21735-7_6
- Sabour S., Frosst N., Hinton G.E. Dynamic routing between capsules // arXiv. 2017. arXiv:1710.09829v2. https://doi. org/10.48550/arXiv.1710.09829
- Hinton G., Sabour S., Frosst N. Matrix capsules with EM routing // Proc. of the 6th International Conference on Learning Representations (ICLR 2018). 2018. P. 1–15.
- LeCun Y., Huang F.J., Bottou L. Learning methods for generic object recognition with invariance to pose and lighting // Proc. of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2004). 2004. P. 97–104. https://doi.org/10.1109/ cvpr.2004.1315150
- Kosiorek A.R., Sabour S., Teh Y.W., Hinton G.E. Stacked capsule autoencoders // arXiv. 2019. arXiv:1906.06818v2. https://doi. org/10.48550/arXiv.1906.06818
- Lee J., Lee Y., Kim J., Kosiorek A.R., Choi S., Teh Y.W. Set Transformer: a framework for attention-based permutation-invariant neural networks // arXiv. 2019. arXiv:1810.00825v3. https://doi. org/10.48550/arXiv.1810.00825
- Haeusser P., Plapp J., Golkov V., Aljalbout E., Cremers D. Associative deep clustering: training a classification network with no labels // Lecture Notes in Computer Science. 2019. V. 11269. P. 18–32. https:// doi.org/10.1007/978-3-030-12939-2_2
- Sun W, Tagliasacchi A., Deng B., Sabour S., Yazdani S., Hinton G., Yi K.M. Canonical capsules: self-supervised capsules in canonical pose // arXiv. 2021. arXiv:2012.04718v2. https://doi.org/10.48550/ arXiv.2012.04718
- Vaswani A., Shazeer N., Parmar N., Uszkoreit J., Jones L., Gomez A.N., Kaiser L., Polosukhin I. Attention is all you need // arXiv. 2017. arXiv:1706.03762v5. https://doi.org/10.48550/ arXiv.1706.03762
- Parmar N., Vaswani A., Uszkoreit J., Kaiser L., Shazeer N., Ku A., Tran D. Image Transformer // arXiv. 2018. arXiv:1802.05751v3. https://doi.org/10.48550/arXiv.1802.05751
- Dosovitskiy A., Beyer L., Kolesnikov A., Weissenborn D., Zhai X., Unterthiner T., Dehghani M., Minderer M., Heigold G., Gelly S.,

aerospace photographs. *Proceedings of SPIE*, 2002, vol. 4741, pp. 351–362. https://doi.org/10.1117/12.478732

- 15. Lutsiv V. Automatic image analysis: Object-independent structured approach. Lambert Academic Publishing, 2011, 308 p. (in Russian)
- 16. Slagle J.R. Artificial Intelligence: The Heuristic Programming Approach. McGraw Hill, 1973, 196 p.
- Lutsiv V.R., Malyshev I.A., Potapov A.S. Hierarchical structural matching algorithms for registration of aerospace images. *Proceedings of SPIE*, 2003, vol. 5238, pp. 164–175. https://doi. org/10.1117/12.511770
- Ponomarev S., Lutsiv V., Malyshev I. Automatic structural matching of 3D image data. *Proceedings of SPIE*, 2015, vol. 9649, pp. 96490M. https://doi.org/10.1117/12.2194312
- Girshick R., Iandola F., Darrell T., Malik, J. Deformable part models are convolutional neural networks. *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015, pp. 437– 446. https://doi.org/10.1109/cvpr.2015.7298641
- Felzenszwalb P.F., Girshick R.B., McAllester D., Ramanan D. Object detection with discriminatively trained part-based models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2010, vol. 32, no. 9, pp. 1627–1645. https://doi.org/10.1109/tpami.2009.167
- Yang Y., Ramanan D. Articulated human detection with flexible mixtures of parts. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2013, vol. 35, no. 12, pp. 2878–2890. https:// doi.org/10.1109/tpami.2012.261
- Wang Y., Tran D., Liao Z. Learning hierarchical poselets for human parsing. CVPR 2011, pp. 1705–1712. https://doi.org/10.1109/ cvpr.2011.5995519
- Wan L., Eigen D., Fergus R. End-to-end integration of a convolutional network, deformable parts model and non-maximum suppression. *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015, pp. 851–859. https://doi.org/10.1109/ cvpr.2015.7298686
- Girshick R., Donahue J., Darrell T., Malik J. Rich feature hierarchies for accurate object detection and semantic segmentation. *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 580–587. https://doi.org/10.1109/CVPR.2014.81
- Hinton G.E., Krizhevsky A., Wang S.D. Transforming auto-encoders. Lecture Notes in Computer Science, 2011, vol. 6791, pp. 44–51. https://doi.org/10.1007/978-3-642-21735-7_6
- Sabour S., Frosst N., Hinton G.E. Dynamic routing between capsules. arXiv, 2017, arXiv:1710.09829v2. https://doi.org/10.48550/ arXiv.1710.09829
- Hinton G., Sabour S., Frosst N. Matrix capsules with EM routing. Proc. of the 6th International Conference on Learning Representations (ICLR 2018), 2018, pp. 1–15.
- LeCun Y., Huang F.J., Bottou L. Learning methods for generic object recognition with invariance to pose and lighting. *Proc. of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2004)*, 2004, pp. 97–104. https://doi.org/10.1109/ cvpr.2004.1315150
- Kosiorek A.R., Sabour S., Teh Y.W., Hinton G.E. Stacked capsule autoencoders. arXiv, 2019, arXiv:1906.06818v2. https://doi. org/10.48550/arXiv.1906.06818
- Lee J., Lee Y., Kim J., Kosiorek A.R., Choi S., Teh Y.W. Set Transformer: a framework for attention-based permutation-invariant neural networks. *arXiv*, 2019, arXiv:1810.00825v3. https://doi. org/10.48550/arXiv.1810.00825
- Haeusser P., Plapp J., Golkov V., Aljalbout E., Cremers D. Associative deep clustering: training a classification network with no labels. *Lecture Notes in Computer Science*, 2019, vol. 11269, pp. 18–32. https://doi.org/10.1007/978-3-030-12939-2_2
- Sun W., Tagliasacchi A., Deng B., Sabour S., Yazdani S., Hinton G., Yi K.M. Canonical capsules: self-supervised capsules in canonical pose. arXiv, 2021, arXiv:2012.04718v2. https://doi.org/10.48550/ arXiv.2012.04718
- Vaswani A., Shazeer N., Parmar N., Uszkoreit J., Jones L., Gomez A.N., Kaiser L., Polosukhin I. Attention is all you need. arXiv, 2017, arXiv:1706.03762v5. https://doi.org/10.48550/ arXiv.1706.03762
- Parmar N., Vaswani A., Uszkoreit J., Kaiser L., Shazeer N., Ku A., Tran D. Image Transformer. *arXiv*, 2018, arXiv:1802.05751v3. https://doi.org/10.48550/arXiv.1802.05751
- Dosovitskiy A., Beyer L., Kolesnikov A., Weissenborn D., Zhai X., Unterthiner T., Dehghani M., Minderer M., Heigold G., Gelly S., Uszkoreit J., Houlsby N. An image is worth 16x16 words:

Uszkoreit J., Houlsby N. An image is worth 16x16 words: transformers for image recognition at scale // arXiv. 2021. https://doi. org/10.48550/arXiv.2010.11929

- Hinton G. How to represent part-whole hierarchies in a neural network // Neural Computation. 2023. V. 35. N 3. P. 413–452. https:// doi.org/10.1162/neco a 01557
- Kohonen T. Self-organization and Associative Memories. Springer, 1984. 255 p.
- 38. Луцив В.Р., Михалькова М.А., Ячная В.О. Компьютерное зрение: учеб. пособие: в 3 ч. Ч. 3. Современные модификации архитектур нейронных сетей и методы улучшения изображений. СПб.: ГУАП, 2024. 196 с.

Авторы

Луцив Вадим Ростиславович — доктор технических наук, профессор, Санкт-Петербургский государственный университет аэрокосмического приборостроения, Санкт-Петербург, 190000, Российская Федерация, вс 6602625465, https://orcid.org/0000-0003-1160-3758, vluciv@mail.ru

Михалькова Мария Анатольевна — младший научный сотрудник, Институт физиологии имени И.П. Павлова РАН, Санкт-Петербург, 199034, Российская Федерация, sc 57218284288, https://orcid.org/0000-0002-7186-5744, mikhalkova_maria@infran.ru

Ячная Валерия Олеговна — младший научный сотрудник, Институт физиологии имени И.П. Павлова РАН, Санкт-Петербург, 199034, Российская Федерация; аспирант, Санкт-Петербургский государственный университет аэрокосмического приборостроения, Санкт-Петербург, 190000, Российская Федерация, вс 57209076316, https://orcid.org/0000-0002-7239-2096, yachnaya_valeria@infran.ru

Статья поступила в редакцию 02.11.2024 Одобрена после рецензирования 27.12.2024 Принята к печати 30.03.2025 transformers for image recognition at scale. *arXiv*, 2021, https://doi. org/10.48550/arXiv.2010.11929

- Hinton G. How to represent part-whole hierarchies in a neural network. *Neural Computation*, 2023, vol. 35, no. 3, pp. 413–452. https://doi.org/10.1162/neco_a_01557
- Kohonen T. Self-Organization and Associative Memories. Springer, 1984, 255 p.
- Lutciv V.R., Mikhalkova M.A., Iachnaia V.O. Computer Vision. A Tutorial in 3 parts. Part 3: Modern Modifications of Neural Network Architectures and Methods of Image Enhancement. St. Petersburg, GUAP, 2024, 196 p. (in Russian)

Authors

Vadim R. Lutsiv — D.Sc., Professor, Saint Petersburg State University of Aerospace Instrumentation, Saint Petersburg, 190000, Russian Federation, Sc 6602625465, https://orcid.org/0000-0003-1160-3758, vluciv@mail.ru

Maria A. Mikhalkova — Junior Researcher, Pavlov Institute of Physiology of the Russian Academy of Sciences, Saint Petersburg, 199034, Russian Federation, SC 57218284288, https://orcid.org/0000-0002-7186-5744, mikhalkova_maria@infran.ru

Valeria O. Yachnaya — Junior Researcher, Pavlov Institute of Physiology of the Russian Academy of Sciences, Saint Petersburg, 199034, Russian Federation; PhD Student, Saint Petersburg State University of Aerospace Instrumentation, Saint Petersburg, 190000, Russian Federation, Sc 57209076316, https://orcid.org/0000-0002-7239-2096, yachnaya_valeria@infran.ru

Received 02.11.2024 Approved after reviewing 27.12.2024 Accepted 30.03.2025



Работа доступна по лицензии Creative Commons «Attribution-NonCommercial» **I/İTMO**

НАУЧНО-ТЕХНИЧЕСКИЙ ВЕСТНИК ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ, МЕХАНИКИ И ОПТИКИ март–апрель 2025 Том 25 № 2 http://ntv.ifmo.ru/ SCIENTIFIC AND TECHNICAL JOURNAL OF INFORMATION TECHNOLOGIES, MECHANICS AND OPTICS March–April 2025 Vol. 25 No 2 http://ntv.ifmo.ru/en/ ISSN 2226-1494 (print) ISSN 2500-0373 (online)

ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ, МЕХАНИКИ И ОПТИКИ

университет итмо

doi: 10.17586/2226-1494-2025-25-2-286-294 УДК 004.056.4

Построение множества плотных информационных совокупностей для кодов Гилберта и их расширений Мария Николаевна Исаева⊠

Национальный исследовательский университет «Высшая школа экономики», Санкт-Петербург, 190008, Российская Федерация

misaeva@hse.ru^{\overlinewy}, https://orcid.org/0009-0007-6228-0617

Аннотация

Введение. При передаче информации по каналам с группирующимися ошибками традиционным подходом является декорреляция канала и использование кодов, исправляющих независимые ошибки. Процедура декорреляции понижает достижимые скорости надежной передачи, поэтому актуальной является задача использования специальных кодов для каналов с памятью и построения эффективных вычислительных методов декодирования для исправления группирующихся ошибок. Для класса случайных кодов известен подход с использованием информационных совокупностей ограниченного диаметра при исправлении пакетов ошибок. Размер множества информационных совокупностей растет линейно с увеличением длины кода, а построение множества описывается вероятностной процедурой. В работе рассматривается построение множества информационных совокупностей для специального класса кодов, исправляющих пакеты ошибок, называемого кодами Гилберта. Метод. Рассмотрены множества кодовых позиций наименьшего возможного диаметра. На основе вычисления рангов подматриц проверочной матрицы кода Гилберта оценена вероятность того, что набор позиций является информационной совокупностью. Для заданного расположения информационной совокупности проанализированы позиции исправляемых пакетов. На основании проведенного анализа предложена методика построения множества плотных информационных совокупностей для кодов Гилберта с целью исправления всех пакетов ошибок в пределах корректирующей способности кода. Используя особенности задания параметров кодов Гилберта, проведена оценка размера полученного множества плотных информационных совокупностей. Основные результаты. При простом размере блока проверочной матрицы квазициклического кода показано, что для кодов Гилберта на любой позиции находится плотная информационная совокупность. Установлено, что в случае расширенных кодов Гилберта совокупности минимального диаметра существуют только на последней позиции каждого блока. Предложена процедура построения множества плотных информационных совокупностей минимального диаметра для кодов Гилберта и их расширений. Проведено сравнение размера множества информационных совокупностей и вероятность его получения для кодов Гилберта и случайных кодов. Установлено, что количество информационных совокупностей, полученных по предложенной процедуре, не растет с увеличением длины кода. Обсуждение. Полученные в работе результаты, показывают возможность разработки вычислительно эффективных декодеров на основе информационных совокупностей при исправлении однократных пакетов ошибок. В отличие от случайных линейных кодов, для которых методы построения информационных совокупностей, в том числе плотных, носят вероятностный характер, для кодов Гилберта определена процедура гарантированного построения множества информационных совокупностей минимального диаметра. Квазициклическая структура кодов Гилберта позволяет строить множества плотных информационных совокупностей меньшей размерности, чем для случайных кодов. Полученные результаты позволяют гарантировать исправление пакетов ошибок в пределах корректирующей способности кодов Гилберта и их расширений с малой вычислительной сложностью. Использование вычислительно эффективных процедур кодирования и декодирования пакетов ошибок позволит повысить надежность доставки сообщений в каналах с памятью.

Ключевые слова

информационные совокупности, корректирующая способность, коды Гилберта, каналы с памятью, пакеты ошибок

© Исаева М.Н., 2025
Благодарности

Работа подготовлена в результате проведения исследования в рамках Программы фундаментальных исследований Национального исследовательского университета «Высшая школа экономики» (НИУ ВШЭ), лаборатория Интернета вещей и киберфизических систем НИУ ВШЭ в Санкт-Петербурге.

Ссылка для цитирования: Исаева М.Н. Построение множества плотных информационных совокупностей для кодов Гилберта и их расширений // Научно-технический вестник информационных технологий, механики и оптики. 2025. Т. 25, № 2. С. 286–294. doi: 10.17586/2226-1494-2025-25-2-286-294

Development of dense information sets for Gilbert codes and its extensions

Maria N. Isaeva⊠

HSE University-St. Petersburg, Saint Petersburg, 190008, Russian Federation

misaeva@hse.ru^{\exist}, https://orcid.org/0009-0007-6228-0617

Abstract

When transmitting information over channels with grouping errors, the traditional approach is channel decorrelation and use of codes correcting independent errors. The decorrelation procedure lowers achievable rates of reliable transmission, therefore the problem of using special codes for channels with memory and construction of computationally effective decoding methods for correction of grouping errors is actual. For the class of random codes, an approach is known using information sets of limited diameter to correct error bursts. The size of the set of information sets grows linearly with increasing code length, and the construction of the set is described by a probabilistic procedure. This article considers the construction of a set of information sets for a special class of codes that correct error bursts called Gilbert codes. The sets of code positions of the smallest possible diameter are considered. Based on the calculation of the ranks of the submatrices of the parity-check matrix of the Gilbert code, the probability that the set of positions is an information set is estimated. For a given location of the information set, the positions of the corrected bursts are analyzed. Based on the analysis, a method for constructing a set of dense information sets for Gilbert codes for correcting all error bursts within the code correcting capacity is proposed. Using the features of setting the parameters of Gilbert codes, an estimate of the size of the resulting set of dense information sets is carried out. For a simple block size of the paritycheck matrix of a quasi-cyclic code, it is shown that for Gilbert codes a dense information set is located at any position. In the case of extended Gilbert codes, it is shown that sets of minimum diameter exist only at the last position of each block. A procedure for constructing a set of dense information sets of minimum diameter for Gilbert codes and their extensions is proposed. A comparison is made of the set size of information sets and the probability of obtaining it for Gilbert codes and random codes. It is shown that the number of information sets obtained by the proposed procedure does not increase with the length of the code. The results obtained in the paper demonstrate the possibility of developing computationally efficient decoders based on information sets when correcting single error bursts. Unlike random linear codes, for which the methods of constructing information sets including dense ones, are probabilistic, a procedure for guaranteed construction of a set of information sets of minimal diameter is specified for Gilbert codes. The quasi-cyclic structure of Gilbert codes allows constructing sets of dense information sets of smaller dimension than for random codes. The obtained results allow us to guarantee the correction of error bursts within the correcting capacity of Gilbert codes and their extensions with low computational complexity. The use of computationally efficient procedures for encoding and decoding error bursts will improve the reliability of message delivery in channels with memory.

Keywords

information sets, error correcting capability, Gilbert codes, channels with memory, error bursts

Acknowledgements

The article was prepared within the framework of the Basic Research Program at HSE University, Internet of Things and Cyber-Physical Systems Laboratory, St. Petersburg School of Physics, Mathematics, and Computer Science.

For citation: Isaeva M.N. Development of dense information sets for Gilbert codes and its extensions. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 2025, vol. 25, no. 2, pp. 286–294 (in Russian). doi: 10.17586/2226-1494-2025-25-2-286-294

Введение

В задачах управления потоками данных, хранения больших объемов информации и ее передачи по каналам связи необходимо обеспечивать надежность обрабатываемых пользовательских данных. Как при передаче информации, так и в процессе ее обработки могут возникать ошибки. Шум влияет как на достоверность передаваемой информации, так и на вероятность доставки. Исходя из этого, актуальной проблемой является разработка методов обеспечения структурной надежности инфокоммуникационных систем и надежной доставки сообщений [1–4], включающих исправление возникших ошибок на принимающей стороне. Надежность доставки может повышаться путем внесения избыточности в систему связи и репликации передаваемых данных с кратностью, зависящей от требований надежности и своевременности доставки передаваемых сообщений [5, 6]. Одним из способов управления избыточностью является добавление контрольной информации непосредственно в само сообщение. Вопросами восстановления поврежденных данных с использованием этой информации занимается теория помехоустойчивого кодирования [7, 8].

Возникновение ошибок далеко не всегда адекватно описывается моделью случайных независимых событий, чаще всего ошибки сгруппированы внутри передаваемой последовательности [8, 9]. Для борьбы с этим явлением в каналах с группирующимися ошибками (пакетами ошибок) необходимо построение специфических кодеров и декодеров, которые учитывают особенности таких каналов.

Классические коды Рида–Соломона в каналах, описываемых марковскими цепями, рассматривались в [9], однако сложность кодирования и декодирования этих кодов сравнительно велика для современных систем связи. Более эффективные как по вероятности ошибки, так и по вычислительной сложности коды с малой плотностью проверок на четность применительно к каналам с памятью изучались в [10–12]. В работах [13–15] рассмотрены полярные коды. Следует отметить, что основное внимание исследователей уделяется вопросам построения кодов для каналов с памятью, а не специальным методам их декодирования. В [16] рассмотрен оконный декодер для исправления пакетов ошибок блочно-перестановочными кодами.

В качестве подхода к декодированию можно использовать декодирование по информационным совокупностям, которые, как показано в [17, 18], могут исправлять однократные пакеты ошибок в соответствии с корректирующей способностью случайных кодов. Но случайные коды не обладают какой-либо структурой, поэтому программная и аппаратная реализации процедур кодирования и декодирования может оказаться недостаточно вычислительно эффективной. В [18] представлено декодирование квазициклических кодов с помощью информационных совокупностей, однако исследование ограничено ансамблем случайных кодов, не оптимизированных под исправление пакетов ошибок. В [19] приведен частный случай квазициклических блочно-перестановочных кодов Гилберта и их модификации и доказано, что эти коды способны исправлять пакеты ошибок. Однако нерешенной остается проблема декодирования кодов Гилберта, реализующего их пакетную корректирующую способность.

При использовании информационных совокупностей для исправления пакетов ошибок количество ошибок может быть уменьшено введением ограничений на их диаметр. Для случайных линейных кодов построение информационных совокупностей заданного диаметра является вероятностным. В настоящей работе решается задача построения множества информационных совокупностей для исправления однократных пакетов ошибок с использованием особенностей конструкции кодов Гилберта и их расширений [19]. Необходимо предложить процедуру построения множества плотных информационных совокупностей для кодов Гилберта, оценить размер полученного множества в зависимости от параметров кода и вероятность его построения.

Блочно-циркулянтные конструкции низкоплотностных кодов

Коды с малой плотностью проверок на четность, или низкоплотностные коды, являются одним из наиболее распространенных классов кодов, используемых в современных инфокоммуникационных стандартах [7]. Эти коды характеризуются разреженной проверочной матрицей, что позволяет строить для них крайне вычислительно эффективные алгоритмы кодирования и декодирования, обеспечивающие низкие вероятности ошибки при исправлении независимых ошибок. Одной из наиболее важных конструкций низкоплотностных кодов является блочно-циркулянтная, в этом случае код является квазициклическим. Рассмотрим особенности данной конструкции.

Блочно-циркулянтные коды задаются с помощью базовой матрицы, состоящей из γ строк и ρ столбцов, причем $\gamma < \rho$. Каждый элемент базовой матрицы задает блок, представляющий собой двоичную квадратную матрицу размера $m \times m$. Базовая матрица \mathbf{H}_{base} блочно-циркулянтного кода имеет вид:

$$\mathbf{H}_{\text{base}} = \begin{bmatrix} t_{11} & \cdots & t_{1\rho} \\ \vdots & \ddots & \vdots \\ t_{\gamma 1} & \cdots & t_{\gamma \rho} \end{bmatrix},$$

где t_{ij} — целые числа. Если каждый элемент \mathbf{H}_{base} заменить на блок $\mathbf{C}^{t_{ij}}$, где \mathbf{C} — матрица циклической перестановки, размер которой $m \times m$, получим двоичную проверочную матрицу \mathbf{H} кода:

$$\mathbf{H} = \begin{bmatrix} \mathbf{C}^{t_{11}} \cdots & \mathbf{C}^{t_{1\rho}} \\ \vdots & \ddots & \vdots \\ \mathbf{C}^{t_{\gamma 1}} \cdots & \mathbf{C}^{t_{\gamma \rho}} \end{bmatrix}.$$

В общем случае в блочно-циркулянтной конструкции в качестве блока может быть выбран любой циркулянт, а в разновидности, называемой блочно-перестановочной конструкцией — произвольная матрица перестановки. Рассмотрим случай использования степеней матрицы С в качестве блоков. Таким образом, для данного случая термины блочно-перестановочный и блочно-циркулянтный взаимозаменяемы и в дальнейшем под этим термином будем подразумевать только матрицы описанного вида.

Известно, что максимальная длина пакета b_{\max} , исправляемого блочно-перестановочными кодами, ограничена размером блока *m*, или *b*_{max} < *m*, поэтому для более эффективного использования избыточности, вносимой помехоустойчивым кодом, необходимо максимизировать размер блока *т* при фиксированной длине *n*, т. е. минимизировать значение р. Это автоматически приводит и к минимизации значения у. Потому, в отличие от ситуации с независимыми ошибками, где для получения наилучших кодов рассматриваются матрицы из большого числа блоков (например, в стандарте 5G), для исправления пакетов необходимо использовать блочно-перестановочные коды, в которых количество блоков мало. В результате рассмотрим коды при $\gamma = 2$ или ү = 3, или, другими словами, с проверочной матрицей из двух или трех полос.

Специальным подклассом блочно-циркулянтных кодов с двумя полосами являются коды Гилберта [19, 20]. Особенность данных кодов — фиксированный способ выбора блоков: в двухполосной матрице кода Гилберта первая полоса состоит из единичных матриц, а вторая — из матриц циклической перестановки со степенями от 0 до $\rho - 1$, где $\rho \le m$. Отметим, что для построения кодов Гилберта используется матрица **С**, обеспечивающая циклический сдвиг влево, например, при m = 4:

$$\mathbf{C} = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

В современных стандартах при описании базовой матрицы кода обычно предполагается, что матрица С обеспечивает сдвиг вправо, однако это дало бы код с совершенно иными (худшими) свойствами с точки зрения исправления пакетов ошибок. Базовая матрица кода Гилберта задается следующим образом:

$$\mathbf{H}_{\text{base}} = \begin{bmatrix} 0 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & \rho - 2 & \rho - 1 \end{bmatrix},$$
(1)

где $\rho \ge 3$. Модификация кода Гилберта для $\rho \ge 4$ может быть получена добавлением третьей строки к базовой матрице [16, 17]:

$$\mathbf{H}_{\text{base}} = \begin{bmatrix} 0 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & \rho - 2 & \rho - 1 \\ \rho - 1 & \rho - 2 & \cdots & 1 & 0 \end{bmatrix}.$$
(2)

Число избыточных символов для кода Гилберта составляет r = 2m - 1, а для расширенного кода: r = 3m - 2.

Способность кода Гилберта или родственных ему кодов исправлять одиночные пакеты ошибок рассматривались во многих научных работах. В частности, в [19] была предложена процедура вычисления точной корректирующей способности таких кодов при исправлении однократных пакетов ошибок. Кратко, можно сказать, что при $\rho = 3$ код Гилберта исправляет пакеты длиной m - 1 и лежит на границе Рейгера (в соответствии с которой для линейного кода $2b_{max} \le r$). Отметим, что при увеличении ρ до примерно m/2 пакетная корректирующая способность b_{max} уменьшается практически по линейному закону также до величины около m/2, при дальнейшем увеличении b_{max} остается неизменной.

В случае расширенных кодов Гилберта известно, что если m — простое число, то они исправляют все пакеты до $b_{\text{max}} = m - 1$ для любого $\rho \in [4, m]$.

В работах [18, 21] рассматривался вопрос декодирования по информационным совокупностям и предлагалась методика построения множества из небольшого количества Т информационных совокупностей, которые гарантировали бы исправление пакетов в пределах корректирующей способности кода. Заметим, что возможность построения такого множества является случайным событием и определяется случайным выбором кода. Проводилась оценка вероятности этого события в предположении, что выбирается случайный линейный двоичный код. Если используется код с выраженной структурой, то полученные в [18, 21] вероятностные оценки могут довольно сильно исказиться, поэтому в дальнейшем рассматривается вопрос разработки методики построения множества плотных информационных совокупностей для блочно-перестановочных конструкций из двух и трех полос, а также оценка вероятности построения такого множества.

Для дальнейшего рассмотрения введем основные понятия информационных совокупностей [22, 23]. Информационной совокупностью называется множество позиций $\chi = \{1 \le j_1 < j_2 < ... < j_k \le n\}$, если элементы из кодового слова длиной *n*, соответствующие этим позициям, однозначно определяют кодовое слово. Если на позициях информационной совокупности не оказалось ошибок, то при декодировании по ней можно корректно восстановить принятое слово.

Пусть существует ($k \times n$)-порождающая матрица G. Для того чтобы некоторое множество из k позиций было информационной совокупностью, необходимо, чтобы ранг подматрицы \mathbf{M}_{χ} , составленной из столбцов матрицы G с номерами из χ , был равен k (т. е. \mathbf{M}_{χ} — невырожденная обратимая матрица). С другой стороны, если рассматривается проверочная ($r \times n$)-матрица кода H и χ — информационная совокупность, и $\overline{\chi} = \{1, ..., n\} - \chi$ образует дополнение к χ до множества $\{1, ..., n\}$, то столбцы проверочной матрицы с номерами из $\overline{\chi}$ формируют невырожденную ($r \times r$)-подматрицу.

При исправлении независимых ошибок выполняется случайный поиск информационной совокупности. Для этого необходимо построить такое множество информационных совокупностей, чтобы при искажении передаваемых символов кодового слова (хотя бы в пределах корректирующей способности кода) нашлась неискаженная информационная совокупность — в предположении, что возможно определить это событие. Для обеспечения вероятностей ошибки декодирования, представляющих практический интерес, мощность такого множества должна быть экспоненциальной. При исправлении однократных пакетов ошибок длиной не более b_{max} достаточно найти по одной информационной совокупности вне позиций пакета для каждого возможного расположения пакета, а таких расположений может быть $n - b_{\text{max}} + 1$. Тем не менее, можно обеспечить дальнейшее уменьшение числа информационных совокупностей, если заметить, что информационная совокупность с небольшим разбросом позиций (диаметром, т. е. разностью между максимальным и минимальным номерами позиции в информационной совокупности) может исправлять пакеты в нескольких позициях. Это приводит к идее построения информационных совокупностей ограниченного диаметра, можно назвать их плотными [17]. Более точно, назовем Δ-плотными информационными совокупностями такие, чьи позиции расположены внутри ограниченного интервала длиной $k + \Delta$, где $\Delta \in \{0, n - k - b_{\max}\}$. Конечно, при возможности следует минимизировать значение Δ . В дальнейшем в работе будет рассмотрен только случай $\Delta = 0$, и для упрощения использован термин «плотный» без указания значения Δ .

Нахождение множества плотных информационных совокупностей для кодов Гилберта и их расширения

Рассмотрим возможность построения множества плотных информационных совокупностей для кодов Гилберта, проверочная матрица которых задается в соответствии с формулой (1). Понятие плотной информационной совокупности связано с заданием некоторой величины Δ . Рассмотрим случай $\Delta = 0$, т. е. множества из k подряд идущих позиций. Исследуем вероятность того, что такие множества образуют информационную совокупность, для этого k подряд идущих столбцов порождающей матрицы кода Гилберта (или, эквивалентно, r подряд идущих столбцов проверочной матрицы) образуют невырожденную подматрицу.

На рис. 1, *а* представлены значения рангов $(r \times r)$ -подматриц, составленных из *r* подряд идущих столбцов проверочной матрицы кода Гилберта в зависимости от положения подматриц (на графике по оси абсцисс отмечены номера их начальных позиций) для $\rho = 5$ и для разных размеров блоков *m*. На рис. 1, *b* представлены аналогичные зависимости для $\rho = 8$.

Как видно из графиков на рис. 1, большинство рассмотренных подматриц не являются вырожденными, и следовательно, их позиции образуют информационные совокупности. На некоторых графиках можно видеть уменьшение ранга в узком диапазоне значений позиций, тогда как на других любая подматрица невырождена. Можно заметить, что последний случай имеет место, если числа m и $\rho - 1$ взаимно просты, т. е. не имеют общих делителей кроме единицы. В дальнейшем будем рассматривать только простые значения m. Таким образом, с учетом $\rho \leq m$ указанное условие всегда выполняется.

Для построения множества плотных информационных совокупностей необходимо указать их местоположение и оценить количество.

В [21] предложена процедура построения множества плотных информационных совокупностей для случайных кодов. Для каждой начальной позиции плотной информационной совокупности диаметром $k + \Delta$ оценивается множество начальных позиций пакетов ошибок длиной b_{max} , не затрагивающих данную информационную совокупность. Такие начальные позиции пакетов ошибок называются «безопасными», а в качестве b_{max}

используется максимальная длина исправляемого кодом пакета. С учетом того, что начальные позиции пакетов ошибок находятся в интервале $[1, n - b_{max} - 1]$, процедура построения множества плотных информационных совокупностей заканчивается, когда все позиции из указанного интервала становятся «безопасными».

Выбор расположения информационных совокупностей происходит следующим образом: к текущему началу *i* информационной совокупности (в начальном случае *i* = 1) прибавляется диаметр $k + \Delta$, а затем защитный интервал из $b_{\text{max}} - 1$ позиций. Таким образом, начало следующей информационной совокупности находится в позиции $i + k + \Delta + b_{\text{max}} - 1$ (позиции рассматриваются циклически). В [21] оценена мощность *T* множества плотных информационных совокупностей, полученных таким образом:

$$T = \left[\frac{n - b_{\max} + 1}{n - b_{\max} + 1 - (k + \Delta)} \right].$$
 (3)

Для кода Гилберта значения b_{max} зависят от конкретных значений ρ , поэтому оценим количество полученных информационных совокупностей экспериментально. На рис. 2 представлены значения *T*, полученные с помощью описанной процедуры, для кодов Гилберта при m = 23 в зависимости от значений ρ .

Из графиков на рис. 2 видно, что полученную кривую можно примерно аппроксимировать как 2р/3. Аналогичный результат был получен и для экспериментов с другими значениями *m*.

Теперь рассмотрим расширенные коды Гилберта для $\gamma = 3$, заданные в соответствии с (2). На рис. 3 представлены ранги подматриц для $\rho = 5$ и некоторых простых значений *m*. Пунктирными линиями указаны максимальные ранги (т. е. равные *r*) для соответствующих кодов.

В каждом из рассмотренных случаев (рис. 3) можно выделить в точности р расположений невырожден-



Рис. 1. Графики рангов подматриц для разных размеров блока кода Гилберта с параметром $\gamma = 2$ при $\Delta = 0$ для $\rho = 5$ (*a*) и $\rho = 8$ (*b*)

Fig. 1. Plots of submatrix ranks for different block sizes of the Gilbert code with parameter $\gamma = 2$ at $\Delta = 0$ for $\rho = 5$ (*a*) and $\rho = 8$ (*b*)



Рис. 2. Число плотных информационных совокупностей для кода Гилберта при *m* = 23

Fig. 2. The number of dense information sets for the Gilbert code at m = 23

ных подматриц, соответствующих пиковым значениям рангов, достигающих r. Это означает, что при выборе плотных информационных совокупностей ($\Delta = 0$) существует ограничение только позициями, соответствующими пикам: в остальных позициях плотных информационных совокупностей диаметром k не существует. Заметим, что значения аргументов в точках пиков приходятся на последние позиции блоков проверочной матрицы.

Рассмотрим расширенные коды Гилберта, базовая матрица которых задана тремя строками и $\rho \ge 4$ столбцами (2). Если в качестве начала плотной информационной совокупности диаметром k выбрать значение пика, то по описанной методике следующее возможное начало информационной совокупности будет в точности соответствовать пику, отстоящему на $\rho - 2$ блоков (как и позиции ранее, пики рассматриваются циклически). Таким образом, параметры, используемые в выражении (3), принимают следующие значения: для любого простого m и натурального $\rho \in [4, m]$ имеем $n = m\rho$, $k = m\rho - 3m + 2$, $b_{max} = m - 1$, $\Delta = 0$. Тогда количество плотных информационных совокупностей оценивается выражением

$$T = \left[\frac{m(\rho - 1) + 2}{2m - 1} \right],$$
 (4)

причем, используя тот факт, что [x] = n если и только если $n - 1 < x \le n$ [24], нетрудно показать, что из (4) справедливо $T = [\rho/2]$.

Заметим, что для четного р построение множества информационных совокупностей после р/2 шагов приведет к первоначальному пику. Из выражения (3) следует, что для полученного множества все позиции кодовых слов будут «безопасными».

В случае нечетного ρ числа $\rho - 2$ и ρ взаимно просты, и необходимо вернуться в первоначальный пик только после прохождения по всем остальным пикам. Однако из (3) и (4) следует, что достаточно рассмотреть только последовательность из ($\rho + 1$)/2 пиков, отстоящих друг от друга на $\rho - 2$ (с учетом цикличности).



позиции нахождения информационной совокупности

Рис. 3. График рангов подматриц для простых *m* и ρ = 5 в расширенном коде Гилберта

Fig. 3. Graph of the ranks of submatrices for primes *m* and $\rho = 5$ in the extended Gilbert code

Приведенные выше рассуждения можно сформулировать в виде следующего метода построения множества плотных информационных совокупностей. Для кода Гилберта (1), m – простое, $\rho \in [3, m]$, исправляющего пакеты длиной до b_{max} , вычислим следующую рекуррентную последовательность:

$$\begin{cases} n_1 = 1, \\ n_i = (n_{i-1} - 2m + b_{\max} - 1) \mod m\rho + 1, i = \overline{2, T}, \end{cases}$$

где T рассчитывается в соответствии с (3). Последовательность n_1, \ldots, n_T задает номера начальных позиций множества плотных информационных совокупностей.

Для расширенных кодов Гилберта последовательность *n*₁, ..., *n*_{*T*} имеет вид:

$$\begin{cases} n_1 = m, \\ n_i = (n_{i-1} - 2m - 1) \text{mod}m\rho + 1, i = \overline{2, T}, \end{cases}$$
где $T = \lceil \rho/2 \rceil.$

В таблице приведены значения *T*, полученные по предложенной методике, для кодов Гилберта, их расширений, а также случайных кодов.

Из таблицы видно, что для случайных кодов нахождение множества информационных совокупностей является вероятностным [21]. Проведенное исследование показывает, что для кодов Гилберта и их расширений множество плотных информационных совокупностей диаметром k может быть построено с вероятностью равной единице для простых значений m. При этом количество информационных совокупностей, необходимых для декодирования пакетов ошибок, меньше, чем при использовании случайных кодов. С учетом того, что длины кодов в ансамблях (2, ρ) и (3, ρ) можно устремить к бесконечности с помощью увеличения m, полученные результаты позволяют строить множества информационных совокупностей, мощность которых не изменяется с ростом длины кода.

Используемый код	Длина кода, <i>п</i>	Длина пакета, b _{max}	Параметр плотности, Δ	Размер множества, Т	Вероятность построения множества				
Скорость кода $R \approx 0,33$									
Код Гилберта (γ = 2, ρ = 3)	51	16	0	2	1				
Случайный код	51	12	5	3	0,90				
Код Гилберта (γ = 3, ρ = 4)	68	16	0	2	1				
Случайный код	68	19	7	3	0,97				
Скорость кода $R \approx 0,5$									
Код Гилберта (γ = 2, ρ = 4)	68	14	0	3	1				
Случайный код	68	13	6	4	0,93				
Код Гилберта (γ = 3, ρ = 6)	102	16	0	3	1				
Случайный код	102	20	8	4	0,97				
Скорость кода $R \approx 0,6$									
Код Гилберта (γ = 2, ρ = 5)	85	13	0	4	1				
Случайный код	85	12	7	5	0,95				
Код Гилберта (ү = 3, <i>ρ</i> = 8)	136	16	0	4	1				
Случайный код	136	19	9	6	0,97				

Таблица. Размер множеств плотных информационных совокупностей и вероятности их построения для различных кодов *Table*. The sizes and construction probabilities of dense information sets for different codes

Заключение

Рассмотрена задача построения множества информационных совокупностей минимального диаметра, состоящего из возможно меньшего числа элементов, для кодов Гилберта и их расширения для исправления пакетов ошибок.

Выполнен анализ возможности нахождения информационных совокупностей диаметром k. Показано, что для простых значений размера блока m кодов Гилберта любые k последовательных позиций являются информационной совокупностью. Это позволило предложить методику выбора плотных информационных совокупностей, количество которых можно примерно оценить, как 2 ρ /3, т. е. их число не зависит от длины кода.

Литература

- Величко В.В., Попков Г.В., Попков В.К. Модели и методы повышения живучести современных систем связи. М.: Горячая линия-Телеком, 2014. 270 с.
- Нетес В.А. Основы теории надежности. М.: Горячая линия-Телеком, 2024. 102 с.
- Половко А.М., Гуров С.В. Основы теории надежности. СПб: БХВ-Петербург, 2006. 704 с.
- Bogatyrev A.V., Bogatyrev V.A., Bogatyrev S.V. The probability of timeliness of a fully connected exchange in a redundant real-time communication system // Proc. of the Wave Electronics and its Application in Information and Telecommunication Systems (WECONF). 2020. P. 1–4. https://doi.org/10.1109/ weconf48837.2020.9131517
- Bogatyrev V.A., Bogatyrev A.V., Bogatyrev S.V. Multipath transmission of heterogeneous traffic in acceptable delays with packet replication and destruction of expired replicas in the nodes that make up the path // Communications in Computer and Information Science. 2023. V. 1748. P. 104–121. https://doi.org/10.1007/978-3-031-30648-8_9

Для расширенных кодов Гилберта было выявлено, что для простых значений *m* существует в точности ρ информационных совокупностей диаметром *k*. Предложена процедура выбора плотных информационных совокупностей, отстоящих друг от друга на $\rho - 2$ блока, для исправления любого пакета ошибок в пределах корректирующей способности кода достаточно [$\rho/2$] информационных совокупностей.

Полученные результаты позволяют проводить декодирование рассмотренных кодов с низкой сложностью и гарантированным исправлением однократных пакетов ошибок в пределах корректирующей способности кода.

References

- Velichko V.V., Popkov G.V., Popkov V.K. Models and Methods of Increasing the Survivability of Modern Communication Systems. Moscow, Hotline–Telecom, 2014, 270 p. (in Russian)
- Netes V.A. Fundamentals of Reliability Theory. Moscow, Hotline– Telecom, 2024, 102 p. (in Russian)
- Polovko A.M., Gurov S.V. Fundamentals of Reliability Theory. St. Petersburg BHV- Petersburg, 2006, 704 p. (in Russian)
- Bogatyrev A.V., Bogatyrev V.A., Bogatyrev S.V. The probability of timeliness of a fully connected exchange in a redundant real-time communication system. *Proc. of the Wave Electronics and its Application in Information and Telecommunication Systems* (WECONF), 2020, pp. 1–4. https://doi.org/10.1109/ weconf48837.2020.9131517
- Bogatyrev V.A., Bogatyrev A.V., Bogatyrev S.V. Multipath transmission of heterogeneous traffic in acceptable delays with packet replication and destruction of expired replicas in the nodes that make up the path. *Communications in Computer and Information Science*, 2023, vol. 1748, pp. 104–121. https://doi.org/10.1007/978-3-031-30648-8_9

- Bogatyrev V.A., Bogatyrev S.V., Bogatyrev A.V. Control of multipath transmissions in the nodes of switching segments of reserved paths // Proc. of the International Conference on Information, Control, and Communication Technologies (ICCT). 2022. P. 1–5. https://doi. org/10.1109/ICCT56057.2022.9976839
- Lin S., Li J. Fundamentals of Classical and Modern Error-Correcting Codes. Cambridge, Cambridge University Press, 2022. 840 p.
- Moon T.K. Error correction coding: Mathematical methods and algorithms. Wiley, 2020. 992 p.
- Kulhandjian M., Kulhandjian H., D'Amours C. Improved soft decoding of Reed-Solomon codes on Gilbert-Elliott channels // Proc. of the IEEE International Symposium on Information Theory (ISIT). 2019. P. 1072–1076. https://doi.org/10.1109/ISIT.2019.8849456
- Xiao X., Vasic B., Lin S., Li J., Abdel–Ghaffar K. Quasi-cyclic LDPC codes with parity-check matrices of column weight two or more for correcting phased bursts of erasures // IEEE Transactions on Communications. 2021. V. 69. N 5. P. 2812–2823. https://doi. org/10.1109/TCOMM.2021.3059001
- Li L., Lv J., Li Y., Dai X., Wang X. Burst error identification method for LDPC coded systems // IEEE Communications Letters. 2024. V. 28. N 7. P. 1–5. https://doi.org/10.1109/LCOMM.2024.3391826
- Yang M., Pan Z., Djordjevic I.B. FPGA-based burst-error performance analysis and optimization of regular and irregular SD-LDPC codes for 50G-PON and beyond // Optics Express. 2023. V. 31. N 6. P. 10936–10946. https://doi.org/10.1364/OE.477546
- Aharoni Z., Huleihel B., Pfister H.D., Permuter H.H. Data-Driven polar codes for unknown Channels with and without memory // Proc. of the IEEE International Symposium on Information Theory (ISIT). 2023. P. 1890-1895. https://doi.org/10.1109/ ISIT54713.2023.10206663
- Fang Y., Chen J. Decoding polar codes for a generalized Gilbert-Elliott channel with unknown parameter // IEEE Transactions on Communications. 2021. V. 69. N 10. P. 6455–6468. https://doi. org/10.1109/TCOMM.2021.3095195
- Ghaddar N., Kim Y.-H., Milstein L.B., Ma L., Yi B.K. Joint channel estimation and coding over channels with memory using polar codes // IEEE Transactions on Communications. 2021. V. 69. N 10. P. 6575–6589. https://doi.org/10.1109/TCOMM.2021.3098822
- Ovchinnikov A.A., Veresova A.M., Fominykh A.A. Decoding of linear codes for single error bursts correction based on the determination of certain events // Information and Control Systems. 2022. N 6. P. 41–52. https://doi.org/10.31799/1684-8853-2022-6-41-52
- Исаева М.Н., Овчинников А.А. Исправление одиночных пакетов ошибок за пределами корректирующей способности кода с использованием информационных совокупностей // Научнотехнический вестник информационных технологий, механики и оптики. 2024. Т. 24. № 1. С. 70–80. https://doi.org/10.17586/2226-1494-2024-24-1-70-80
- Исаева М.Н. Поиск информационных совокупностей при исправлении пакетов ошибок квазициклическими кодами // Т-Comm: Телекоммуникации и транспорт. 2023. Т. 17. № 7. С. 4–12. https:// doi.org/10.36724/2072-8735-2023-17-7-4-12
- Крук Е.А., Овчинников А.А. Точная корректирующая способность кодов Гилберта при исправлении пакетов ошибок // Информационно-управляющие системы. 2016. № 1(80). С. 80–87. https://doi.org/10.15217/issn1684-8853.2016.1.80
- Veresova A., Ovchinnikov. Burst detection and correction for Gilbert Codes and its QC-LDPC extensions // Proc. of the IEEE International Multi-Conference on Engineering, Computer and Information Sciences (SIBIRCON). 2024. P. 47–51. https://doi.org/10.1109/ SIBIRCON63777.2024.10758475
- 21. Исаева М.Н. Разработка и анализ методики построения множества плотных информационных совокупностей для исправления пакетов ошибок // Т-Сотт: Телекоммуникации и транспорт. 2024. Т. 18. № 10. С. 20–26. https://doi.org/10.36724/2072-8735-2024-18-10-20-26
- Крук Е.А., Федоренко С.В. Декодирование по обобщенным информационным совокупностям // Проблемы передачи информации. 1995. Т. 31. № 2. С. 54–61.
- Barg A., Krouk E., van Tilborg H.C.A. On the complexity of minimum distance decoding of long linear codes // IEEE Transactions on Information Theory. 1999. V. 45. N 5. P. 1392–1405. https://doi. org/10.1109/18.771141
- Грэхем Р., Кнут Д., Паташник О. Конкретная математика. Основание информатики. М.: Мир, 1998. 703 с.

- Bogatyrev V.A., Bogatyrev S.V., Bogatyrev A.V. Control of multipath transmissions in the nodes of switching segments of reserved paths. *Proc. of the International Conference on Information, Control, and Communication Technologies (ICCT)*, 2022, pp. 1–5. https://doi. org/10.1109/ICCT56057.2022.9976839
- 7. Lin S., Li J. Fundamentals of Classical and Modern Error-Correcting Codes. Cambridge, Cambridge University Press, 2022, 840 p.
- 8. Moon T.K. Error Correction Coding: Mathematical methods and algorithms. Wiley, 2020, 992 p.
- Kulhandjian M., Kulhandjian H., D'Amours C. Improved soft decoding of Reed-Solomon codes on Gilbert-Elliott channels. *Proc. of the IEEE International Symposium on Information Theory* (ISIT), 2019, pp. 1072–1076. https://doi.org/10.1109/ ISIT.2019.8849456
- Xiao X., Vasic B., Lin S., Li J., Abdel–Ghaffar K. Quasi-cyclic LDPC codes with parity-check matrices of column weight two or more for correcting phased bursts of erasures. *IEEE Transactions on Communications*, 2021, vol. 69, no. 5, pp. 2812–2823. https://doi. org/10.1109/TCOMM.2021.3059001
- Li L., Lv J., Li Y., Dai X., Wang X. Burst error identification method for LDPC coded systems. *IEEE Communications Letters*, 2024, vol. 28, no. 7, pp. 1–5. https://doi.org/10.1109/ LCOMM.2024.3391826
- Yang M., Pan Z., Djordjevic I.B. FPGA-based burst-error performance analysis and optimization of regular and irregular SD-LDPC codes for 50G-PON and beyond. *Optics Express*, 2023, vol. 31, no. 6, pp. 10936–10946. https://doi.org/10.1364/OE.477546
- Aharoni Z., Huleihel B., Pfister H.D., Permuter H.H. Data-Driven polar codes for unknown Channels with and without memory. *Proc.* of the IEEE International Symposium on Information Theory (ISIT), 2023, pp. 1890–1895. https://doi.org/10.1109/ ISIT54713.2023.10206663
- Fang Y., Chen J. Decoding polar codes for a generalized Gilbert-Elliott channel with unknown parameter. *IEEE Transactions on Communications*, 2021, vol. 69, no. 10, pp. 6455–6468. https://doi. org/10.1109/TCOMM.2021.3095195
- Ghaddar N., Kim Y.-H., Milstein L.B., Ma L., Yi B.K. Joint channel estimation and coding over channels with memory using polar codes. *IEEE Transactions on Communications*, 2021, vol. 69, no. 10, pp. 6575–6589. https://doi.org/10.1109/TCOMM.2021.3098822
- Ovchinnikov A.A., Veresova A.M., Fominykh A.A. Decoding of linear codes for single error bursts correction based on the determination of certain events. *Information and Control Systems*, 2022, no. 6. pp. 41–52. https://doi.org/10.31799/1684-8853-2022-6-41-52
- Isaeva M.N., Ovchinnikov A.A. Correction of single error bursts beyond the code correction capability using information sets. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 2024, vol. 24, no. 1, pp. 70–80. (in Russian). https://doi.org/10.17586/2226-1494-2024-24-1-70-80
- Isaeva M.N. Finding information sets when correcting error bursts with quasi-cyclic codes. *T-Comm*, 2023, vol. 17, no. 7, pp. 4–12. (in Russian). https://doi.org/10.36724/2072-8735-2023-17-7-4-12
- Krouk E.A., Ovchinnikov A.A. Exact burst-correction capability of Gilbert Codes. *Information and Control Systems*, 2016, no. 1(80), pp. 80–87. (in Russian). https://doi.org/10.15217/issn1684-8853.2016.1.80
- Veresova A., Ovchinnikov. Burst detection and correction for Gilbert Codes and its QC-LDPC extensions. Proc. of the IEEE International Multi-Conference on Engineering, Computer and Information Sciences (SIBIRCON), 2024, pp. 47–51. https://doi.org/10.1109/ SIBIRCON63777.2024.10758475
- Isaeva M.N. Development and analysis of a method for constructing dense information sets for error bursts correction. *T-Comm*, 2024, vol. 18, no. 10, pp. 20–26. (in Russian). https://doi. org/10.36724/2072-8735-2024-18-10-20-26
- Kruk E.A., Fedorenko S.V. Decoding by generalized information sets. *Problemy Peredachi Informatsii*, 1995, vol. 31, no. 2, pp. 54–61. (in Russian)
- Barg A., Krouk E., van Tilborg H.C.A. On the complexity of minimum distance decoding of long linear codes. *IEEE Transactions* on *Information Theory*, 1999, vol. 45, no. 5, pp. 1392–1405. https:// doi.org/10.1109/18.771141
- Graham R.L., Knuth D.E., Patashnik O. Concrete Mathematics: A Foundation for Computer Science. Addison-Wesley Professional, 1994, 672 p.

Автор

Исаева Мария Николаевиа — младший научный сотрудник, Национальный исследовательский университет «Высшая школа экономики», Санкт-Петербург, 190008, Российская Федерация, sc 57243599200, https://orcid.org/0009-0007-6228-0617, misaeva@hse.ru

Статья поступила в редакцию 10.12.2024 Одобрена после рецензирования 19.01.2025 Принята к печати 27.03.2025

Работа доступна по лицензии Creative Commons «Attribution-NonCommercial»

Approved after reviewing 19.01.2025

Received 10.12.2024

Accepted 27.03.2025

Author

Maria N. Isaeva — Junior Researcher, HSE University-St. Petersburg, Saint Petersburg, 190008, Russian Federation, sc 57243599200, https://orcid.org/0009-0007-6228-0617, misaeva@hse.ru

VİTMO

НАУЧНО-ТЕХНИЧЕСКИЙ ВЕСТНИК ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ, МЕХАНИКИ И ОПТИКИ март–апрель 2025 Том 25 № 2 http://ntv.ifmo.ru/ SCIENTIFIC AND TECHNICAL JOURNAL OF INFORMATION TECHNOLOGIES, MECHANICS AND OPTICS March–April 2025 Vol. 25 № 2 http://ntv.ifmo.ru/en/ ISSN 2226-1494 (print) ISSN 2500-0373 (online)

ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ, МЕХАНИКИ И ОПТИКІ

doi: 10.17586/2226-1494-2025-25-2-295-302 УДК 004.021

Предобработка траекторий ключевых точек скелета в задаче автоматизации регистрации поведения лабораторных животных Дмитрий Игоревич Краснов¹[∞], Максим Александрович Волынский², Александр Алексеевич Гусев³

1,2,3 Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация

¹ dmitriy krasnov@outlook.com^{\Box}, https://orcid.org/0009-0003-9469-6618

² maxim.volynsky@gmail.com, https://orcid.org/0000-0003-1679-4450

³ gusew@internet.ru, https://orcid.org/0000-0003-2437-8537

Аннотация

Введение. Автоматизация регистрации поведения лабораторных животных необходима для упрощения проведения поведенческих тестов в исследованиях в области патофизиологии и реабилитации. Регистрация поведения чаще всего выполняется с помощью анализа траекторий ключевых точек скелета. При этом существующие методы привязаны к конкретному виду животного, выбранным скелетным точкам и набору движений для распознавания. В настоящее время отсутствует математическая формулировка задачи, а алгоритмы фильтрации полученных траекторий недостаточно исследованы. В работе создан набор данных для детекции ключевых точек крыс линии Wistar и выполнена оценка алгоритмов фильтрации траекторий по зашумленным измерениям. При этом в данной задаче важна не только близость отфильтрованной траектории к реальной, но и гладкость полученной кривой. Метод. В рассмотренной скелетной модели крысы выбраны 13 точек, позволяющих оценить ее поведение по траекториям. Полученный набор данных позволяет обучить сверточную нейронную сеть для детекции ключевых точек скелета. Представлено математическое описание динамики движения точек между кадрами для использования в фильтре Калмана. Проведено сравнение четырех алгоритмов фильтрации по критериям точности и гладкости кривой. Основные результаты. Разработана методика построения ковариационной матрицы шума детектора по анализу ошибок детекции ключевых точек. Показана применимость фильтра Калмана в задаче оценки траекторий ключевых точек скелета по зашумленным измерениям. Сравнение алгоритмов фильтрации показало, что в данной задаче лучшие результаты показывает Unscented фильтр Калмана с нелинейной моделью и фильтр скользящего среднего. Обсуждение. Полученное математическое описание динамики движения ключевых точек скелета может быть использовано для оценки реальной траектории по зашумленным измерениям. При этом описанные методы применимы не только для лабораторных животных, но и для человека. Результаты оценки нескольких алгоритмов фильтрации могут в дальнейшем облегчить выбор алгоритма для использования в подобных задачах. Предобработка траекторий поможет в будущем уменьшить ошибки классификатора поведения и повысить его точность.

Ключевые слова

компьютерное зрение, лабораторные животные, классификация поведения, анализ траекторий, сглаживание траекторий, фильтр Калмана, фильтр скользящего среднего

Благодарности

Авторы выражают благодарность д.м.н. Алексею Андреевичу Шмонину (Первый Санкт-Петербургский государственный медицинский университет им. акад. И.П. Павлова) за постановку задачи и предоставление записей видеонаблюдения за поведением крыс.

Ссылка для цитирования: Краснов Д.И., Волынский М.А., Гусев А.А. Предобработка траекторий ключевых точек скелета в задаче автоматизации регистрации поведения лабораторных животных // Научнотехнический вестник информационных технологий, механики и оптики. 2025. Т. 25, № 2. С. 295–302. doi: 10.17586/2226-1494-2025-25-2-295-302

[©] Краснов Д.И., Волынский М.А., Гусев А.А., 2025

Preprocessing of skeletal keypoints trajectories in the task of laboratory animal behavior recording automation

Dmitrii I. Krasnov^{1⊠}, Maxim A. Volynsky², Alexander A. Gusev³

^{1,2,3} ITMO University, Saint Petersburg, 197101, Russian Federation

¹ dmitriy krasnov@outlook.com^{\Box}, https://orcid.org/0009-0003-9469-6618

² maxim.volynsky@gmail.com, https://orcid.org/0000-0003-1679-4450

³ gusew@internet.ru, https://orcid.org/0000-0003-2437-8537

Abstract

The automation of action recognition in laboratory animals is a crucial step in simplifying behavioral tests in the fields of pathophysiology and rehabilitation research. The most common method of action recognition is to analyze the trajectories of key skeletal points. However, the existing methods are strongly tied to the specific animal species, selected skeletal points, and set of activities to be recognized. Furthermore, there is a dearth of mathematical formulations of this problem and research on algorithms for filtering obtained trajectories. The research task involves the collection of a dataset for key points detection of Wistar rats and evaluation of algorithms for filtering trajectories from noisy measurements. In considered skeletal model of the rat, a total of thirteen points were selected for the purpose of estimating the behavior along trajectories. A mathematical description of the dynamics of point movement between frames for use in a Kalman filter is provided. Four filtering algorithms are evaluated in terms of accuracy and curve smoothness. The technique of constructing the covariance matrix of the detector noise by analyzing the key point detection errors is developed. The comparison of filtering algorithms shows that the Unscented Kalman filter with nonlinear model and moving average filter yield the most optimal results in this task. The findings of this study allow the use of a mathematical description of system dynamics to estimate the actual trajectory from noisy measurements. Furthermore, the described methodologies are not exclusive to laboratory animals, but can also be applied to human subjects.

Keywords

computer vision, laboratory animals, behavior classification, trajectory analysis, trajectory smoothing, Kalman filter, moving average filter

Acknowledgements

The authors are grateful to M.D., D.Sc. (Medicine) Alexey Andreevich Shmonin (First Pavlov State Medical University of St. Petersburg) for setting the task and providing video recordings of rats' behavior.

For citation: Krasnov D.I., Volynsky M.A., Gusev A.A. Preprocessing of skeletal keypoints trajectories in the task of laboratory animal behavior recording automation. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 2025, vol. 25, no. 2, pp. 295–302 (in Russian). doi: 10.17586/2226-1494-2025-25-2-295-302

Введение

Лабораторные животные активно используются для изучения механизмов возникновения и протекания различных заболеваний человека, исследования и разработки возможных подходов к лечению. Крысы широко применяются в исследованиях, направленных на лечение инсульта, поскольку физиология крыс позволяет в лабораторных условиях воспроизводить состояния, аналогичные человеческим. В большинстве опытов с моделированием инсульта у крыс развиваются симптомы в виде слабости в передней и задней лапах с одной стороны, снижения поведенческой активности и координации движений. Однако через некоторое время все симптомы проходят, а по неврологическим тестам изменения не выявляются. Наблюдения за животными в течение времени показали, что крыса не ухаживает за собой, мало двигается, меньше взаимодействует с другими крысами. Похожие наблюдения можно сделать у людей при инсульте, инфаркте миокарда, онкологических заболеваниях и хронических болях. Таким образом, формируется инвалидность, где причиной потери навыков являются как симптомы заболевания, так и потеря деятельности, т. е. вторичные изменения. При этом поведенческие изменения у крыс в процессе реабилитации не являются до конца изученными, поскольку для каждого заболевания и каждой особи характерны уникальные поведенческие профили.

Например, в работе [1] исследовано качество моделей в экспериментах инсульта у крыс и сделан вывод о том, что в данной области не хватает поведенческих тестов. В работах [2, 3] сделан акцент в необходимости автоматизации для повышения воспроизводимости поведенческих тестов, поскольку чаще всего такие тесты выполняются научными сотрудниками вручную.

В настоящее время существует ряд способов регистрации поведения лабораторных животных методами компьютерного зрения. Подавляющее большинство из них основывается на анализе пространственно-временной информации, полученной из последовательности изображений. В качестве такой информации наиболее часто выступают траектории ключевых точек, образующих скелетную модель животного [4] или человека [5]. В существующих подходах к определению ключевых точек лабораторных животных для последующей регистрации поведения возникает большое количество особенностей.

Главная особенность — ограниченность применения того или иного решения в рамках конкретной задачи вместе с большой вариативностью таких задач (в зависимости от вида лабораторных животных). В задачах регистрации поведения грызунов вариативность исходит из вида животного (крысы или мыши), цвета (белый, черный, серый или пятнистый) и количества особей (одна, две или более). Например, в работе [6] рассмотрены взаимодействия черной и белой мышей, в [7] выполнен анализ поведения черной и коричневой мышей, а в [8] — поведение пятнистой темной крысы и т. п. В результате предложенные методы определения ключевых точек не способны работать с особями, отличными от тех, на изображениях которых обучалась модель. Ситуацию усложняет возможное наличие посторонних объектов в клетке с животными (миски с едой и водой, домики и т. д.).

Вторая особенность — определение ключевых точек, образующих скелет. Чаще всего определяемый скелет включает в себя нос, глаза, уши, крайние точки пояса задних конечностей и основание хвоста [9], но может включать и другие точки, такие как отделенные точки позвоночника, середина и кончик хвоста, пояс передних конечностей. При этом каждый существующий метод использует уникальный набор точек, что усложняет их адаптацию.

Третья особенность — ракурс. Чаще всего для регистрации ключевых точек используется одиночная камера с ракурсом снизу-вверх [7] или сверху-вниз [6], при этом точки распознаются в 2D пространстве. Однако существуют методы [10], использующие несколько камер с последующей обработкой для реконструкции трехмерной траектории. Перечисленные особенности приводят к тому, что в случае существенного отличия в постановке задачи (другой цвет крысы, другое расположение камеры и т. п.), детектор не способен с достаточной точностью находить требуемые ключевые точки, что приводит к необходимости сбора нового набора данных.

Немаловажным фактором для правильной регистрации поведения является гладкость траекторий точек скелета. Движение крысы между кадрами можно охарактеризовать как отрывистое с частым изменением направления. Это приводит к тому, что траектории точек и их производные не являются гладкими. Такие выбросы вместе с погрешностью детектора точек могут привести к ложным срабатываниям классификатора поведения. В работе [10] показано, что фильтрация точек скелета необходима для более точного определения их положения в кадре. Для фильтрации и сглаживания траекторий в задачах трекинга часто используется фильтр Калмана с соответствующей математической моделью движения [11, 12]. Однако такой подход ранее не применялся в задаче трекинга точек скелета лабораторных животных.

Таким образом, в настоящей работе рассматривается задача фильтрации траекторий ключевых точек скелета, которые в будущем будут использоваться для классификации поведения взрослых особей крыс линии Wistar обоих полов. Целями работы являются математическое описание кинематики движения ключевых точек крысы и сравнительный анализ границ применимости существующих алгоритмов фильтрации для данной задачи. Для этого использовались две модели движения: линейная модель равноускоренного движения с линейным фильтром Калмана и нелинейная модель Constant Turn Rate and Acceleration (CTRA) [13] с Unscented Kalman Filter (UKF) [14]. Для сравнения с описанными выше методами была выполнена фильтрация с помощью One Euro Filter [15] и фильтра скользящего среднего.

Поскольку задача анализа поведения крыс является достаточно специфической, в работе использовался собранный и размеченный вручную набор данных, на котором обучался детектор ключевых точек.

Описание ключевых точек скелета крысы

Для обучения детектора ключевых точек был собран и размечен набор данных, содержащий изображения особей крыс Wistar. Скелетная модель крысы состоит из 13-ти точек: нос (1); правый и левый глаз (2 и 3); правое и левое ухо (5 и 6); средняя точка головы (4); средняя точка пояса передних конечностей (7); средняя точка позвоночника (8); правая и левая передняя конечность (9 и 10); основание хвоста (11); правая и левая задняя конечность (12 и 13). Отметим, что в аналогичных работах [7-9] для анализа поведения выбраны разные ключевые точки, что определяется конкретной задачей, однако некоторые точки совпадают. В настоящей работе ключевые точки скелета выбраны после консультации с экспертом, который их использует для визуального определения поведения крысы. Пример расположения точек скелета на крысе представлен на рис. 1. Данные были размечены в программе Computer Vision Annotation Tool (CVAT). Итоговый набор данных содержит 1762 изображения разрешением 1060 × 548 пикселов с соответствующими аннотациями ключевых точек.

Для определения координат ключевых точек скелета на видео была обучена сверточная нейронная сеть YOLOv8m [16] на итоговом наборе данных. На рис. 2 представлена столбчатая диаграмма дисперсий ошибок обученной модели для координат *x* и *y* каждой ключевой точки скелета. Дисперсия ошибок для одной ключевой точки вычислялась по формуле

$$\sigma_x^2 = \frac{1}{N-1} \sum_{i=1}^{N} (\Delta x_i - \Delta x_m)^2,$$

где N — количество изображений в выборке; Δx_i — разность между истинным и предсказанным сверточной сетью положением ключевой точки по координате x на *i*-ом изображении из выборки; Δx_m — среднее значение Δx для всех изображений из выборки.

Для координаты у дисперсия ошибок вычислена аналогично. Эти дисперсии могут использоваться для



Puc. 1. Расположение точек скелета на крысе *Fig. 1.* Skeletal points position on the rat



Puc. 2. Диаграмма дисперсий ошибок детектора *Fig. 2.* Diagram of detector errors variances

оценки точности детектора ключевых точек, а также для построения ковариационной матрицы шума измерения в фильтре Калмана.

Математическая постановка задачи фильтрации траекторий ключевых точек

Траектории движения точек скелета крысы на последовательности изображений не являются гладкими. Это происходит из-за переменного резкого движения крысы в совокупности с ошибками детектора точек. Перед классификацией поведения такие траектории необходимо обработать, чтобы исключить влияние выбросов на результат. Для этого можно использовать линейный фильтр Калмана [17] с линейной моделью движения. Фильтр Калмана широко используется в задачах, где необходима оценка неизмеряемых переменных состояния и фильтрация зашумленных показаний детектора.

Сделаем допущение, что точки скелета крысы имеют постоянное ускорение между кадрами, т. е. между измерениями. Тогда их движение можно считать равноускоренным и описать соответствующим кинематическим уравнением. Представим вектор состояния *i*-ой точки скелета крысы как вектор-столбец

$$\mathbf{p}_i = [x \quad y \quad v_x \quad v_y \quad a_x \quad a_y]^T,$$

где x, y — координаты точки в кадре; v_x, v_y — линейные скорости; a_x, a_y — линейные ускорения; $i = \overline{1,13}$.

Тогда модель равноускоренного движения для линейного дискретного фильтра Калмана описывается уравнением

p ₁		Α	0	•••	0	p ₁	
p ₂	_	0	A	•••	0	p ₂	+ 11
÷		:	÷	·.	:	:	••• <i>k</i>
p ₁₃	k	0	0		A	p ₁₃	<i>k</i> -1

где **А** — матрица движения для одной точки; **w**_k — вектор шума процесса; *k* — номер итерации фильтра. Матрица движения точки определяется как

	1	0	Δt	0	$0,5\Delta t^2$	0
	0	1	0	Δt	0	$0,5\Delta t^2$
Δ =	0	0	1	0	Δt	0
A -	0	0	0	1	0	Δt
	0	0	0	0	1	0
	0	0	0	0	0	1

где Δt — интервал времени между измерениями.

Вектор измерения представляет собой вектор-столбец с координатами точек скелета, а модель измерения имеет вид

$$\begin{bmatrix} x_1 \\ y_1 \\ \vdots \\ x_{13} \\ y_{13} \end{bmatrix}_k = \begin{bmatrix} \mathbf{H} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{H} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{H} \end{bmatrix} \begin{bmatrix} \mathbf{p}_1 \\ \mathbf{p}_2 \\ \vdots \\ \mathbf{p}_{13} \end{bmatrix}_{k-1} + \mathbf{r}_k$$

где **H** — матрица измерения для одной точки; \mathbf{r}_k — вектор шума измерения.

В свою очередь матрица измерения для одной точки определяется как

$$\mathbf{H} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

Ковариационная матрица шума измерения состоит из дисперсий ошибок детектора для каждой координаты каждой точки на главной диагонали (рис. 2). Определим ковариационную матрицу шума процесса для одной точки скелета и для всего скелета, которая составляется вдоль главной диагонали аналогично матрице перехода **A**. Пусть шум процесса определяется шумом ускорения w_a с дисперсией σ_a^2 , возникающим из-за неучтенных в модели факторов. Тогда вектор шума процесса для одной точки можно представить в виде:

$$\mathbf{w}_k = \begin{bmatrix} 0, 5\Delta t^2 & 0, 5\Delta t^2 & \Delta t & \Delta t & 1 & 1 \end{bmatrix}^T w_a$$

Ковариационная матрица для одной точки может быть вычислена через произведение

Научно-технический вестник информационных технологий, механики и оптики, 2025, том 25, № 2 Scientific and Technical Journal of Information Technologies, Mechanics and Optics, 2025, vol. 25, no 2

$$\mathbf{Q} = \operatorname{cov}(\mathbf{w}_k) = \sigma_a^2 [0, 5\Delta t^2 \quad 0, 5\Delta t^2 \quad \Delta t \quad \Delta t \quad 1 \quad 1]^T \times \\ \times [0, 5\Delta t^2 \quad 0, 5\Delta t^2 \quad \Delta t \quad \Delta t \quad 1 \quad 1],$$

а для всех точек скелета строится аналогично матрице перехода динамической модели.

Для более подробного описания динамики точек можно воспользоваться нелинейной моделью CTRA с UKF, что широко используется для трекинга, например, маневрирующих автомобилей. Данная модель использует допущение о том, что точки скелета крысы перемещаются с постоянным линейным ускорением и угловой скоростью. Тогда вектор состояния *i*-ой точки скелета можно представить в виде:

$$\mathbf{p}_i = [x \quad y \quad v \quad a \quad \phi \quad \omega]^T,$$

где *v* — линейная скорость; *a* — линейное ускорение; φ — угол между вектором скорости и осью *OX*; ω – угловая скорость поворота.

Модель СТRА с UKF движения точки скелета описывается уравнениями, приведенными в работе [13]. Модель измерения и ковариационная матрица шума измерения аналогичны линейной модели.

Методы

Описанные в разделе «Математическая постановка задачи фильтрации траекторий ключевых точек» линейная и нелинейная кинематические модели движения использовались для построения линейного фильтра Калмана [17] и UKF. Для сравнения с фильтрами Калмана были взяты более простые фильтры, такие как One Euro фильтр и фильтр скользящего среднего. Координаты ключевых точек скелета на каждом кадре определялись с помощью предварительно обученной сверточной нейронной сети на размеченном наборе данных (рис. 2). Совокупность координат ключевых точек на последовательных кадрах видеозаписи формировали траектории этих ключевых точек, которые затем фильтровались перечисленными фильтрами. Фильтры реализованы на языке программирования Python с использованием библиотек NumPy и PyTorch. Вычисления проводились на процессоре Intel Core i7-11700F, обучение сверточной нейронной сети — на NVIDIA GeForce RTX 3060 12GB.

Для сравнительного анализа алгоритмов фильтрации траекторий точек скелета была выбрана последовательность из 900 изображений, где крыса непрерывно перемещается в клетке. Для анализа использовалась наиболее часто перемещающаяся точка крысы — нос. В эксперименте оценивались близость отфильтрованной траектории к истинной траектории и гладкость. Близость траектории определялась по евклидовому расстоянию между точкой отфильтрованной траектории и точкой истинной траектории на каждом кадре (ошибка)

$$\Delta e = \sqrt{(x_{pred} - x_{true})^2 + (y_{pred} - y_{true})^2},$$

где *x_{pred}*, *y_{pred}* — координаты *x*, *y* точки отфильтрованной траектории; *x_{true}*, *y_{true}* — координаты *x*, *y* точки истинной траектории.

Гладкость траектории рассчитывалась по евклидовому расстоянию между одной и той же точкой на соседних кадрах (смещение)

$$\Delta l = \sqrt{(x_i - x_{i-1})^2 + (y_i - y_{i-1})^2},$$

где x_i, y_i — координаты x, y точки отфильтрованной траектории на текущем кадре; x_{i-1}, y_{i-1} — координаты x, yточки отфильтрованной траектории на предыдущем кадре.

Результаты и обсуждение

В качестве примера на рис. 3 представлены результаты фильтрации CTRA с UKF. Для остальных методов полученные траектории аналогичны. Для сравнения различных методов фильтрации была построена коробчатые диаграммы ошибок (рис. 4, *a*) и смещений точек между соседними кадрами (рис. 4, *b*). Данные диаграммы показывают точность метода фильтрации



Puc. 3. Результаты фильтрации траектории UKF *Fig. 3.* Results of trajectory filtering with UKF



Puc. 4. Коробчатые диаграммы для сравнения фильтров: отклонение траектории (*a*); оценка гладкости траектории (*b*) *Fig. 4.* Box plots for filters comparison: trajectory error (*a*); smoothness evaluation (*b*)

и гладкость отфильтрованной траектории соответственно.

На рис. 4 видно, что нелинейная модель CTRA с UKF лучше описывает сложное и изменчивое движение точек скелета крысы, чем линейная модель. Линейный фильтр Калмана не успевает перестроиться при резком изменении направления, поскольку оно меняется слишком часто. При этом нелинейная модель CTRA с UKF дает хорошее приближение даже на сложных участках с частой сменой направления и скорости. Однако можно увидеть участки, где отфильтрованная траектория значительно расходится с реальной, что также связано с длительной перестройкой фильтра. Опе Euro фильтр показывает лучшее приближение к реальной траектории и худшую гладкость кривой. Похожие результаты можно увидеть при использовании фильтра скользящего среднего, при этом траектория получается более гладкой. Можно заметить, что результаты фильтрации UKF и более простыми фильтрами близки. На практике выбор фильтра предполагает компромисс между точностью и гладкостью, поскольку высокая точность соответствует меньшей гладкости и наоборот. В данном случае по рис. 4 видно, что СТКА с UKF, One Euro фильтр и фильтр скользящего среднего имеют примерно близкие характеристики, однако фильтр скользящего среднего наиболее просто реализуем, вычислительно менее сложен и позволяет распараллелить вычисления.

Заключение

В работе создан набор данных с изображениями особей крыс линии Wistar и соответствующей разметкой ключевых точек скелета, на котором обучена сверточная нейронная сеть для детекции этих точек. Проведен анализ алгоритмов предобработки траекторий точек скелета для последующей регистрации поведения крысы. Описаны математические модели динамики движения ключевых точек скелета для использования фильтра Калмана.

Результаты анализа траекторий показали, что нелинейная модель движения CTRA с UKF описывает динамику процесса движения ключевых точек скелета лучше, чем линейная модель равноускоренного движения. Более простые фильтры, такие как One Euro фильтр и фильтр скользящего среднего, также применимы в данной задаче, при этом они оказываются менее затратными вычислительно и достаточно сглаживают траекторию.

В данной работе описан вариант наблюдения за белой особью крысы Wistar с ракурса сверху-вниз.

Описанные динамические модели движения и алгоритмы фильтрации применимы для любого лабораторного животного (крысы, мыши, мухи и т. п.), наблюдаемого с ракурса сверху-вниз, поскольку от вида и цвета животного зависит только метод обнаружения ключевых точек на каждом кадре.

Перспективной задачей является расширение набора данных, включение более разнообразных активностей крысы (где она умывается, чешется, ест и т. п.) и изображений с несколькими особями вместе (особи отличаются полом и наличием или отсутствием заболеваний). Дальнейшая работа по регистрации поведения на основе отфильтрованных траекторий требует

Литература

- Friedrich J., Lindauer U., Höllig A. Procedural and methodological quality in preclinical Stroke Research-A Cohort Analysis of the Rat MCAO model comparing periods before and after the publication of STAIR/ARRIVE // Frontiers in Neurology. 2022. V. 13. P. 834003. https://doi.org/10.3389/fneur.2022.834003
- Gulinello M., Mitchell H.A., Chang Q., O'Brien W.T., Zhou Z., Abel T., Wang L., Corbin J.G., Veeraragavan S., Samaco R.C., Andrews N.A., Fagiolini M., Cole T.B., Burbacher T.M., Crawley J.N. Rigor and reproducibility in rodent behavioral research // Neurobiology of Learning and Memory. 2019. V. 165. P. 106780 https://doi.org/10.1016/j.nlm.2018.01.001
- Altun M., Bergman E., Edström E., Johnson H., Ulfhake B. Behavioral impairments of the aging rat // Physiology & Behavior. 2007. V. 92. N 5. P. 911–923. https://doi.org/10.1016/j. physbeh.2007.06.017
- Lauer J., Zhou M., Ye S., Menegas W., Schneider S., Nath T., Rahman M.M., Di Santo V., Soberanes D., Feng G.P., Murthy V.N., Lauder G., Dulac C., Mathis M.W., Mathis A. Multi-animal pose estimation, identification and tracking with DeepLabCut // Nature Methods. 2022. V. 19. N 4. P. 496–504. https://doi.org/10.1038/ s41592-022-01443-0
- Song L., Yu G., Yuan J., Liu Z. Human pose estimation and its application to action recognition: A survey // Journal of Visual Communication and Image Representation. 2021. V. 76. P. 103055 https://doi.org/10.1016/j.jvcir.2021.103055
- Segalin C., Williams J., Karigo T., Hui M., Zelikowsky M., Sun J.J., Perona P., Anderson D.J., Kennedy A. The Mouse Action Recognition System (MARS) software pipeline for automated analysis of social behaviors in mice // eLife. 2021. V. 10. P. e63720. https://doi. org/10.7554/eLife.63720
- Hsu A.I., Yttri E.A. B-SOiD, an open-source unsupervised algorithm for identification and fast prediction of behaviors // Nature Communications. 2021. V. 12. N 1. P. 5188. https://doi.org/10.1038/ s41467-021-25420-x
- Monsees A., Voit K.M., Wallace D.J., Sawinski J., Charyasz E., Scheffler K., Macke J.H., Kerr J.N.D. Estimation of skeletal kinematics in freely moving rodents // Nature Methods. 2022. V. 19. N 11. P. 1500–1509. https://doi.org/10.1038/s41592-022-01634-9
- Pereira T.D., Tabris N., Matsliah A. Turner D.M., Li J., Ravindranath S., Papadoyannis E.S., Normand E., Deutsch D.S., Wang Z.Y., McKenzie-Smith G.C., Mitelut C.C., Castro M.D., D'Uva J., Kislin M., Sanes D.H., Kocher S.D., Wang S.S.H., Falkner A.L., Shaevitz J.W., Murthy M. SLEAP: A deep learning system for multi-animal pose tracking // Nature Methods. 2022. V. 19. N 4. P. 486–495. https://doi.org/10.1038/s41592-022-01426-1
- Karashchuk P., Rupp K.L., Dickinson E.S., Walling-Bell S., Sanders E., Azim E., Brunton B.W., Tuthill J.C. Anipose: A toolkit for robust markerless 3D pose estimation // Cell Reports. 2021. V. 36. N 13. P. 109730. https://doi.org/10.1016/j.celrep.2021.109730
- Kam H.C., Yu Y.K., Wong K.H. An improvement on ArUco marker for pose tracking using kalman filter // Proc. of the 19th IEEE/ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD). 2018. P. 65–69. https://doi.org/10.1109/SNPD.2018.8441049

математической постановки задачи оценки вектора состояния. Планируется классифицировать широкий набор движений, состоящий из индивидуального и социального поведений, например умывание, чесание, сон, прием пищи, драка, обнюхивание и т. п. В результате предполагается достичь достаточной точности распознавания всех классов поведения для автоматизации поведенческих тестов.

Набор данных с изображениями крыс линии Wistar, используемый в работе, является уникальным и может быть предоставлен по запросу вместе с исходным кодом (Краснов Д.И., dmitriy_krasnov@outlook.com).

References

- Friedrich J., Lindauer U., Höllig A. Procedural and methodological quality in preclinical Stroke Research-A Cohort Analysis of the Rat MCAO model comparing periods before and after the publication of STAIR/ARRIVE. *Frontiers in Neurology*, 2022, vol. 13, pp. 834003. https://doi.org/10.3389/fneur.2022.834003
- Gulinello M., Mitchell H.A., Chang Q., O'Brien W.T., Zhou Z., Abel T., Wang L., Corbin J.G., Veeraragavan S., Samaco R.C., Andrews N.A., Fagiolini M., Cole T.B., Burbacher T.M., Crawley J.N. Rigor and reproducibility in rodent behavioral research. *Neurobiology* of Learning and Memory, 2019, vol. 165, pp. 106780. https://doi. org/10.1016/j.nlm.2018.01.001
- Altun M., Bergman E., Edström E., Johnson H., Ulfhake B. Behavioral impairments of the aging rat. *Physiology & Behavior*, 2007, vol. 92, no. 5, pp. 911–923. https://doi.org/10.1016/j. physbeh.2007.06.017
- Lauer J., Zhou M., Ye S., Menegas W., Schneider S., Nath T., Rahman M.M., Di Santo V., Soberanes D., Feng G.P., Murthy V.N., Lauder G., Dulac C., Mathis M.W., Mathis A. Multi-animal pose estimation, identification and tracking with DeepLabCut. *Nature Methods*, 2022, vol. 19, no. 4, pp. 496–504. https://doi.org/10.1038/s41592-022-01443-0
- Song L., Yu G., Yuan J., Liu Z. Human pose estimation and its application to action recognition: A survey. *Journal of Visual Communication and Image Representation*, 2021, vol. 7, pp. 103055. https://doi.org/10.1016/j.jvcir.2021.103055
- Segalin C., Williams J., Karigo T., Hui M., Zelikowsky M., Sun J.J., Perona P., Anderson D.J., Kennedy A. The Mouse Action Recognition System (MARS) software pipeline for automated analysis of social behaviors in mice. *eLife*, 2021, vol. 10, pp. e63720. https://doi. org/10.7554/eLife.63720
- Hsu A.I., Yttri E.A. B-SOiD, an open-source unsupervised algorithm for identification and fast prediction of behaviors. *Nature Communications*, 2021, vol. 12, no. 1, pp. 5188. https://doi. org/10.1038/s41467-021-25420-x
- Monsees A., Voit K.M., Wallace D.J., Sawinski J., Charyasz E., Scheffler K., Macke J.H., Kerr J.N.D. Estimation of skeletal kinematics in freely moving rodents. *Nature Methods*, 2022, vol. 19, no. 11, pp. 1500–1509. https://doi.org/10.1038/s41592-022-01634-9
- Pereira T.D., Tabris N., Matsliah A. Turner D.M., Li J., Ravindranath S., Papadoyannis E.S., Normand E., Deutsch D.S., Wang Z.Y., McKenzie-Smith G.C., Mitelut C.C., Castro M.D., D'Uva J., Kislin M., Sanes D.H., Kocher S.D., Wang S.S.H., Falkner A.L., Shaevitz J.W., Murthy M. SLEAP: A deep learning system for multi-animal pose tracking. *Nature Methods*, 2022, vol. 19, no. 4, pp. 486–495. https://doi.org/10.1038/s41592-022-01426-1
- Karashchuk P., Rupp K.L., Dickinson E.S., Walling-Bell S., Sanders E., Azim E., Brunton B.W., Tuthill J.C. Anipose: A toolkit for robust markerless 3D pose estimation. *Cell Reports*, 2021, vol. 36, no. 13, pp. 109730. https://doi.org/10.1016/j.celrep.2021.109730
- Kam H.C., Yu Y.K., Wong K.H. An improvement on ArUco marker for pose tracking using kalman filter. Proc. of the 19th IEEE/ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD), 2018, pp. 65–69. https://doi.org/10.1109/ SNPD.2018.8441049

- Buizza C., Fischer T., Demiris Y. Real-Time Multi-Person pose tracking using data assimilation // Proc. of the IEEE Winter Conference on Applications of Computer Vision (WACV). 2020. P. 438–447. https://doi.org/10.1109/WACV45572.2020.9093442
- Svensson D. Derivation of the discrete-time constant turn rate and acceleration motion model // Proc. of the Sensor Data Fusion: Trends, Solutions, Applications (SDF). 2019. P. 1–5. https://doi.org/10.1109/ SDF.2019.8916654
- Wan E.A., Van Der Merwe R. The unscented Kalman filter for nonlinear estimation // Proc. of the IEEE 2000 Adaptive Systems for Signal Processing, Communications, and Control Symposium. 2000. P. 153–158. https://doi.org/10.1109/ASSPCC.2000.882463
- Casiez G., Roussel N., Vogel D. 1 € filter: a simple speed-based lowpass filter for noisy input in interactive systems // Proc. of the SIGCHI Conference on Human Factors in Computing Systems (CHI '12). 2012. P. 2527–2530. https://doi.org/10.1145/2207676.2208639
- Varghese R., Sambath M. YOLOv8: a novel object detection algorithm with enhanced performance and robustness // Proc. of the International Conference on Advances in Data Engineering and Intelligent Computing Systems (ADICS). 2024. P. 1–6. https://doi. org/10.1109/ADICS58448.2024.10533619
- Kalman R.E. A new approach to linear filtering and prediction problems // Journal of Fluids Engineering. 1960. V. 82. N 1. P. 35–45. https://doi.org/10.1115/1.3662552

Авторы

Краснов Дмитрий Игоревич — аспирант, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, sc 59411982500, https://orcid.org/0009-0003-9469-6618, dmitriy_krasnov@outlook.com Вольнский Максим Александрович — кандидат технических наук, доцент, директор научно-образовательной лаборатории «Техническое зрение», доцент, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, sc 23006901100, https://orcid.org/0000-0003-1679-4450, maxim.volynsky@gmail.com

Гусев Александр Алексеевич — кандидат технических наук, ведущий инженер, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, вс 57207731147, https://orcid.org/0000-0003-2437-8537, gusew@internet.ru

Статья поступила в редакцию 04.11.2024 Одобрена после рецензирования 21.01.2025 Принята к печати 18.03.2025



- Buizza C., Fischer T., Demiris Y. Real-Time Multi-Person pose tracking using data assimilation. Proc. of the IEEE Winter Conference on Applications of Computer Vision (WACV), 2020, pp. 438–447. https://doi.org/10.1109/WACV45572.2020.9093442
- Svensson D. Derivation of the discrete-time constant turn rate and acceleration motion model. *Proc. of the Sensor Data Fusion: Trends, Solutions, Applications (SDF)*, 2019, pp. 1–5. https://doi.org/10.1109/ SDF.2019.8916654
- Wan E.A., Van Der Merwe R. The unscented Kalman filter for nonlinear estimation. *Proc. of the IEEE 2000 Adaptive Systems for Signal Processing, Communications, and Control Symposium*, 2000, pp. 153–158. https://doi.org/10.1109/ASSPCC.2000.882463
- Casiez G., Roussel N., Vogel D. 1 € filter: a simple speed-based lowpass filter for noisy input in interactive systems. *Proc. of the SIGCHI Conference on Human Factors in Computing Systems (CHI '12)*, 2012, pp. 2527–2530. https://doi.org/10.1145/2207676.2208639
- Varghese R., Sambath M. YOLOv8: a novel object detection algorithm with enhanced performance and robustness. *Proc. of the International Conference on Advances in Data Engineering and Intelligent Computing Systems (ADICS)*, 2024, pp. 1–6. https://doi. org/10.1109/ADICS58448.2024.10533619
- Kalman R.E. A new approach to linear filtering and prediction problems. *Journal of Fluids Engineering*, 1960, vol. 82, no. 1, pp. 35– 45. https://doi.org/10.1115/1.3662552

Authors

Dmitrii I. Krasnov — PhD Student, ITMO University, Saint Petersburg, 197101, Russian Federation, sc 59411982500, https://orcid.org/0009-0003-9469-6618, dmitriy_krasnov@outlook.com

Maxim A. Volynsky — PhD, Associate Professor, Director (Technical Vision Laboratory), Associate Professor, ITMO University, Saint Petersburg, 197101, Russian Federation, sc 23006901100, https://orcid. org/0000-0003-1679-4450, maxim.volynsky@gmail.com

Alexander A. Gusev — PhD, Leading Engineer, ITMO University, Saint Petersburg, 197101, Russian Federation, sc 57207731147, https://orcid. org/0000-0003-2437-8537, gusew@internet.ru

Received 04.11.2024 Approved after reviewing 21.01.2025 Accepted 18.03.2025

Работа доступна по лицензии Creative Commons «Attribution-NonCommercial» **VİTMO**

НАУЧНО-ТЕХНИЧЕСКИЙ ВЕСТНИК ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ, МЕХАНИКИ И ОПТИКИ март–апрель 2025 Том 25 № 2 http://ntv.ifmo.ru/ SCIENTIFIC AND TECHNICAL JOURNAL OF INFORMATION TECHNOLOGIES, MECHANICS AND OPTICS March–April 2025 Vol. 25 № 2 http://ntv.ifmo.ru/en/ ISSN 2226-1494 (print) ISSN 2500-0373 (online)

ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ, МЕХАНИКИ И ОПТИКІ

doi: 10.17586/2226-1494-2025-25-2-303-310 УДК 004.932

Поиск трехмерных изображений методом сравнения контуров в задачах геологического моделирования пласта

Павел Александрович Литвинов¹, Игорь Александрович Бессмертный²

1,2 Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация

¹ p.litvinov7@gmail.com^{\Box}, https://orcid.org/0009-0009-5246-3713

² bessmertny@itmo.ru, https://orcid.org/0000-0001-6711-6399

Аннотация

Введение. Исследованы методы сравнения трехмерных изображений в задачах геологического моделирования пласта с целью повышения их качества. Предлагаемый в данной работе метод сочетает в себе такие преимущества, как глобальное представление формы, инвариантность к трансформациям, устойчивость к шумам и вычислительную эффективность. Разработан и обоснован подход, основанный на применении моментов изображений для анализа геологических данных в задачах геологического моделирования пласта. Метод. Решение задачи сравнения трехмерных изображений достигается применением математического аппарата алгебраических инвариантов. Сущность предлагаемого подхода состоит в вычислении моментов трехмерных изображений для сравнения инвариантов контуров эталона и образца. Результатом сравнения является количественная метрика соответствия сравниваемого контура искомому эталону. Основные результаты. Разработаны программные средства, встроенные в общий конвейер моделирования и анализа пакета Gempy. Метод показал удовлетворительные результаты работы на тестовой геологической модели. Точность распознавания позволяет использовать разработанные средства как рекомендательную систему. Подтверждена возможность применения предлагаемого метода для поиска заданного объекта в геологической модели и ограниченная применимость для верификации упрощенной модели в ходе итерационного расчета. Обсуждение. Выполнено сравнение разработанного метода с метрикой Хаусдорфа, методом сравнения поперечных срезов, методом сравнения ключевых точек с помощью алгоритмов SIFT и SURF, а также методом интерполяции на сетке. Показано, что для представленного метода возможно расширение области применения на более сложных геологических образованиях для работы с неоднородными структурами. Разработанные средства можно интегрировать с системами геологического моделирования, системами управления базами данных и аналитическими платформами.

Ключевые слова

поиск трехмерных изображений, сравнение контуров, моменты изображений, алгебраические инварианты, геологическое моделирование

Благодарности

Авторы выражают благодарность Гололобовой Нелли Игоревне за консультирование в области геологического моделирования.

Ссылка для цитирования: Литвинов П.А., Бессмертный И.А. Поиск трехмерных изображений методом сравнения контуров в задачах геологического моделирования пласта // Научно-технический вестник информационных технологий, механики и оптики. 2025. Т. 25, № 2. С. 303–310. doi: 10.17586/2226-1494-2025-25-2-303-310

[©] Литвинов П.А., Бессмертный И.А., 2025

Search for three-dimensional images using the contour comparison method in problems of geological reservoir modeling

Pavel A. Litvinov^{1⊠}, Igor A. Bessmertny²

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

¹ p.litvinov7@gmail.com^{\Box}, https://orcid.org/0009-0009-5246-3713

² bessmertny@itmo.ru, https://orcid.org/0000-0001-6711-6399

Abstract

Methods for comparing three-dimensional images in problems of geological modeling of a reservoir are studied in order to improve their quality. The proposed method combines such advantages as global representation of shape, invariance to transformations, noise resistance, and computational efficiency. An approach based on the use of image moments for analyzing geological data in problems of geological modeling of a reservoir is developed and substantiated. The problem of comparing three-dimensional images is solved using the mathematical apparatus of algebraic invariants. The essence of the proposed approach is to calculate the moments of three-dimensional images for comparing the invariants of the contours of the standard and sample. The result of the comparison is a quantitative metric of the conformity of the compared contour to the desired standard. Developed software tools were built into the overall modeling and analysis pipeline of the Gempy package. The method showed satisfactory results on the test geological model. The recognition accuracy allows using the developed tools as a recommender system. The possibility of using the proposed method to search for a given object in a geological model and limited applicability for verifying a simplified model during iterative calculations are confirmed. The proposed method is compared with the Hausdorff metric, the cross-section comparison method, the SIFT and SURF methods, and the grid interpolation method. It is shown that the proposed method can be expanded to more complex geological formations for working with heterogeneous structures. The developed tools can be integrated with geological modeling systems, database management systems, and analytical platforms.

Keywords

three-dimensional image search, contour comparison, image moments, algebraic invariants, geological modeling

Acknowledgements

The authors express their acknowledgements to Nelly Igorevna Gololobova for consulting in the field of geological modeling.

For citation: Litvinov P.A., Bessmertny I.A. Search for three-dimensional images using the contour comparison method in problems of geological reservoir modeling. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 2025, vol. 25, no. 2, pp. 303–310 (in Russian). doi: 10.17586/2226-1494-2025-25-2-303-310

Введение

Современный подход в создании цифрового двойника актива (месторождения или группы месторождений) предполагает разработку ряда моделей, одной из которых является геологическая модель. Данная модель может быть представлена в виде трехмерного изображения, описывающего месторождение, и сформирована на основе данных сейсморазведки и телеметрии. Геологическая модель представляет собой математическое отражение пласта и происходящие в нем физические процессы; при этом расчет модели требует значительных вычислительных мощностей. Геологическое моделирование применяют в первую очередь с целью обоснования и корректировки технических мероприятий среднесрочной и долгосрочной разработок. Такое моделирование используется на «зрелых» активах, с целью оптимизации методов их разработки. Модель строится приближенно, и от точности моделирования зависит качество прогноза, поскольку полученное в ходе моделирования трехмерное изображение пласта используется впоследствии в геофизическом анализе. В ходе анализа возникают задачи сравнения элементов модели с эталонами (например, при анализе построенной геологической модели, ее верификации), которые можно свести к автоматизированному поиску заданного трехмерного изображения эталонного объекта среди трехмерных изображений объектов модели.

При решении задачи сравнения элементов модели с эталонами работа ведется с оцифрованным физическим

объектом, трехмерное изображение которого не является нормированным и упорядоченным, поскольку было создано на основе данных геофизического исследования скважин и сейсморазведки. Соответственно, для успешного распознавания по образцу искомого объекта в полученном трехмерном изображении необходимо, чтобы это сравнение отвечало требованиям инвариантности к масштабу, ориентации и сдвигу. Для сравнения изображений предлагается анализировать их контуры. Целью исследования является повышение отдачи месторождений за счет увеличения качества моделей геологических пластов на основе математического аппарата алгебраических инвариантов для сопоставления трехмерных изображений. В настоящей работе выполнено сравнение предлагаемого метода с другими методами, применяемыми для решаемой задачи: метрика Хаусдорфа, метод сравнения поперечных срезов, метод интерполяции на сетке, а также методами сравнения ключевых точек Scale-invariant feature transform (SIFT) и Speeded-Up Robust Features (SURF)).

Наиболее перспективными для применения математического аппарата сравнения трехмерных изображений возможны две задачи, возникающие в ходе геологического моделирования: поиск заданного объекта в уже посчитанной модели и верификация упрощенных моделей в ходе их итерационного расчета. В первом случае примером использования может быть поиск «карманов» заданной структуры в посчитанной модели пласта. Прогнозирование строения пласта состоит из ряда операций и результатов измерений, которые анализирует геолог. Поскольку использование результатов единственного источника (например, интерпретация данных сейсморазведки) недостаточно для корректного построения модели [1], то проведение сравнения объектов модели с эталонами, и комбинация результатов сравнений, может выступить дополнительной проверкой, которая подтвердит достоверность и точность геологического моделирования. Второй задачей в рамках геологического моделирования, где предлагаемый подход может быть применим, является верификация создаваемой упрощенной трехмерной модели [2]. Такие модели создаются для проведения быстрых оценок дебита и его прогноза или проведения исследовательских работ. При таком упрощенном моделировании оценка корректности построения модели может быть реализована путем сравнения одного из смоделированных объектов модели с существующим физическим объектом. Тогда искомый физический объект выступит в роли эталона, с которым необходимо будет сравнивать объекты модели. Такой подход позволяет итерационно увеличивать точность моделирования, постепенно доводя ее до минимально требуемой и достаточной. Рассмотренные две задачи предлагается решить путем сравнения контуров трехмерных изображений, применив математический аппарат моментов изображений.

Представленный метод отличается от аналогов тем, что оперирует не расстояниями между точками изображения и эталона, а метриками, описывающими форму, что делает его менее чувствительным к шумам и имеющим меньшую вычислительную сложность O(n). Разработанный метод может быть использован для автоматизации процессов геологического анализа. Предложен метод для быстрого поиска объектов интереса в больших моделях, что актуально для задач разработки месторождений. Сформулирован метод оценки точности и достоверности упрощенных моделей, что снижает риски ошибок на этапах планирования проектирования ввода новых мощностей и добычи.

Использование моментов изображений для решения обозначенной задачи

Одним из способов сравнения изображений, с помощью анализа контуров эталона и идентифицируемого объекта, является сравнение их абсолютных инвариантов. Математическая база предлагаемого метода позволяет обеспечить требуемые условия инвариантности при сравнении контуров трехмерных изображений. На вход метода подаются черно-белые контуры изображений с толщиной в один пиксел (или воксель). В результате сравнения вычисляется количественная метрика соответствия того, насколько сравниваемый контур изображения соответствует искомому эталону. Предлагаемый метод на выходе формирует количественную метрику соответствия одного контура другому. В рамках данного исследования порог метрики был принят 0,87. Метод работает на базе алгебраических инвариантов, а искомые моменты изображений определяются через интеграл на основе функции плотности. Функцию плотности можно выразить через интенсивность пиксела или вокселя: поскольку обрабатываемое

изображение является бинарным, где контуры выделены одним цветом, а остальное изображение другим цветом, то интегрирование сводится к суммированию характеристик всех пикселов контура. Отметим, что идентификация контуров в трехмерных изображениях представляет отдельную задачу и выходит за рамки настоящей работы.

Для контуров двумерных изображений существуют моменты [3], сравнение которых решает поставленную задачу на плоскости. В русскоязычных научных работах за ними закрепилось название Ни-моменты. В работе [3] описан математический аппарат алгебраических инвариантов, который позволяет обеспечить требуемое сравнение, но только в двумерном пространстве. После публикации работы [3] автором был проведен ряд исследований, где расширялась область применения разработанного метода, увеличивался перечень задач обработки изображений, в котором Ни-моменты и их производные применимы. Например, кроме сравнения контуров, были разработаны подходы для анализа оригинального изображения и его копии с уменьшенным разрешением, с целью реконструкции оригинала [4]. Также были подробно исследованы подходы к наведению двумерного изображения на трехмерное, в том числе в [5, 6], с целью последующей идентификации и обработки, когда наведенное двумерное изображение совпадало со срезом исследуемого трехмерного. Активно изучались и разрабатывались подходы к снижению вычислительной сложности моментов изображения, например для моментов, инвариантных только к ориентации распознаваемого объекта [7]. В результате моменты изображений применимы как часть в больших конвейерах обработки изображений [8].

В работе [9] была развита теория, описанная в [3] до трехмерного пространства. Математическая база метода была расширена, она также основана на алгебраических инвариантах, но уже для трехмерного пространства. Получившийся в ходе разработки метод также инвариантен к масштабу, ориентации и сдвигу. Это, в свою очередь, дало старт исследованиям по расширению обозначенной в работе [9] области применения. В [10] предложен подход для распознавания лиц, позволяющий идентифицировать лица без привязки к позе и освещению, а в [11] выполнено сравнение моментов трехмерных изображений для решения задачи сравнения формы участков связывания белков. В работе [12] разработан метод, который получил применение в компьютерной графике для работы с геометрическими примитивами, используя который можно искать в меше (сетке полигонов) заданные примитивы. Как и в случае с двумерным пространством также были исследования по обеспечению инвариантности моментов к ориентации, но уже с учетом особенностей трех измерений [13]. Отдельно отметим, что в работе [14] был разработан математический аппарат для моментов в *п*-мерном пространстве.

Предлагаемый метод решения задач

Учитывая то, что в настоящей работе поиск контуров изображений не выполнялся, то отправной точкой будут служить трехмерные черно-белые изображения с идентифицированными контурами толщиной в один воксель. Сравнение контуров трехмерных изображений выполнено при анализе их инвариантов. Авторы работы [9] основывались на идее в [3] расширили теоретическую базу до трехмерного пространства, развивая оригинальный метод. Для решения обозначенной в рамках настоящего исследования задачи введем обозначение момента трехмерного изображения. Момент изображения m_{pqr} порядка (p, q, r) функции плотности $\rho(x, y, z)$ существует и определяется через интеграл Римана как

$$m_{(pqr)} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x^p y^q z^r \rho(x, y, z) dx \, dy \, dz,$$

где p, q, r = 0, 1, 2, ..., n; функция плотности $\rho(x, y, z)$ является кусочно-непрерывной, а значит, ограниченной, а также имеет ненулевые значения в пространстве \mathbb{R}^3 только в конечной части пространства (x, y, z).

Как и в случае с двумерным пространством и моментами m_{pq} порядка (p, q) для моментов трехмерного изображения вводится определение центральных моментов, инвариантных к масштабу

$$\mu_{(pqr)} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - \overline{x})^p (y - \overline{y})^q (z - \overline{z})^{r \times} \\ \times \rho(x, y, z) d(x - \overline{x}) d(y - \overline{y}) d(z - \overline{z}),$$
(1)

где $\overline{x} = \frac{m_{100}}{m_{000}}, \, \overline{y} = \frac{m_{010}}{m_{000}} \,_{\mathrm{H}} \, \overline{z} = \frac{m_{001}}{m_{000}}.$

В работе [9] приведены три абсолютных инварианта — инварианты к масштабу, ориентации и сдвигу.

$$I_1 = \mu_{200} + \mu_{020} + \mu_{002}, \tag{2}$$

$$I_2 = \mu_{020}\mu_{002} - \mu_{011}^2 + \mu_{200}\mu_{002} - \mu_{101}^2 + \mu_{200}\mu_{020} - \mu_{110}^2,$$
(3)

$$I_{3} = \mu_{200}\mu_{020}\mu_{002} - \mu_{200}\mu_{011}^{2} - \mu_{002}\mu_{110}^{2} + 2\mu_{110}\mu_{011}\mu_{101} - \mu_{020}\mu_{101}^{2}.$$
 (4)

Также в [9] выведены выражения для сравнения контуров на основе абсолютных инвариантов (2), (3), (4): $\frac{I_1^2}{I_2}$ и $\frac{I_3}{I_1^3}$. Их вычисление дает количественную ха-

рактеристику соответствия эталона трехмерного изображения и искомого объекта.

Отметим, что для дискретного изображения можно выразить центральные моменты трехмерного изображения (1) через сумму ряда:

$$\mu_{(pqr)} = \sum_{X} \sum_{Y} \sum_{Z} (x - \overline{x})^p (y - \overline{y})^q (z - \overline{z})^r \rho(x, y, z),$$

где *X*, *Y*, *Z* — области, где определены точки трехмерного изображения. Это применимо в том числе и к формуле (1) для трехмерного изображения.

Помимо рассмотрения метода сравнения контуров, стоит дать оценку применимости другим методам, используемым для решения обозначенной задачи. Рассмотрим четыре метода, которые по мнению авторов настоящей работы наиболее часто применяются.

Метрика Хаусдорфа. Метод основан на измерении максимального расстояния между ближайшими точками двух объектов. Расчет расстояния служит оценкой сходства или расхождения двух объектов. Вследствие ее чувствительности к артефактам и шумам удобно находить локальные различия между объектами. Моменты же, напротив, описывают форму, и менее чувствительны к локальным искажениям. Для применения данного метода в настоящей работе потребуются итерационные вычисления расстояния между множеством объектов, в то время как сравнение моментов — это анализ числовых характеристик.

Метод интерполяции на сетке. Метод предполагает интерполяцию двух поверхностей на одинаковые сетки. Затем, на основе значений в узлах сеток, строятся численные описания поверхности, вычисляются разности (например, среднеквадратичные) для каждой пары узлов. Необходимо учитывать, что интерполяция сглаживает мелкие детали и границы, в то время как применяемые моменты напрямую работают с геометрией объектов. По аналогии с метрикой Хаусдорфа, сравнение поверхностей методом интерполяции на сетке итерационное, и требует больших вычислительных ресурсов, чем сравнение числовых характеристик.

Метод сравнения поперечных срезов. Метод предполагает последовательное исследование формы объектов посредством анализа срезов. Фактически, каждый срез представляет собой отдельный двумерный контур, который сопоставляется с двумерного контуром, полученным из эталона. Существуют разные подходы в сравнении двумерных срезов: от сравнения Ни-моментов до сравнения таких метрик как площадь, длина и т. п. В любом случае результаты анализа зависят от изначальной ориентации срезов и их дискретизации. При сравнении трехмерных изображений такой зависимости нет. Также, сравнение моментов позволяет учесть топологию изображения целиком, а не разбивать его на отдельные несвязные двумерные срезы. С примерами использования подобных методов можно ознакомиться в материалах [5, 6].

Метод сравнения ключевых точек, где наибольшую известность имеют алгоритмы SIFT [15] и SURF [16]. Суть метода заключается в том, чтобы найти ключевые точки эталонного и полученного изображений и сравнить их. Алгоритмы SIFT и SURF обеспечивают инвариантность к масштабированию, ориентации и сдвигу. Они также дают незначительную устойчивость к искажению, но в настоящей работе задача найти не точно такой же объект, а только похожий. Например, если требуется найти прямоугольник, и используется эталон с углами в 90°, а образец со скругленными углами, то алгоритмы построят разные наборы ключевых точек, и, как следствие, получили бы ложноотрицательные результаты сравнения. Это связано с тем, что при наличии мелких деталей алгоритмы вычисляют дескрипторы ключевых точек с ошибками из-за зашумленности. Сравнение рассматриваемых методов представлено в таблице.

Метод	Инвариантность	Вычислительная сложность	Устойчивость к шуму
Метрика Хаусдорфа	Нет	$O(n^2)$	Низкая
Метод сравнения поперечных срезов	Частичная	O(n)	Средняя
Метод сравнения ключевых точек (SIFT и SURF)	Дa	$O(n \log n)$	Низкая
Метод интерполяции на сетке	Нет	O(n)	Высокая
Предлагаемый метод	Дa	O(n)	Высокая

Таблица. Сравнительная таблица рассматриваемых методов *Table*. Comparative table of the methods under consideration

Разработанный метод сочетает в себе ключевые преимущества: глобальное представление формы, инвариантность к трансформациям, устойчивость к шумам и вычислительную эффективность. На основе предложенного метода разработана методика сравнения объектов в геологическом моделировании, превосходящая как локальные, так и сеточные подходы по отдельным показателям

Результаты

В настоящей работе для геологического моделирования использовался пакет с открытым кодом Gempy, реализованный на языке программирования Python. Пакет позволяет моделировать сложные трехмерные геологические модели, с такими включениями как разломы, складчатые структуры, сети разломов, магматические тела, складки и прочее. Благодаря тому, что программное обеспечение распространяется по модели open source, можно легко встроить собственный модуль в конвейер расчета. Также пакет содержит в себе встроенные инструменты визуализации получившихся моделей. В качестве опытной модели используется готовая публичная модель. Она обладает достаточным количеством слоев для наглядности работы метода в рамках исследования, а также позволяет проверить корректность работы метода при сравнении криволинейных трехмерных объектов. Визуализация опытной модели приведена на рис. 1. Модель содержит три слоя и один нефтяной карман. Искомый карман находится между слоями Formation 1 (Порода 1) и Formation 3 (Порода 3) (рис. 1). Значения координатных осей соответствуют координатам вокселей в границах моделируемого экстента. Стоит отметить, что исходная модель не содержит достаточно данных для определения топологии пласта встроенными средствами Gempy.

Схематичное изображение исследуемого пласта в проекции оси ординат приведено на рис. 2.







Puc. 2. Схематичное изображение структуры пласта *Fig. 2.* Schematic representation of the formation structure

Приведем два сценария работы предлагаемого метода.

Сценарий 1. Выполним поиск заданного объекта в уже посчитанной модели. Примем за искомый объект нефтяной карман. В ходе исследования итерационно сравнивались все контуры модели с контуром эталона. Для сравнения контуров было вычислено значение выражения $\frac{I_1^2}{I_2}$ для контура эталона и 156 контуров в трехмерном изображении, где I_1 и I_2 описаны в уравнениях (2) и (3). Значение для контура эталона составило 3,476. В силу специфики решаемой задачи вместо метрики среднеквадратичного отклонения была выбрана метрика точности. Порог точности был задан в 87 %. Результаты работы созданных средств поиска контуров: ложноположительный результат — 2; истинно положительный результат — 1; ложноотрицательный результат — 153.

Применяемая методика показала удовлетворительные результаты работы, однако на один искомый контур пришлось два ложноположительных результата. Точность распознавания позволяет использовать разработанные на основе методики средства как рекомендательную систему, поскольку позволяет устранить ложноотрицательные результаты, однако она недостаточна для автоматизированной чистовой разметки имеющихся моделей. Методика также применима как вспомогательная система черновой разметки, которая впоследствии верифицируется геологом. При этом необходима база искомых эталонов, используемых для автоматизированного сравнения.

Сценарий 2. Использован метод верификации упрощенной модели в ходе ее итерационного расчета. Примем за искомый контур тот же, что и в рамках сценария 1. Использование пакета Gempy позволяет встроить в конвейер расчета модели дополнительные этапы. Критерием останова является сходимость расчетов: достигая нужного уровня сходимости при заданной точности моделирование считается завершенным, а модель рассчитанной. В ходе расчета должны сходиться сразу же множество параметров, это обуславливает необходимость учета всей сцены целиком для каждой итерации. В настоящей работе сравнение трехмерных изображений осуществлено на каждой итерации расчета. Первоначально предполагалось использовать существующий эталон для сравнения. Однако на начальных итерациях расчета геометрическая форма сравниваемых объектов значительно была упрощена, и провести сравнение качественно не удалось. Для решения этой проблемы была создана база эталонов с постепенным усложнением меша, конечным элементом которой был исходный эталон.

Изначальный расчет модели требовал 5 итераций расчета с разделением на 6, 6, 42, 9 и 6 частей (chunks). На вычислительном узле с характеристиками СРU 1,1 GHz Intel Core i5, GPU Intel Iris Plus Graphics 1536 МБ и RAM 16 ГБ 3733 MHz LPDDR4X расчет занимал 1 мин 34 с. После модификации конвейера Gempy, с усложняющейся поэтапно проверкой сходимости, требуемая сходимость получена на четвертой итерации, а расчет длился на 1100 мс меньше, что составляет 1,1 % от общей продолжительности расчета. При этом временные затраты для подготовки базы эталонов, очевидно, несопоставимы, и предлагаемый метод неэффективен в случае разовых вычислений. Можно констатировать, что в сценарии 2 разработанный метод применим в том случае, если модель необходимо постоянно пересчитывать, и затраты на пересчет превышают затраты на создание базы упрощенных эталонов.

Заключение

Метод сравнения контуров применим для реализации поиска трехмерных изображений в задачах геологического моделирования пласта. Предложенный метод позволяет решать задачи оценки корректности построения упрощенной модели и достоверности построенной модели. Метод применим в общем конвейере моделирования и анализа, имеет высокую устойчивость к шуму, а также инвариантен к ориентации, масштабированию и сдвигу. При этом предложенный метод только дополняет, а не заменяет полный расчет модели из-за наличия преимуществ и ограничений.

К преимуществам данного метода можно отнести выявление изменений в моделируемом объекте, низкую ресурсоемкость и интуитивное представление. Сравнение контуров применимо для выявления изменений и аномалий в геологической структуре, такое применение упрощает первичные автоматизированные диагностику и анализ. Сравнение контуров также менее ресурсоемко по сравнению с более комплексными методами анализа геологических моделей. Предлагаемый метод имеет вычислительную сложность O(n). Сами контуры представляют собой понятное и наглядное представление формы и структуры пласта, что облегчает первичную интерпретацию данных.

Стоит отметить, что помимо выделенных преимуществ данный метод обладает явными ограничениями, которые необходимо учитывать при использовании, а именно потерю информации, зависимость от качества данных и отсутствие подходов в работе с многозначностью контуров. В предлагаемом методе не используется информация о внутренней структуре пласта, полученная в ходе сейсморазведки, контуры

Литература

- Билибин С.И., Дьяконова Т.Ф., Исакова Т.Г., Истомин С.Б., Юканова Е.А. Трехмерная геологическая модель — необходимый и обязательный этап изучения нефтегазового месторождения // Недропользование XXI век. 2007. № 4. С. 38–42.
- Григорьев Ю.М., Харбанов М.В. Разработка модуля по созданию упрощенной структурной 3D модели нефтяного пласта // Математические заметки ЯГУ. 2013. Т. 20. № 2. С. 246–255.
- Hu M.-K. Visual pattern recognition by moment invariants // IRE Transactions on Information Theory. 1962. V. 8. N 2. P. 179–187. https://doi.org/10.1109/Tit.1962.1057692
- Boyce J.F., Hossack W.J. Moment invariants for pattern recognition // Pattern Recognition Letters. 1983.V. 1. N 5-6. P. 451–456. https://doi. org/10.1016/0167-8655(83)90085-5
- Нгуен З.Т., Хачумов М.В. Метод наведения 3D-модели объекта на 2D-изображение на основе инвариантных моментов // Программные системы: теория и приложения. 2017. Т. 8. № 4(35). С. 209–220.
- Mukundan R., Ramakrishnan K.R. An iterative solution for object pose parameters using image moments // Pattern Recognition Letters. 1996. V. 17. N 12. P. 1279–1284. https://doi.org/10.1016/0167-8655(96)00099-2
- Flusser J. On the independence of rotation moment invariants // Pattern Recognition. 2000. V. 33. N 9. P. 1405–1410. https://doi. org/10.1016/S0031-3203(99)00127-2
- Zitova B., Flusser J. Image registration methods: a survey // Image and Vision Computing. 2003. V. 21. N 11. P. 977–1000. https://doi. org/10.1016/S0262-8856(03)00137-9
- Sadjadi F.A., Hall E.L. Three-dimensional moment invariants // IEEE Transactions on Pattern Analysis and Machine Intelligence. 1980. V. 2. N 2. P. 127–136. https://doi.org/10.1109/TPAMI.1980.4766990
- Thakare N.M., Thakare V.M. A Supervised Hybrid Methodology for Pose and Illumination Invariant 3D Face Recognition // International Journal of Computer Applications. 2012. V. 47. N 25. P. 24–29. https://doi.org/10.5120/7537-474
- Sommer I., Muller O., Domingues F.S., Sander O., Weickert J., Lengauer T. Moment invariants as shape recognition technique for comparing protein binding sites // Bioinformatics. 2007. V. 23. N 23. P. 3139–3146. https://doi.org/10.1093/bioinformatics/btm503
- Xu D., Li H. 3-D affine moment invariants generated by geometric primitives // Proc. of the 18th International Conference on Pattern

представляют собой упрощенную проекцию данных. Точность сравнения контуров сильно зависит от качества исходных данных и метода построения контуров. Некачественные данные с большим количеством помех приводят к неверной идентификации контуров, а значит и неудовлетворительным результатами работы метода. В некоторых случаях, один и тот же контур может представлять разные геологические объекты, и принадлежность к конкретному объекту можно определить только в контексте, что усложняет интерпретацию.

Предложенный метод и разработанные средства могут быть усовершенствованы путем расширения области применения и обеспечения совместимости с другими системами. Возможно расширение области применения метода на более сложные геологические образования для работы с неоднородными геологическими структурами. К таким структурам можно отнести фации, разломы и тектонические структуры. Также, возможно внедрение методов для отслеживания изменений геологических образований с течением времени. Разработанные средства можно интегрировать с системами геологического моделирования, системами управления базами данных и аналитическими платформами.

References

- Bilibin S.I., Diakonova T.F., Isakova T.G., Istomin S.B., Iukanova E.A. A three-dimensional geological model is a necessary and mandatory stage in studying an oil and gas field. *Nedra 21*, 2007, no. 4, pp. 38–42. (in Russian)
- Grigorev Yu.M., Kharbanov M.V. The development of a program module for creating simplified structural 3d oil reservoirs model. *Mathematical notes of Yakutsk State University*, 2013, vol. 20, no. 2, pp. 246–255. (in Russian)
- Hu M.-K. Visual pattern recognition by moment invariants. *IRE Transactions on Information Theory*. 1962. vol. 8, no. 2, pp. 179–187. https://doi.org/10.1109/Tit.1962.1057692
- 4. Boyce J.F., Hossack W.J. Moment invariants for pattern recognition. *Pattern Recognition Letters*, 1983, vol. 1, no. 5-6, pp. 451–456. https://doi.org/10.1016/0167-8655(83)90085-5
- Nguyen D.T., Khacumov M.V. The method of comparing a 3d object model with a 2d image based on invariant moments. *Program Systems: Theory and Applications*, 2017, vol. 8, no. 4(35), pp. 209– 220. (in Russian)
- Mukundan R., Ramakrishnan K.R. An iterative solution for object pose parameters using image moments. *Pattern Recognition Letters*, 1996, vol. 17, no. 12, pp. 1279–1284. https://doi.org/10.1016/0167-8655(96)00099-2
- Flusser J. On the independence of rotation moment invariants. *Pattern Recognition*, 2000, vol. 33, no. 9, pp. 1405–1410. https://doi. org/10.1016/S0031-3203(99)00127-2
- Zitova B., Flusser J. Image registration methods: a survey. *Image and Vision Computing*, 2003, vol. 21, no. 11, pp. 977–1000. https://doi.org/10.1016/S0262-8856(03)00137-9
- Sadjadi F.A., Hall E.L. Three-dimensional moment invariants. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1980, vol. 2, no. 2, pp. 127–136. https://doi.org/10.1109/TPAMI.1980.4766990
- Thakare N.M., Thakare V.M. A Supervised Hybrid Methodology for Pose and Illumination Invariant 3D Face Recognition. *International Journal of Computer Applications*, 2012, vol. 47, no. 25, pp. 24–29. https://doi.org/10.5120/7537-474
- Sommer I., Muller O., Domingues F.S., Sander O., Weickert J., Lengauer T. Moment invariants as shape recognition technique for comparing protein binding sites. *Bioinformatics*, 2007, vol. 23, no. 23, pp. 3139–3146. https://doi.org/10.1093/bioinformatics/btm503

Recognition (ICPR'06). 2006. P. 544–547. https://doi.org/10.1109/ icpr.2006.21

- Suk T., Flusser J., Boldyš J. 3D rotation invariants by complex moments // Pattern Recognition. 2015. V. 48. N 11. P. 516–3526. https://doi.org/10.1016/j.patcog.2015.05.007
- Mamistvalov A.G. n-Dimensional moment invariants and conceptual mathematical Theory of recognition n-Dimensional solids // IEEE Transactions on Pattern Analysis and Machine Intelligence. 1998. V. 20. N 8. P. 819–831. https://doi.org/10.1109/34.709598
- Lowe D.G. Distinctive image features from scale-invariant keypoints // International Journal of Computer Vision. 2004. V. 60. N 2. P. 91–110. https://doi.org/10.1023/b:visi.0000029664.99615.94
- Bay H., Tuytelaars T., Van Gool L. SURF: Speeded up robust features // Lecture Notes in Computer Science. 2006. V. 3951. P. 404– 417. https://doi.org/10.1007/11744023_32

Авторы

Литвинов Павел Александрович — аспирант, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, https://orcid. org/0009-0009-5246-3713, p.litvinov7@gmail.com

Бессмертный Игорь Александрович — доктор технических наук, профессор, профессор, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, sc 36661767800, https://orcid.org/0000-0001-6711-6399, bessmertny@itmo.ru

Статья поступила в редакцию 30.09.2024 Одобрена после рецензирования 26.02.2025 Принята к печати 26.03.2025 Received 30.09.2024 Approved after reviewing 26.02.2025 Accepted 26.03.2025



Работа доступна по лицензии Creative Commons «Attribution-NonCommercial»

- Xu D., Li H. 3-D affine moment invariants generated by geometric primitives. Proc. of the 18th International Conference on Pattern Recognition (ICPR'06), 2006, pp. 544–547. https://doi.org/10.1109/ icpr.2006.21
- Suk T., Flusser J., Boldyš J. 3D rotation invariants by complex moments. *Pattern Recognition*, 2015, vol. 48, no. 11, pp. 516–3526. https://doi.org/10.1016/j.patcog.2015.05.007
- Mamistvalov A.G. n-Dimensional moment invariants and conceptual mathematical Theory of recognition n-Dimensional solids. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1998, vol. 20, no. 8, pp. 819–831. https://doi.org/10.1109/34.709598
- Lowe D.G. Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision, 2004, vol. 60, no. 2, pp. 91–110. https://doi.org/10.1023/b:visi.0000029664.99615.94
- Bay H., Tuytelaars T., Van Gool L. SURF: Speeded up robust features. Lecture Notes in Computer Science, 2006, vol. 3951, pp. 404–417. https://doi.org/10.1007/11744023_32

Authors

Pavel A. Litvinov — PhD Student, ITMO University, Saint Petersburg, 197101, Russian Federation, https://orcid.org/0009-0009-5246-3713, p.litvinov7@gmail.com

Igor A. Bessmertny — D.Sc., Full Professor, ITMO University, Saint Petersburg, 197101, Russian Federation, SC 36661767800, https://orcid. org/0000-0001-6711-6399, bessmertny@itmo.ru

νίτμο

НАУЧНО-ТЕХНИЧЕСКИЙ ВЕСТНИК ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ, МЕХАНИКИ И ОПТИКИ март–апрель 2025 Том 25 № 2 http://ntv.ifmo.ru/ SCIENTIFIC AND TECHNICAL JOURNAL OF INFORMATION TECHNOLOGIES, MECHANICS AND OPTICS March–April 2025 Vol. 25 № 2 http://ntv.ifmo.ru/en/ ISSN 2226-1494 (print) ISSN 2500-0373 (online)

ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ, МЕХАНИКИ И ОПТИКІ

doi: 10.17586/2226-1494-2025-25-2-311-320

Integration of enhanced qualified electronic signature to the fast healthcare interoperability resources (FHIR RU-core) protocol Nikolai A. Gorbunov^{1⊠}, Valeria O. Kuleshova², Viktoriia M. Korzhuk³

¹ "Netrika Medicine" LLC, Saint Petersburg, 191015, Russian Federation

^{1,3} ITMO University, Saint Petersburg, 197101, Russian Federation

² CSRI Elektropribor, Saint Petersburg, 197046, Russian Federation

¹ gorb-2157@mail.ru^{\Box}, https://orcid.org/0009-0004-2973-9594

² valeria.kuleshova@yahoo.com, https://orcid.org/0000-0003-1377-6003

³ vmkorzhuk@itmo.ru, https://orcid.org/0000-0002-0240-9067

Abstract

The article discusses how to use the Russian profile of the Fast Healthcare Interoperability Resources (FHIR) RU-core protocol for medical information systems developing. An enhanced qualified electronic signature has been used for information protection for a long time; however, it is currently being implemented for the first time with the FHIR RU-core protocol to protect medical information systems. The goal of the research is enhanced qualified electronic signature integration for organizations developing secure software for medical information systems. To reach the goal, the following tasks are solved: previous works including foreign ones are analyzed and the table with different variants of FHIR protocol using is presented; the step-by-step plan of an enhanced qualified electronic signature integration has elaborated. A software code has been created to ensure the safe transmission of sensitive medical data to meet the challenge of implementing an enhanced qualified electronic signature. Russian standards were used to implement cryptographic protection of information in various medical information systems. To ensure secure data exchange, an enhanced qualified electronic signature was incorporated into the domestic version of the FHIR protocol. The use of Russian version of the protocol and certificates result in the correct exchange of medical documents. New functionality for medical information systems was standardized through the application of the Russian profile of FHIR protocol. Medical information systems deployed in the certified data processing centers are now using the FHIR RU-core protocol. The medical community easily uses FHIR RU-core, which is the most advanced tool for domestic medical systems. The method is aimed at integrating health information systems safely to develop regional services for doctors, patients, and digital health care organizers. The scientific novelty and relevance of the research lies in the field of adaptation international experience of using FHIR protocol under Russian circumstances and refinement of an enhanced qualified electronic signature integration method without capacity loss. The practical result demonstrates that the use of the Russian enhanced qualified electronic signature satisfies the information security requirements of new medical information systems and allows sensitive data to be transmitted without loss of quality and speed. It has been concluded that a systematic approach to using the Russian profile of the FHIR RU-core protocol for new medical information systems, with the aim of implementing digital healthcare, is highly recommended. This article is a valuable resource for medical information systems software architects and developers, as well as information security specialists.

Keywords

medical information systems, Russian FHIR protocol, electronic signature integration

Acknowledgements

Funding from the research project "Autonomous Intelligent Systems" No. 623106 at ITMO University.

For citation: Gorbunov N.A., Kuleshova V.O., Korzhuk V.M. Integration of enhanced qualified electronic signature to the fast healthcare interoperability resources (FHIR RU-core) protocol. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 2025, vol. 25, no. 2, pp. 311–320. doi: 10.17586/2226-1494-2025-25-2-311-320

© Gorbunov N.A., Kuleshova V.O., Korzhuk V.M., 2025

УДК 004.056

Интеграция усиленной квалифицированной электронной подписи с ресурсами быстрого взаимодействия в сфере здравоохранения FHIR RU-core

Николай Александрович Горбунов^{1⊠}, Валерия Олеговна Кулешова², Виктория Михайловна Коржук³

¹ ООО «Нетрика Медицина», Санкт-Петербург, 191015, Российская Федерация

1,3 Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация

² АО «Концерн «ЦНИИ «Электроприбор», Санкт-Петербург, 197046, Российская Федерация

¹ gorb-2157@mail.ru^{\equiv}, https://orcid.org/0009-0004-2973-9594

² valeria.kuleshova@yahoo.com, https://orcid.org/0000-0003-1377-6003

³ vmkorzhuk@itmo.ru, https://orcid.org/0000-0002-0240-9067

Аннотация

Введение. Рассматривается возможность использования российского профиля протокола Fast Healthcare Interoperability Resources (FHIR) RU-core для разработки медицинских информационных систем. Усиленная квалифицированная электронная подпись применяется для защиты информации уже длительное время, однако, совместно с протоколом FHIR RU-соге для защиты медицинских информационных систем впервые. В работе предложен процесс внедрения усиленной квалифицированной электронной подписи для организаций, разрабатывающих безопасное программное обеспечение для медицинских информационных систем. Для достижения цели были решены следующие задачи: проанализированы русские и зарубежные научные работы, и составлена таблица возможностей использования протокола FHIR, выработан пошаговый план внедрения усиленной квалифицированной электронной подписи. Метод. Для решения задачи внедрения усиленной квалифицированной электронной подписи был разработан программный код, который позволяет передавать чувствительные медицинские данные безопасно. Проведена реализация криптографической защиты информации при помощи российских стандартов интеграции различных медицинских информационных систем. В отечественной версии протокола FHIR встроена усиленная квалифицированная электронная подпись для защищенного обмена данными. Корректный обмен медицинскими документами возможен при использовании российских версий протокола и сертификатов. Проведена стандартизация применения российского профиля протокола FHIR RU-соге для медицинских информационных систем с новым функционалом. Протокол FHIR RU-соге был внедрен в медицинские информационные системы, развернутые в аттестованных центрах обработки данных. Стандарт FHIR RU-core является наиболее инновационным инструментом для отечественных медицинских систем, так как находится в открытом доступе, и удобен для применения медицинским сообществом. Метод ориентирован на безопасную интеграцию медицинских информационных систем для развития региональных сервисов для врачей, пациентов и организаторов цифрового здравоохранения. Основные результаты. Полученные результаты заключаются в адаптации международного опыта использования протокола FHIR для российской действительности и в уточнении метода внедрения усиленной квалифицированной электронной подписи без потери пропускной способности в медицинских информационных системах. Практический результат показывает, что использование российской усиленной квалифицированной электронной подписи удовлетворяет требованиям информационной безопасности для новых медицинских информационных систем и позволяет передавать чувствительные данные без потери качества и скорости. Обсуждение. Сделан вывод о необходимости внедрения системного подхода к использованию российского профиля протокола FHIR RU-core для новых медицинских информационных систем с целью дальнейшего развития цифрового здравоохранения. Результаты работы могут быть ценным ресурсом для архитекторов и разработчиков программного обеспечения медицинских информационных систем, а также специалистов по информационной безопасности.

Ключевые слова

медицинские информационные системы, российский протокол FHIR, интеграция электронной подписи

Благодарности

Финансирование выполнено за счет средств научно-исследовательской работы «Автономные интеллектуальные системы» № 623106 в Университете ИТМО.

Ссылка для цитирования: Горбунов Н.А., Кулешова В.О., Коржук В.М. Интеграция усиленной квалифицированной электронной подписи с ресурсами быстрого взаимодействия в сфере здравоохранения FHIR RU-core // Научно-технический вестник информационных технологий, механики и оптики. 2025. Т. 25, № 2. С. 311–320 (на англ. яз.). doi: 10.17586/2226-1494-2025-25-2-311-320

Introduction

Public and private Medical Information Systems (MIS) are forced to constantly transfer large volumes of private information, namely personal data¹. The source [1] has

confirmed how to generate a list of personal data processed in MIS. When discussing the transfer of medical data, it is important to emphasize the importance of the Fast Healthcare Interoperability Resources (FHIR) protocol. In digital health, the exchange of medical information and the interoperability of healthcare resources can be achieved through the use of the standard protocol FHIR according to [2].

¹ Федеральный закон России № 152 от 27.07.2006 «О персональных данных». Federal Law of Russia no. 152 at 27.07.2006 "About personal data".

One of the well-known companies in Russia engaged in secure integration for MIS in the Information Technology (IT) sector is "Netrika Medicine". In fact, this company is a Russian IT solution provider and developer. It is involved in developing web portals that are compliant with the latest FHIR standards. The company facilitates access for doctors and medical institution workers in Russian regions to an integrated electronic medical record¹. Many new tools for practical work have been developed by doctors who work in private and public medical organizations, clinics, and diagnostic centers. MIS is where these tools are utilized to achieve new goals and issues. User web portals enable a significant increase in the speed and accuracy of disease diagnosis as well as facilitating doctor-patient interactions. Patients are no longer required to collect certificates, extracts, test data, ultrasound, etc. before receiving advisory or medical assistance due to this fact, which greatly simplifies their lives.

During the project, the specialists at "Netrika Medicine" solved multiple problems. Electronic versions of medical documents can be obtained from MIS used in healthcare institutions and provided to doctors in other institutions through a graphical interface. In addition, the system statistical indicators were expanded by significantly enhancing the databases.

When creating the interaction between web portals and the integrated electronic medical record module, the initial focus was on the development of various types of MIS. To utilize this opportunity, FHIR was selected as the most advanced standard for exchanging medical data². Health Level Seven International is working on the development of the FHIR protocol. The Table 1 below shows how the FHIR protocol has been utilized in previous scientific works.

The FHIR protocol global significance lies in its ability to enhance the interaction between diverse health systems. This allows MISs, which are developed by non-cooperative corporations, to effectively cooperate, share data, and interpret general information.

It is important to emphasize the ideological similarities of the decisions when comparing the integration of Enhanced Qualified Electronic Signature (EQES) with FHIR RU-core and existing solutions in the United States of America (USA). The algorithm GOST R 34.10-2012³ is the sole distinguishing factor of EQES which is applicable to MISs operating in Russia. MISs operating in the US employ the Digital Signature Standard algorithm in the EQES. The main similarity of EQES integration with Russian and American MISs is in the introduction of EQES into the structure of the FHIR protocol. The aim of this integration is unambiguous identification of authorship as well as confirmation of the integrity of the signed data.

The FHIR standard-based MIS types and properties vary among the sources presented in the Table 1. However, they demonstrate the consistency and necessity of utilizing a universal FHIR protocol in any MIS. After analyzing six articles on the FHIR protocol, we have determined that it is used in all MIS. Moreover, FHIR protocol universality and user-friendliness make it convenient for MIS developers. It serves as a basis for further research on the idea of utilizing its Russian version.

By using FHIR, interaction with various MIS can be completely unified. This fact allows for the storage of an integrated electronic medical record through a specific

Study title	Research feature
FHIR profile-based patient decision aid system [3]	It uses the FHIR standard for semantic compatibility with different systems
FHIR profile-based patient decision support system [4]	It uses the FHIR standard to create a decision support MIS for patients
FHIR medical data management platform [5]	It uses the FHIR standard to provide clinical decision support
FHIR profile-based model for collecting clinical study data [6]	It uses the FHIR standard to achieve interoperability of metadata and clinical trial data
FHIR profile-based pharmacogenomics clinical decision support service [7]	It uses the FHIR standard to combine pharmacogenomics clinical decision support service with clinical decision support system
FHIR profile-based practice guidelines [8]	It uses the FHIR standard to improve practice guidelines through the use of patient-specific recommendations
FHIR profile-based approach to clinical decision support system integration in critical care [9]	It describes the experiences of nurses and doctors using a FHIR profile-based clinical decision support system

Table 1. Implementation of the FHIR protocol

¹ Программные продукты ООО «Нетрика Медицина» [Электронный ресурс]. URL: https://n3med.ru/ (дата обращения: 20.05.2024). Software products of the company "Netrika Medicine", URL: https://n3med.ru (accessed: 20.05.2024).

² Рыжиков М. Введение в HL7 FHIR. EverCare. 2016 [Электронный ресурс]. URL: https://evercare.ru/vvedenie-vhl7-fhir (дата обращения: 20.05.2024). Ryzhikov M. Introduction to HL7 FHIR. EverCare. 2016. URL: https://evercare.ru/vvedeniev-hl7-fhir (accessed: 20.05.2024).

³ Национальный стандарт РФ ГОСТ Р 34.10-2012 «Информационная технология. Криптографическая защита информации. Процессы формирования и проверки электронной цифровой подписи». National standard of the Russian Federation GOST R 34.10-2012 "Information technology. Cryptographic protection of information. Electronic digital signature generation and verification processes".

web portal and also directly from any medical institution information system.

EQES meets the security requirements because it provides the highest level of protection and legal significance. This is identical to a handwritten signature. EQES is used to transmit confidential information to banks, courts, tax authorities, medical institutions, and other organizations with strict security requirements. This is confirmed by the fact that only accredited certifying centers are entitled to create EQES certificates. The format and storage of EQES keys require additional requirements. The organization EQES private key must be written to a token that is held solely by the owner. Cryptographic means are utilized to create EQES, which confirms authorship and document integrity without the need for additional agreements between the parties. It is widely accepted that the EQES is a universal tool for electronic document processing worldwide.

Synchronization using two cryptographic transformation principles can make integration compatibility with international health systems possible.

The FHIR RU-core protocol is the model used by the Russian Federation for transferring sensitive personal data in healthcare¹. The interaction of information systems in Russian healthcare requires unified understanding of data which it serves as a basis for.

Method

The article aims to develop a distinctive approach for utilizing Russian EQES in the FHIR RU-core protocol when constructing MISs to expand integration capabilities².

This goal requires the integration of cryptographic information protection tools in the Russian national standard for exchanging medical records among MISs. The integration task has arisen because the Russian version of the FHIR protocol utilizes domestic cryptographic information protection tools in a distinct manner. This scientific article was created through the use of logical inference, system analysis, search and cognition, and methodological design.

The research uniqueness lies in its method for integrating EQES into the FHIR RU-core protocol which will enable data exchange in various MISs. Transferring manipulations that were previously unused in the FHIR protocol for MIS is now possible thanks to this new method. The use of cryptographic information protection tools for transferring personal data makes it unique by adhering to the general principle. However, at the same time, we emphasized the use of the Russian tools for exchanging medical records.

The relevance of the study lies in the necessity to integrate EQES certificates into the FHIR RU-core

framework. Systematic work resulted in the standardization of indicators for the use of the Russian profile of the FHIR RU-core protocol in a new MIS created to implement new functionality.

Using the inalienability principle of the organization component and production needs, the FHIR RU-core protocol was incorporated into MIS that is currently deployed in several data processing centers certified for various information security requirements [10].

It becomes possible due to the Russian version of the tool for exchanging medical records meeting the following information security requirements:

- Government decree on the protection of personal data³;
- Order of the Federal Service for Technical and Export Control of the Russian Federation regarding the protection of information in state information systems⁴;
- Government decree on the categorization of critical information infrastructure facilities in the Russian Federation⁵.

Scientific articles on healthcare MIS confirmed the advantages of adding EQES to the FHIR RU-core protocol, indicating that MIS without EQES is not very useful [11]. The MIS subsystems are continuously being developed, which includes updating the list of information security requirements [12]. The development of various MIS resulted in the need to establish a coordinated data exchange protocol. In Russia, the FHIR RU-core standard is the most advanced and developed tool for MIS. Both, MIS software developers and representatives from the medical industry are interested in this standard. This protocol is a necessary addition to the system of medical standards. In fact, HL7 FHIR is currently the only standard with a large professional community and a profound comprehension of medical records. Its primary advantage is a resource diagram that is well-designed to describe all required entities including patients and doctors, immunizations, treatment plans, etc.

¹ Манифест FHIR Ru. FHIR RU-core [Электронный реcypc]. URL: https://fhirru.github.io/core/docs/manifest/basic. html. (дата обращения: 20.05.2024). Manifesto FHIR Ru. FHIR RU-core, URL: https://fhirru.github.io/core/docs/manifest/basic. html (accessed: 20.05.2024).

² Федеральный закон России № 63 от 06.04.2011 «Об электронной подписи». Federal Law of Russia No. 63 at 04.06.2011 "About electronic signature".

³ Постановление Правительства России от 01.11.2012 № 1119 «Об утверждении Требований к защите персональных данных при их обработке в информационных системах персональных данных». Decree of the government of Russia No. 1119 at 11.01.2012 "About approval of requirements for the protection of personal data during their processing in personal data information systems".

⁴ Приказ ФСТЭК России от 02.11.2013 № 17 «Об утверждении требований о защите информации, не составляющей государственную тайну, содержащейся в государственных информационных системах». Order of the FSTEC of Russia at 11.02.2013 No. 17 "About approval of the requirements for the protection of information that does not constitute a state secret contained in state information systems".

⁵Постановление Правительства России № 127 от 08.02.2018 «Об утверждении Правил категорирования объектов критической информационной инфраструктуры Российской Федерации, а также перечня показателей критериев значимости объектов критической информационной инфраструктуры Российской Федерации и их значений». Decree of the government of Russia No. 127 at 02.08.2018 "About approval of the rules for categorizing objects of critical information infrastructure of the Russian Federation, as well as the list of indicators of criteria for the significance of objects of critical information infrastructure of the Russian Federation and their values".

At the regional level, HL7 standard of FHIR RUcore has already been implemented by almost all leading medical system developers in one or another of their solutions for MIS. The research examines the possibility of implementing the FHIR RU-core protocol in the MIS of "Netrika Medicine" to address new challenges and demands. The "Netrika Medicine" software currently has multiple components, including laboratory and instrumental tests, prescriptions, telemedicine consultations, various types of appointments, but the most significant one is the secure integration of MIS. The software has been successfully used in over 20 regions in Russia.

One of the research tasks is to assist the medical community in attaining long term goals for the digitalization of healthcare. Our method is primarily focused on the safe integration of medical information systems and the development of regional services for doctors, patients and healthcare providers. It is necessary to resolve the contradiction between the level of technological development and the problem of developing a scientific and methodological apparatus.

The Russian EQES was utilized to ensure safe integration with the state unified healthcare system¹. MIS components were responsible for the functions and structures of services of Federal Register of Medical Organizations (FRMO)², federal register of medical workers³, which were posted on the portal for the operational interaction with participants. The implementation of these components involved using enhanced EQES for information interaction.

The implementation of information interaction with the FRMO subsystem has been achieved through the use of CryptoPro Cryptographic Service Provider (CSP) and EQES. If an EQES certificate is available, the following data about medical organizations will be transmitted:

- obtaining basic information about a medical organization;
- obtaining information about buildings;
- obtaining a range of divisions and departments of a medical organization;
- obtaining information about the staffing schedules of a medical organization;
- obtaining information about the medical equipment of a medical organization;
- obtaining information about mobile units;
- obtaining information about households.

³ Федеральный регистр медицинских работников [Электронный ресурс]. URL: https://portal.egisz.rosminzdrav.ru/ materials/4135 (дата обращения: 20.05.2024). Federal register of medical workers, URL: https://portal.egisz.rosminzdrav.ru/ materials/4135 (accessed: 20.05.2024). The innovation of the method lies in the introduction of a certificate into the FHIR RU-core protocol. Previously, these solutions were not used due to the need to adapt the domestic FHIR protocol to the method.

The EQES integration improves the system by promoting digitalization of healthcare. Doctors and MIS owners are able to obtain information about patients, their stories of diseases, treatment methods, and other information in a quick and qualitative manner. The management of departments and healthcare facilities can enhance their control over staff performance. EQES enhances data storage security by protecting it from unauthorized access, modification and destruction.

The software code for EQES integration into MIS required the creation of a function that connected a cryptographic information protection tool — CryptoPro CSP and the use of an EQES certificate to uniquely identify authorship and ensure the integrity of transmitted data.

Generating and sending an Object Identifier (OID) to the FRMO using a certificate and the GOST algorithm is the main aspect of the method. The method can be schematically presented in the form of an algorithm, where *I* means the counter required to set the limit of password submission attempts. The limit is three. The algorithm is presented in Fig. 1. If you enter a PIN code



Fig. 1. Algorithm of signing with EQES data for sending to a medical organization

¹ Единая государственная информационная система в сфере здравоохранения [Электронный ресурс]. URL: https://egisz.rosminzdrav.ru/#firstPag (дата обращения: 20.05.2024). Unified state information system in the field of healthcare, URL: https://egisz.rosminzdrav.ru/#firstPag (accessed: 20.05.2024).

² Федеральный регистр медицинских работников [Электронный ресурс]. URL: https://portal.egisz.rosminzdrav.ru/ materials/4133 (дата обращения: 20.05.2024). Federal register of medical organizations, URL: https://portal.egisz.rosminzdrav. ru/materials/4133 (accessed: 20.05.2024).

three times in a row incorrectly, the use of the EQES will be suspended until unlocked.

Documents must be signed to upload information about medical workers and institutions to the federal services. Cryptographic Message Syntax format is required for the signature to be EQES. EQES should be formed using the algorithms of GOST R 34.10-2012. Before FHIR RU-core is implemented, it is important to create separate files for the signatures of medical professionals and organizations (each electronic signature must correspond to a separate file).

The limited validity of signature certificates is a potential cause of difficulties with using EQES. The closed key is valid for one year. Due to the current demographic situation, there is a severe shortage of IT specialists in the field, which could make it challenging to support and maintain the EQES.

Automated complex debugging testing was performed by generating and sending the object identification to the FRMO using the certificate and algorithm of GOST R 34.10-2012.

When discussing the quality of data transmission, it is important to emphasize that EQES, which is part of the FHIR RU-core structure, ensures the integrity of transmitted data. The speed is significantly faster than manual processing. Without using EQES as part of the FHIR RU-core structure, scalability in actual health applications is not achievable due to the absence of legally legitimate EQES analogues. The process of integrating EQES into the FHIR RU-core structure may require further development after being certified by the quantum cryptographic regulator.

Software implementation

Information interaction is ensured using cryptographic measures of information security in the developed program code¹. EQES certificates provide:

- formation of a qualified electronic signature using GOST R 34.10-2012 algorithms;
- verification of a qualified electronic signature using GOST R 34.10-2012 algorithms.

Thus, the Russian EQES implemented in the HL7 FHIR RU-core protocol meets information security requirements. The unique source process is presented below.

After entering the Internet portal, the user was redirected to a URL generated on the portal side. The URL contains the following data:

- client_id the identifier of the client system (mnemonic of the Unified Identification and Automation System (UIAS) indicated in capital letters);
- *client_secret* the request signature in Public-Key Cryptographic System #7 (PKCS#7) detached signature format. In UTF-8 encoding from the values of four HTTP request parameters: scope, timestamp, clientId, and state (without delimiters). <*client_secret>* must be encoded base64 URL safely. The certificate used to verify the signature must be pre-registered in the UIAS and linked to the health committee of the client system in the UIAS. UIAS uses certificates in the X.509 format and interacts with the algorithms for generating an electronic signature and cryptographic hashing using GOST R 34.10-2012 and GOST R 34.11-2012²;
- *redirect_uri* the link that the user is directed to after being given permission to access the resource;
- scope access area, i.e., rights requested. For example, if a client system requests access to information about employees of an organization, then the access scope should be as in http://esia.gosuslugi.ru/org_emps with the necessary parameters. If the access scope id_doc (user data) is requested, then there is no need to specify the OID of this user as a parameter;
- *response_type* is the type of response that is expected from the UIAS, and it has a value code if the client system needs to receive an authorization code;
- *state* a 128-bit request identifier that is randomly generated and meets the universally unique identifier standard to prevent interceptions;
- timestamp the time of the authorization code request in the format yyyy.MM.dd HH:mm:ss Z (for example, 2013.01.25 14:36:11 +0400), it is necessary to record the beginning of the time period during which the request with this data is a valid identifier (<state>);
- access_type takes the value "offline" if access to resources is required and when the owner cannot be called (in this case, a refresh token is issued). The value is "online" — access is required only if the owner is present.

The program structure for implementing EQES into the FHIR RU-core protocol is presented above. Since the integration platform was created with C Sharp (C#), the programming language choice was predetermined.

Below is an example of obtaining a list of medical organizations that are available. Request type: POST. Listing 1 and 2 show the request and the generated response.

```
Listing 1. Request
```

```
URL-запроса: [base]/term/ValueSet/$expand?_
format=json
Header: Authorization: [GUID-токен]
BODY-запроса:
{"resourceType": "Parameters",
    "parameter": [{
    "name": "system",
```

² Национальный стандарт РФ ГОСТ Р 34.11-2012 «Информационная технология. Криптографическая защита информации. Функция хэширования». National standard of the Russian Federation GOST R 34.11-2012 "Information technology. Cryptographic data security. Hash function".

¹ Приказ ФСБ России от 10.07.2014 № 378 «Об утверждении Состава и содержания организационных и технических мер по обеспечению безопасности персональных данных при их обработке в информационных системах персональных данных с использованием средств криптографической защиты информации, необходимых для выполнения установленных Правительством Российской Федерации требований к защите персональных данных для каждого из уровней защищенности». Order of the FSB of Russia at 10.07.2014 No. 378 "About approval of the composition and content of organizational and technical measures to ensure the security of personal data during their processing in personal data information systems using cryptographic information protection tools necessary to fulfill the requirements established by the government of the Russian Federation personal data protection for each security level".

```
"valueString":"urn:o
id:1.2.643.2.69.1.1.1.84.2"
}
]
}
```

Listing 2. Response

```
"parameter":[{
        "name": "return",
        "resource": {
            "id": "Oda121d9-eeb4-4695-9b3a-
            cb64c3fe9b4a",
            "url": "urn:o
id:1.2.643.2.69.1.1.1.84.2",
            "meta": {
                "versionId": "0e277f6e-
                ed52-405a-afd6-
                5683a19b7b8c",
                «lastUpdated»: «2024-05-
                28T11:02:16.547447+03:00»
            },
            «name»: «Медицинские работники
            медицинских организаций»,
            "status": "active",
            "contact": [{
                "telecom": [{
                     "value": "",
                    «system»: «email»
                }]
            }],
            «version»: «0»,
            «expansion»: {
                «contains»: [{
                    «code»: «1.2.643.5.1.13
                     .13.12.2.78.8000»,
                     «display»: «СПб ГБУЗ
                     \»Стом. пол. №11\"",
                     "version": "0",
                     "contains": [{
                         "code": "mr",
                         "display": "Создано
                         записей: 2″
                     }]
                }],
                "parameter": [{
                    "name": "total",
                     "valueString": "1"
                }1,
                "timestamp": "2024-05-
                28T14:04:25.07663+03:00"
            },
            "publisher": "",
            "experimental": true,
            "resourceType": "ValueSet"
        }
    }],
    "resourceType": "Parameters"
}
```

The received response means that a Medical Organization (MO) is available to the user from OID 1.2.643.5.1.13.13.12.2.78.8000.

Experimental study

The 'Experimental study' includes information about the experiment. The characteristics of the test stand are given in relation to the computer system and the software used. Table 2 presents the steps.

"Netrika Medicine", an IT company that operates in over twenty regions of our country, created the first real example of using the FHIR protocol in health care. In particular, these are such software products as Access Management System, Patient Flow Management, Telemedicine, Patient Portal, and Integrated Electronic Medical Card.

The benchmarks are presented in this section: Case No. 1, Case No. 2, Case No. 3.

The experiment is described in the process of updating information in the Regional Segment (RS) of the FRMO. Table 2 presents the experimental steps.

The experiment requires logistics support.

The experiment should be carried out on personal computers with the following indicators:

- Intel Celeron Processor G1850 2.90 GHz or comparable;
- RAM volume should not be less than 4 GB;
- disk subsystem should not be lower than 80 GB;
- network adapter with at least 100 Mbit;
- Windows 10 and higher, MacOS X 10.10 and higher;
- Google Chrome browser, version 45.0 and higher;
- Google Chrome Advanced REST Client browser plugin with URL version 7.43 and higher;
- CryptoPRO EQES Browser Plug-in;
- CryptoPRO CSP software version 5.00. The following test cases were used: Test case No. 1. Medical Organization Data OID MO: 1.2.643.5.1.13.13.12.2.41.3842 Name of MO: "Kamchatka MO" Test case No. 2. Obtaining Complete Data of a MO Request type: GET URL: [haso]/fbir/tarm/gat_resource2_for

URL: [base]/fhir/term/get_resource?_form at=json&system=1.2.643.2.69.1.1.1.86.2&co de=1.2.643.5.1.13.13.12.2.41.3842

Test case No. 3. Updating MO Data in Register of Medical Organizations (RMO) with Data from FRMO. New Division of the MO

OID MO: 1.2.643.5.1.13.13.12.2.41.3842

Name: Outpatient department

Type of unit: Outpatient

Type of unit: Outpatient clinics (including mobile ones) Buildings and premises: <arbitrary buildings and premises>

Scheduled visits per shift (Total number): 33

Attached Residents (Total Number): 300

Form of medical care: Planned

Conditions for providing medical care: Outpatient

Medical services provided by the unit: <voluntary services>

Updating the information of the RS FRMO with the data from FRMO Uniform State Health Information System (USHIS) is presented in Table 2.

	^	-
No	Steps	Results
1	Open the RS FRMO directory, OID 1.2.643.2.69.1.1.1.86.2. Find an MO from test case No. 1. Open a directory entry to view details	In the graphical user interface of the reference data in directory 1.2.643.2.69.1.1.1.86.2, in the MO data from the test case No. 1, there is no data about the department from the test case No. 3
2	Check if the structure of the medical organization contains the unit from test case No. 3. Execute the request from test case No. 2 in the mail application	In the resulting JSON object of the MO, obtained as a result of executing the request of test case No. 2, in the "departs" information block there is no data about the department from test case No. 3
3	 Add a new division to the federal system FRMO USHIS for Moscow Region from test case No. 1: Log in to the system. Open the MO card from the test case. Add a new department for the ministry using data from test case No. 3 	The department from test case No. 3 is successfully created in the USHIS 2.0
4	 Start the process of obtaining data from the FRMO: Launch and log in to the web application "FRMO". Set parameters for synchronization from test case No. 3 and begin the process of synchronization. Wait for the data acquisition process to complete and the process status changes from "saving result" to "stopped" 	The process of updating MO data from test case No. 1 is successfully launched and completed
5	Access the regulatory reference information web application. Open the RS FRMO directory, OID 1.2.643.2.69.1.1.1.86.2	In the graphical user interface of the reference data in directory 1.2.643.2.69.1.1.1.86.2, the MO data from test case No. 1 contains data about the new unit
6	Find a medical organization with an OID from test case No. 1. Open the directory entry to view details. Find the latest uploaded information. Check if the structure of the medical organization contains the unit from test case No. 3. Execute the request from test case No. 2 in a mail application	In the resulting JSON object obtained as a result of executing the request of test case No. 2, the information block "departs" contains data about new department

Table 2. The experimental steps

The experiment is run as Administrator. Role: Administrator. Updating the data of MO in the RMO with data from the FRMO: adding a unit.

Results

The MIS of "Netrika Medicine" is a real illustration of how cryptographic information security can be integrated into the HL7 FHIR RU-core protocol. The "integration platform", "patient flow management", "access management system", "telemedicine", "unified electronic medical card", and "patient portal" have been deployed in more than 20 regions of the Russian Federation. A quantitative analysis is done to determine the rapid



Fig. 2. A graph showing the annual increase in the number of MIS

increase in MIS interactions through an integration platform. Starting from 2020, there has been a significant annual increase in the number of MIS operating on the HL7 FHIR RU-core standard. This rapid increase is shown in Fig. 2.

Discussion

Practical suggestions for integrating EQES into the HL7 FHIR RU-core standard incorporate the design of a protected MIS. Furthermore, the types of certificates were identified and the necessity for additional information security measures when processing personal data was presented [13].

Moreover, FHIR RU-core is the only standard that is currently popular and has a large professional that is wellversed in medical records description. The key advantage is a clearly thought-out resource structure that enables to describe all the necessary entities: from patients and doctors to immunizations, treatment plans, etc. By combining the same resources in different ways, it is possible to create descriptions of completely different processes using the EQES certificate implemented in FHIR as if they were building blocks.

Limited funding for medical facilities is a potential limitation, as CryptoPRO CSP and EQES certificates will require extra costs. These restrictions are not a significant threat since data transfer to the 'Unified Public Health System' is currently mandatory. Compatibility of the domestic and international versions of FHIR protocol in MIS is possible. To achieve this, it is necessary to acknowledge the initial structures and synchronize goals and objectives in the development of MIS.

Conclusion

The Russian Fast Healthcare Interoperability Resources (FHIR) community has existed for several years. During this time, many important decisions were made. The development of community-created content is currently underway. It is now possible to apply certain advancements. For example, a patient's card makes it possible to integrate Medical Information Systems (MIS) created by different vendors. Commercial companies that support the community and invest both financially and with expertise will have a positive impact on the speed with which all decisions and materials are described, worked out, and recorded. The Russian HL7 FHIR RU-core is faster, but requires appropriate legislation to be available everywhere.

References

- Gorbunov N., Kuleshova V., Korzhuk V. Personal data processed in medical information systems. *International Research Journal*, 2023, no. 9 (135), pp. 10. (in Russian). https://doi.org/10.23670/ IRJ.2023.135.3
- Mikhailenko O.V., Staykov G.B., Dorrer G.A. Using the fast healthcare interoperability resources health information exchange standard in digital healthcare. *ITNOU*, 2021, no. 1 (17), pp. 43–49. (in Russian). https://doi.org/10.47501/ITNOU.2021.1.043-049
- Semenov I., Kopanitsa G., Denisov D., Yakovenko A., Osenev R., Andreychuk Y. Patients decision aid system based on FHIR profiles. *Journal of Medical Systems*, 2018, vol. 42, no. 9, pp. 166. https://doi. org/10.1007/s10916-018-1016-4
- Semenov I., Kopanitsa G. Decision support system based on FHIR profiles. *Studies in Health Technology and Informatics*, 2018, vol. 249, pp. 117–121. https://doi.org/10.3233/978-1-61499-868-6-117
- Semenov I., Osenev R., Gerasimov S., Kopanitsa G., Denisov D., Andreychuk Y. Experience in developing an FHIR medical data management platform to provide clinical decision support. *International Journal of Environmental Research and Public Health*, 2019, vol. 17, no. 1, pp. 73. https://doi.org/10.3390/ijerph17010073
- Leroux H., Metke-Jimenez A., Lawley M.J. Towards achieving semantic interoperability of clinical study data with FHIR. *Journal of Biomedical Semantics*, 2017, vol. 8, pp. 41. https://doi.org/10.1186/ s13326-017-0148-7
- Dolin R.H., Boxwala A., Shalaby J. A Pharmacogenomics Clinical Decision Support service based on FHIR and CDS Hooks. *Methods* of *Information in Medicine*, 2018, vol. 57 (S 02), pp. e115–e123. https://doi.org/10.1055/s-0038-1676466
- Owens D.K. Improving practice guidelines with patient-specific recommendations. *Annals of Internal Medicine*, 2011, vol. 154, no. 9. pp. 638–639. https://doi.org/10.7326/0003-4819-154-9-201105030-00010
- Weber S., Crago E.A., Sherwood P.R., Smith T. Practitioner approaches to the integration of clinical decision support system technology in critical care. *The Journal of Nursing Administration*, 2009, vol. 39, no. 11, pp. 465–469. https://doi.org/10.1097/ NNA.0b013e3181bd5fc2
- Gorbunov N., Kuleshova V., Korzhuk V. Organizational and law aspects of medical information systems. *International Research Journal*, 2024, no. 4 (142), pp. 11. https://doi.org/10.23670/ IRJ.2024.142.38
- Vaganova E.V. Hospital information systems as the object of evaluation: factors and development tendencies. *Tomsk State University Journal of Economics*, 2017, no. 37, pp. 113–130. (in Russian). https://doi.org/10.17223/19988648/37/9

FHIR is the only standard that has detailed descriptions of medical records. In 80 % of cases, the FHIR knowledge base satisfies Russian medical regulatory requirements and health care practitioners. There is a lack of global unifying motive for practical synchronization, which may prevent the Russian and international versions of different companies from being harmonized.

The security of Enhanced Qualified Electronic Signature integration into the FHIR structure is confirmed by the author's identification and ensuring the integrity of the information.

This should become a strong basis for the further digitalization of medicine in Russia.

In conclusion, it is crucial to emphasize that the materials that describe the systematization of the Russian FHIR RU-core protocol for new MIS have been validated for their novelty, relevance, methodological, and practical value. Thus, when developing MIS with new functionality, it is necessary to use the innovations presented in this article.

Литература

- Горбунов Н.А., Кулешова В.О., Коржук В.М. Персональные данные, обрабатываемые в медицинских информационных системах // Международный научно-исследовательский журнал. 2023. № 9 (135). С. 10. https://doi.org/10.23670/IRJ.2023.135.3
- Михайленко О.В., Стайков Г.Б., Доррер Г.А. Использование стандарта обмена медицинской информацией fast healthcare interoperability resources в цифровом здравоохранении // ИТНОУ: Информационные технологии в науке, образовании и управлении. 2021. № 1. С. 43–49. https://doi.org/10.47501/ ITNOU.2021.1.043-049
- Semenov I., Kopanitsa G., Denisov D., Yakovenko A., Osenev R., Andreychuk Y. Patients decision aid system based on FHIR profiles // Journal of Medical Systems. 2018. V. 42. N 9. P. 166. https://doi. org/10.1007/s10916-018-1016-4
- Semenov I., Kopanitsa G. Decision support system based on FHIR profiles // Studies in Health Technology and Informatics. 2018. V. 249. P. 117–121. https://doi.org/10.3233/978-1-61499-868-6-117
- Semenov I., Osenev R., Gerasimov S., Kopanitsa G., Denisov D., Andreychuk Y. Experience in developing an FHIR medical data management platform to provide clinical decision support // International Journal of Environmental Research and Public Health. 2019. V. 17. N 1. P. 73. https://doi.org/10.3390/ijerph17010073
- Leroux H., Metke-Jimenez A., Lawley M.J. Towards achieving semantic interoperability of clinical study data with FHIR // Journal of Biomedical Semantics. 2017. V. 8. P. 41. https://doi.org/10.1186/ s13326-017-0148-7
- Dolin R.H., Boxwala A., Shalaby J. A Pharmacogenomics Clinical Decision Support service based on FHIR and CDS Hooks // Methods of Information in Medicine. 2018. V. 57 (S 02). P. e115–e123. https:// doi.org/10.1055/s-0038-1676466
- Owens D.K. Improving practice guidelines with patient-specific recommendations // Annals of Internal Medicine. 2011. V. 154. N 9. P. 638–639. https://doi.org/10.7326/0003-4819-154-9-201105030-00010
- Weber S., Crago E.A., Sherwood P.R., Smith T. Practitioner approaches to the integration of clinical decision support system technology in critical care // The Journal of Nursing Administration. 2009. V. 39. N 11. P. 465–469. https://doi.org/10.1097/ NNA.0b013e3181bd5fc2
- Gorbunov N., Kuleshova V., Korzhuk V. Organizational and law aspects of medical information systems // International Research Journal. 2024. N 4 (142). P. 11. https://doi.org/10.23670/IRJ.2024.142.38
- Ваганова Е.В. Медицинские информационные системы как объект оценки: факторы и тенденции развития // Вестник Томского государственного университета. Экономика. 2017. № 37. С. 113– 130. https://doi.org/10.17223/19988648/37/9

- 12. Mavlyanova L.T. Protection of information on the Internet. *Oriental Renaissance: Innovative, Educational, Natural and Social Sciences*, 2022, vol. 2, no. 1, pp. 568–575. (in Russian)
- Alsafwani N., Fazea Y., Alnajjar F. Strategic approaches in network communication and information security risk assessment. *Information*, 2024, vol. 15, no. 6, pp. 353. https://doi.org/10.3390/ info15060353

Authors

Nikolai A. Gorbunov — Head of Department, "Netrika Medicine" LLC, Saint Petersburg, 191015, Russian Federation; PhD Student, ITMO University, Saint Petersburg, 197101, Russian Federation, https://orcid. org/0009-0004-2973-9594, gorb-2157@mail.ru

Valeria O. Kuleshova — PhD (Philology), CSRI Elektropribor, Saint Petersburg, 197046, Russian Federation, https://orcid.org/0000-0003-1377-6003, valeria.kuleshova@yahoo.com

Viktoriia M. Korzhuk — PhD, Associate Professor, Associate Professor, ITMO University, Saint Petersburg, 197101, Russian Federation, SC 56875395200, https://orcid.org/0000-0002-0240-9067, vmkorzhuk@ itmo.ru

Received 20.06.2024 Approved after reviewing 21.01.2025 Accepted 18.03.2025

- 12. Мавлянова Л.Т. Защита информации в интернет // Oriental Renaissance: Innovative, Educational, Natural and Social Sciences. 2022. V. 2. N 1. C. 568–575.
- Alsafwani N., Fazea Y., Alnajjar F. Strategic approaches in network communication and information security risk assessment // Information. 2024. V. 15. N 6. P. 353. https://doi.org/10.3390/ info15060353

Авторы

Горбунов Николай Александрович — руководитель отдела, ООО «Нетрика Медицина», Санкт-Петербург, 191015, Российская Федерация; аспирант, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, https://orcid.org/0009-0004-2973-9594, gorb-2157@mail.ru

Кулешова Валерия Олеговиа — кандидат филологических наук, преподаватель, АО «Концерн «ЦНИИ «Электроприбор», Санкт-Петербург, 197046, Российская Федерация, https://orcid.org/0000-0003-1377-6003, valeria.kuleshova@yahoo.com

Коржук Виктория Михайловна — кандидат технических наук, доцент, доцент, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, sc 56875395200, https://orcid.org/0000-0002-0240-9067, vmkorzhuk@itmo.ru

Статья поступила в редакцию 20.06.2024 Одобрена после рецензирования 21.01.2025 Принята к печати 18.03.2025



Работа доступна по лицензии Creative Commons «Attribution-NonCommercial» **VİTMO**

НАУЧНО-ТЕХНИЧЕСКИЙ ВЕСТНИК ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ, МЕХАНИКИ И ОПТИКИ март–апрель 2025 Том 25 № 2 http://ntv.ifmo.ru/ SCIENTIFIC AND TECHNICAL JOURNAL OF INFORMATION TECHNOLOGIES, MECHANICS AND OPTICS March–April 2025 Vol. 25 № 2 http://ntv.ifmo.ru/en/ ISSN 2226-1494 (print) ISSN 2500-0373 (online)

ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ, МЕХАНИКИ И ОПТИКІ

doi: 10.17586/2226-1494-2025-25-2-321-327 УДК 004.032.26

Метод увеличения разрешения изображения с использованием референсных изображений на основе диффузионной модели Алексей Константинович Денисов^{1⊠}, Сергей Вячеславович Быковский², Павел Валерьевич Кустарев³

1,2,3 Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация

¹ denisov@itmo.ru^{\Box}, https://orcid.org/0000-0001-8135-1723

² sergei bykovskii@itmo.ru, https://orcid.org/0000-0003-4163-9743

³ kustarev@itmo.ru, https://orcid.org/0000-0001-9326-0837

Аннотация

Введение. В настоящий момент активно развиваются различные методы восстановления изображений на основе методов глубокого машинного обучения. С помощью таких методов решаются задачи восстановления утраченных областей, подавления шумов и увеличения разрешения изображений. В задаче увеличения разрешения важную роль играют методы, основанные на применении референсных изображений, позволяющих восстановить недостающую информацию на основном изображении. Такие методы реализуются с использованием сверточных нейронных сетей, широко востребованных в задачах компьютерного зрения. В применяемых в настоящее время методах область изображения, не представленная на референсном изображении, часто отличается худшим качеством по сравнению с остальным изображением, что заметно визуально. Наряду со сверточными нейронными сетями в задачах восстановления изображений активно применяются диффузионные модели, позволяющие генерировать изображения с высоким качеством и четкостью, однако их недостатком часто бывает несоответствие сгенерированных деталей реальным. В работе обсуждается проблема улучшения качества восстановления изображений на основе применения референсных изображений с использованием диффузионной модели. Метод. Для получения хорошего конечного результата предложена гибридная архитектура нейронной сети диффузионной модели, состоящая из трех основных блоков: базового модуля диффузионной модели, модуля использования референсной информации и модуля слияния. Обучение предложенной гибридной модели, а также сравниваемой с ней сверточной нейронной сети, использующей референсные изображения, и диффузионной моделью выполнено с использованием набора данных Large-Scale Multi-Reference Dataset (LMR). Основные результаты. По результатам тестирования обученных моделей на тестовой выборке набора данных LMR проведено качественное (визуальное) и количественное сравнение работы трех моделей. Гибридная модель продемонстрировала более высокое качество, четкость и однородность изображения в сравнении со сверточной нейронной сетью с использованием референсных изображений и лучшее восстановление реальных деталей по сравнению с диффузионной моделью. Количественные оценки подтвердили, что гибридная модель также показала более высокие результаты по сравнению с остальными моделями. Обсуждение. Результаты работы могут быть использованы для увеличения разрешения любых изображений с использованием референсной информации.

Ключевые слова

обработка изображений, диффузионные модели, super-resolution, глубокое обучение, восстановление изображений

Ссылка для цитирования: Денисов А.К., Быковский С.В., Кустарев П.В. Метод увеличения разрешения изображения с использованием референсных изображений на основе диффузионной модели // Научнотехнический вестник информационных технологий, механики и оптики. 2025. Т. 25, № 2. С. 321–327. doi: 10.17586/2226-1494-2025-25-2-321-327

[©] Денисов А.К., Быковский С.В., Кустарев П.В., 2025

Reference-based diffusion model for super-resolution

Aleksei K. Denisov^{1⊠}, Sergei V. Bykovskii², Pavel V. Kustarev³

1,2,3 ITMO University, Saint Petersburg, 197101, Russian Federation

¹ denisov@itmo.ru^{\boxdots}, https://orcid.org/0000-0001-8135-1723

² sergei_bykovskii@itmo.ru, https://orcid.org/0000-0003-4163-9743

³ kustarev@itmo.ru, https://orcid.org/0000-0001-9326-0837

Abstract

This article is devoted to digital image processing algorithms, namely, super-resolution task. Currently, various methods of image restoration based on deep learning are actively developing. These methods are used to solve image restoration problems, such as inpainting, denoising, and super-resolution. One important class of super-resolution methods is reference-based super-resolution that allows restoring the missing information in the main image using reference images. Methods of this class are mainly represented by convolutional neural networks which are widely used in computer vision problems. Despite the significant achievements of existing methods, they have one significant drawback: the image area not represented in the reference image often has worse quality compared to the rest of the image, which is clearly visible to the observer. In addition to convolutional neural networks, diffusion models are actively used in image restoration problems. They are capable of generating images with high quality and diverse fine details but suffer from a lack of fidelity between the generated details and the real ones. The aim of this work is to improve the quality of the reference-based image restoration method using the diffusion model. A hybrid architecture of the diffusion model denoising neural network is proposed consisting of three main blocks: the basic denoising module, the reference-based module, and the fusion module for the final result generation. Three models were trained: a diffusion model, a referencebased convolutional neural network, and a proposed hybrid model. All three models were trained and evaluated on the Large-Scale Multi-Reference Dataset dataset. Based on the results of the trained models testing, a qualitative (visual) and quantitative comparison of the three models was done. The hybrid model demonstrated higher image quality, clarity, and consistency compared to the convolutional neural network using references and better restoration of real details compared to the diffusion model. According to the quantitative evaluation, the hybrid model also showed higher results compared to pure models. The results of this work can be used to increase the resolution of any images using reference information.

Keywords

image processing, diffusion models, super-resolution, deep learning, image restoration

For citation: Denisov A.K., Bykovskii S.V., Kustarev P.V. Reference-based diffusion model for super-resolution. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 2025, vol. 25, no. 2, pp. 321–327 (in Russian). doi: 10.17586/2226-1494-2025-25-2-321-327

Введение

В эпоху быстрого развития цифровых технологий постоянно растет спрос на высококачественные изображения в различных приложениях, от медицинской диагностики до сферы развлечений. Несовершенство процесса получения цифровых изображений, начиная от конструктивных особенностей оптических систем и специфики работы сенсоров до сложных условий съемки, связанных с погодными условиями, низкой освещенностью, движением объектов сцены и т. д., делает необходимыми алгоритмы восстановления и улучшения качества изображений. Одной из важных подзадач улучшения изображений является увеличение разрешения (Super-Resolution, SR). Выделяются методы, работающие с одним изображением (Single Image Super-Resolution, SISR), и использующие референсные изображения (Reference-Based Super-Resolution, RefSR). RefSR выступает в качестве перспективного решения, которое использует дополнительные референсные изображения высокого разрешения для повышения качества целевого изображения с низким разрешением. В отличие от традиционных SISR-методов, RefSR использует большое количество деталей, доступных в референсных изображениях для более реалистичной реконструкции. Глубокое обучение, в частности сверточные нейронные сети (CNN), произвело революцию в области обработки изображений, предоставив мощные инструменты для моделирования сложных отображений между входными и выходными изображениями. Одной из особенностей RefSR-методов на основе CNN является существенное различие в качестве деталей восстановленного изображения в областях, присутствующих и отсутствующих на референсных изображениях.

Одновременно с этим развиваются SISR-методы, использующие в основе диффузионные модели. Такие модели способны за счет итеративного процесса генерации выходного изображения восстанавливать большое количество реалистичных деталей. Проблемой этих методов является нежелательная генерация неверных деталей в тех случаях, когда заведомо известно, что должно быть изображено в данном месте (например, известного архитектурного сооружения). Иными словами, классические RefSR-методы страдают от наличия на результирующем изображении областей с низкой четкостью и качеством деталей, а мощные SISR-методы на основе диффузионных моделей генерируют ложные детали, которые хорошо заметны в некоторых случаях. В настоящей работе делается попытка исправить недостатки двух SR-методов и использовать их сильные стороны для решения задачи SR-изображений с использованием референсной информации.

Обзор научных публикаций

Методы SR-изображений за последние годы претерпели значительные изменения, особенно с внедрением методов глубокого обучения, которые преобразовали область обработки изображений. В работе [1] были заложены основы использования CNN в задачах SR
и продемонстрированы значительные улучшения по сравнению с традиционными методами интерполяции. Основываясь на этом успехе, авторы работы [2] предложили архитектуру VDSR, которая использовала очень глубокую сеть для увеличения качества с помощью обучения восстановлению остаточной информации (residual learning). В [3] была представлена улучшенная глубокая остаточная сеть (EDSR). В настоящей работе архитектура сети была оптимизирована для удаления ненужных модулей с целью повышения эффективности обучения и точности SR. Генеративно-состязательные сети (GAN) также оказали значительное влияние на SR с появлением SRGAN [4]. Целью работы [4] было создание визуально привлекательных результатов SR с помощью применения состязательного обучения для улучшения перцептивного качества изображений. В [5] значительное внимание уделено синтезированию тренировочных данных для имитации различных деградаций изображений для улучшения качества работы в реальных сценариях. Эти исследования иллюстрируют быстрый прогресс в методах SR-изображений, подчеркивая переход от классических технологий к более сложным моделям глубокого обучения, которые используют возможности CNN для повышения качества изображения.

Параллельно с методами SISR важной областью исследований стали методы RefSR, особенностью которого является использование дополнительных изображений, содержащих часть той же сцены в высоком разрешении. Метод SRNTT [6] использует многоуровневую структуру для переноса текстур между фрагментами входного и референсного изображений. Фреймворк C2-Matching, предложенный в [7], использует три стадии обучения, в первых двух из которых с помощью контрастивного обучения тренируется сеть извлечения признаков из целевого и референсного изображений, а на третьей стадии обучается сеть, восстанавливающая результирующее изображение на основе извлеченных и сопоставленных признаков. Еще одним заметным вкладом является метод DATSR [8], в котором применяется механизм адаптивного внимания (deformable attention) для динамического фокуса на соответствующих признаках в референсных изображениях, позволяя модели генерировать изображения с высоким разрешением, лучше справляясь с пространственными вариациями в данных. Важной проблемой RefSR-методов является дефицит данных для обучения. В работе [9] представлен LMR: большой многообразный набор данных для обучения и бенчмаркинга различных RefSR-моделей, кроме того, в работе был предложен метод, позволяющий использовать сразу несколько референсных изображений и улучшить таким образом качество выходного изображения. В [10] используется предобученная диффузионная модель для генерации референсного изображения, детали с которого затем переносятся на целевое изображение. Эти исследования подчеркивают значительный прогресс в методах RefSR и показывают перспективы их использования. При этом одной из существенных проблем остается восстановление областей изображения, отсутствующих на референсе и из-за этого страдающих от недостатка реалистичных текстур.

В то время как большинство методов обработки изображений использовали в своей основе CNN и трансформеры, в последнее время активно стали применяться диффузионные модели. Такие модели стали мощной основой в генеративном моделировании. Диффузионные модели направлены на усваивание структуры сложных распределений данных с помощью процесса, который итеративно преобразует зашумленные изображения в образцы данных из целевого распределения. Основополагающим вкладом в диффузионные модели является работа Denoising Diffusion Probabilistic Models (DDPM) [11], где представлен фреймворк, в котором процесс прямой диффузии постепенно добавляет шум к данным, в то время как обучаемый обратный процесс удаляет этот шум для генерации новых образцов данных. В работе DDPM подчеркивается важность зависимости интенсивности шума от шага диффузии и типов параметризации, устанавливаются базовые уровни качества, на которых основаны последующие исследования. Основываясь на DDPM, в [12] представлен метод Denoising Diffusion Implicit Models (DDIM), который направлен на повышение эффективности генерации путем предложения неявного процесса сэмплинга. Использование DDIM приводит к более быстрому процессу генерации при сохранении высокого качества. Работа DDIM подчеркивает универсальность подходов, основанных на диффузии, предлагая практические идеи относительно вычислительной эффективности.

Дальнейшие достижения видны в разработке моделей скрытой диффузии (Latent Diffusion Models, LDM) [13]. Ключевая идея LDM состоит в выполнении процессов диффузии в сжатом скрытом пространстве, например, вариационного автоэнкодера, а не напрямую в пространстве данных высокой размерности. Это значительно снижает вычислительную сложность при сохранении высокого качества генерации, демонстрируя применимость модели к таким задачам, как синтез изображений и видео высокого разрешения. Помимо использования в задачах генерации изображений, диффузионные модели нашли свое применение в различных image-to-image задачах, например восстановлении изображений. Используются два основных подхода. В первом — обучение диффузионной модели с нуля целевой задаче, при этом изображение для восстановления обычно подается внутрь денойзера диффузионной модели вместе с шумом. Так решаются image-to-image задачи в таких методах как SRDiff [14], LDM [13]. Другим подходом является использование больших предобученных генеративных моделей и дальнейшее обучение целевой задаче специальных адаптеров к ним. Примером такого метода является работа SUPIR [15], использующая Stable Diffusion XL в качестве базовой модели.

Предложенный метод

В данной работе предлагается гибридная диффузионная модель, использующая помимо целевого изображения низкого разрешения одно или несколько референсных изображений.

В качестве основы используется LDM, представляющая собой вероятностную модель, которая выучивает распределение данных p(x) путем постепенного удаления шума из сэмпла, полученного из нормального распределения, что соответствует выучиванию обратного процесса фиксированной марковской цепи длины Т. Эти модели можно интерпретировать как равновзвешенную последовательность шумоподавляющих автоэнкодеров $\epsilon_{\theta}(x_t, t); t = 1, ..., T$, которые обучены предсказывать чистую версию входа x_t, где x_t — зашумленная версия входа х. Для эффективного сжатия входных данных и возможности работать с более большими изображениями используется VQ-VAE, состоящий из энкодера \mathcal{E} и декодера \mathcal{D} с коэффициентом сжатия f = 4. Таким образом, диффузионная модель работает с латентной версией входа $z = \mathcal{E}(x)$, а для получения результирующего изображения в пространстве пикселов необходимо пропустить выход модели через декодер. При этом обучение происходит в латентном пространстве, минимизируется следующая функция потерь: L_{LDM} : = $\mathbb{E}_{z,\epsilon,t}[\|\epsilon - \epsilon_{\theta}(z_t, t)\|_2^2]$, где $\epsilon \sim \mathcal{N}(0, 1)$ — гауссов шум, добавляемый на каждом шаге; 0 — параметры модели; *t* — шаг времени, случайно выбираемый во время обучения; $\mathbb{E}[\cdot]$ — математическое ожидание.

Для переноса информации с референсных изображений используется механизм, предложенный в работе [9]. Для каждого референсного изображения отдельно проводится следующая процедура: из изображения низкого разрешения (LR) и референсного изображения извлекаются признаки на разных уровнях; между фрагментами этих карт признаков производится сопоставление; для каждого фрагмента LR выбирается наилучший фрагмент референсного изображения (в пространстве признаков); с помощью адаптивной свертки (deformable convolution) производится уточнение и улучшение агрегированного референсного изображения. После того как полностью выполнена процедура, применяется специальный модуль слияния, комбинирующий несколько агрегированных референсных изображений в одно, после чего производится восстановление LR с использованием полученного комбинированного референсного изображения. В настоящей работе используется претренированная модель данного модуля.

Архитектура предложенной модели представлена на рис. 1. Модель состоит из трех модулей: U-Net сети денойзера (Diffusion Unet), извлечения информации из референсных изображений (модуль обозначен как RefSR, референсные изображения — как Refs) и сли-



Рис. 1. Предложенная гибридная модель восстановления изображения.

Красным цветом выделены обучаемые модули, синим — претренированные

Fig. 1. Proposed hybrid restoration model architecture. Blue denotes frozen modules, red denotes trained modules



Рис. 2. Схема предложенного модуля слияния (Fusion Module).

VQE — VQ Encoder, RFE — Reference Feature Extractor, WG — Weights Generator. Красным цветом выделены обучаемые модули, синим — претренированные

Fig. 2. Proposed fusion module.

VQE — VQ Encoder, RFE — Reference Feature Extractor, WG — Weights Generator. Blue denotes frozen modules, red denotes trained modules

яния (Fusion Module). На вход поступают LR и референсные изображения. В результате работы модели получается восстановленное изображение SR. На рисунке не показаны предобученные энкодер и декодер, используемые для работы в латентном пространстве.

На рис. 2 представлена схема разработанного модуля слияния. Изображение с перенесенной референсной информацией проходит через VQ энкодер, затем через блок извлечения признаков (RFE), состоящий из пяти остаточных блоков, после чего конкатенируется с признаками из U-Net и идет в модуль генерации весовых коэффициентов (WG), состоящий из трех сверточных слоев. Финальный результат (SR) получается путем взвешенного суммирования по следующей формуле (обозначения соответствуют рис. 2): $x_f = Wx + (1 - W)r$, где x — исходные признаки из U-Net; x_f — измененные признаки U-Net; W — весовые коэффициенты; r — признаки, извлеченные из изображения с перенесенной референсной информацией.

Результаты

Для сравнения предлагаемого метода со стандартной диффузионной моделью (LDM) и сверточной сетью с использованием референсной информации (LMR) был использован набор данных из работы [9]. Он содержит 112 142 набора [целевое изображение, референсное изображение 1 ... референсное изображение 5] для обучения, а также 142 набора с переменным (2–6) количеством референсных изображений для тестирования. В качестве модели деградации использовался 4x bicubic downsampling.

Так как используемый набор данных был предложен в той же работе, что метод LMR, была использована предобученная авторами модель. LDM и предлагаемая гибридная модель были обучены по следующей схеме: целевое изображение низкого разрешения подавалось в диффузионную модель, его каналы склеивались (конкатенировались) с каналами зашумленного целевого изображения высокого разрешения. При этом в модуль слияния гибридной модели одновременно подавалось изображение, восстановленное с помощью референсной информации. Обучение проводилось в течение 400 000 итераций с размером батча 4 на 8 NVIDIA V100 на изображениях размером 512 × 512 пикселов. В качестве метрик для оценки качества классификации использовались следующие метрики: Learned Perceptual Image Patch Similarity (LPIPS) [16], Contrastive Language-Image Pre-training Image Quality Analysis (CLIP-IQA) [17], Fréchet Inception Distance (FID) [18]. Метрика LPIPS представляет собой расстояние между признаками, вычисленными от эталонного и оцениваемого изображений. Признаки вычисляются с помощью предобученной сети, например, AlexNet. Авторы данной метрики показывают значительную корреляцию с перцептивной оценкой качества изображения. Метрика CLIP-IQA построена на основе мультимодальной модели CLIP, связывающей изображения с их описаниями. Данная метрика основана на вычислении косинусного расстояния между описанием оцениваемого изображения и некоторым набором описаний заведомо «хорошего» и «плохого» изображений. Таким образом, для оценки качества изображения не требуется эталон. Метрика FID используется для оценки качества генерации и показывает расстояние между целевым распределением и распределением сгенерированных сэмплов после обработки их с помощью предобученной Inception v3 и вычисления средних и ковариаций последнего слоя.

В таблице приведены результаты обучения трех рассматриваемых методов: сверточной сети с использованием референсных изображений (LMR), обычной диффузионной модели (LDM) и предлагаемой гибрид-

Таблица. Результаты сравнения качества восстановленных изображений *Table*. Quantitative evaluation of the proposed method

Метод	LPIPS-Alexnet ↓	CLIP-IQA↑	FID ↓
LMR	0,1488	0,4678	15,225
LDM	0,1480	0,5857	11,400
LDM-Ref	0,1156	0,6063	10,169

Примечание. Стрелки у названия метрик указывают на уменьшение (↓) или увеличение (↑) величины метрики с увеличением качества изображения.



Рис. 3. Изображения, полученные в результате работы трех сравниваемых методов: LMR (a); LDM (b); гибридная модель (LDM-Ref) (c)
Fig. 3. Visual results of trained models: LMR (a); LDM (b); hybrid model (LDM-Ref) (c)

Научно-технический вестник информационных технологий, механики и оптики, 2025, том 25, № 2 Scientific and Technical Journal of Information Technologies, Mechanics and Optics, 2025, vol. 25, no 2 ной диффузионной модели с использованием референсных изображений (LDM-Ref).

Из таблицы видно, что предлагаемый метод улучшает все рассматриваемые метрики, используя сильные стороны каждой из составных частей гибридной модели. По сравнению с LDM и LMR улучшается попиксельное сходство текстур выходного изображения и эталонного (в LDM эта метрика страдает от ненастоящих деталей, в LMR — от плохо восстановленных областей, где отсутствует референс), что показывает уменьшение LPIPS более чем на 20 %. CLIP-IQA и FID особенно улучшаются по сравнению с LMR (увеличение почти на 30 % CLIP-IQA и уменьшение FID на 33 %) благодаря использованию диффузионной модели и генерации более четких текстур.

На рис. 3 представлены результаты работы трех сравниваемых методов. Использованы изображения из набора данных LMR [9]. Красными стрелками показаны области интереса. Видно, что LDM генерирует ненастоящие детали — в области колеса, стрельчатой арки, в то время как на результате работы LMR присутствуют слаботекстурированные области. Предлагаемый метод стремится исправить оба этих недостатка, балансируя между количеством деталей и их реалистичностью, увеличивая четкость и резкость по сравнению с LMR, и однородность, и согласованность деталей и текстур по сравнению с LDM.

Литература

- Dong C., Loy C.C., He K., Tang X. Image Super-Resolution using deep convolutional networks // IEEE Transactions on Pattern Analysis and Machine Intelligence. 2016. V. 38. N 2. P. 295–307. https://doi. org/10.1109/TPAMI.2015.2439281
- Kim J., Lee J.K., Lee K.M. Accurate image Super-Resolution using very deep convolutional networks // Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2016. P. 1646– 1654. https://doi.org/10.1109/CVPR.2016.182
- Lim B., Son S., Kim H., Nah S., Lee K.M. Enhanced deep residual networks for single image Super-Resolution // Proc. of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). 2017. P. 1132–1140. https://doi.org/10.1109/ CVPRW.2017.151
- Ledig C., Theis L., Huszár F., Caballero J., Cunningham A., Acosta A., Aitken A., Tejani A., Totz J., Wang Z., Shi W. Photo-realistic single image Super-Resolution using a generative adversarial network // Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2017. P. 105–114. https://doi.org/10.1109/10.1109/ CVPR.2017.19
- Wang X., Xie L., Dong C., Shan Y. Real-ESRGAN: training real-world blind Super-Resolution with pure synthetic data // Proc. of the IEEE/ CVF International Conference on Computer Vision Workshops (ICCVW). 2021. P. 1905–1914. https://doi.org/10.1109/ ICCVW54120.2021.00217
- Zhang Z., Wang Z., Lin Z., Qi H. Image Super-Resolution by neural texture transfer // Proc. of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2019. P. 7974–7983. https:// doi.org/10.1109/CVPR.2019.00817
- Jiang Y., Chan K.C.K., Wang X., Loy C.C., Liu Z. Robust Referencebased Super-Resolution via C2-Matching // Proc. of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2021. P. 2103–2112. https://doi.org/10.1109/CVPR46437.2021.00214
- Cao J., Liang J., Zhang K., Li Y., Zhang Y., Wang W., Van Gool L. Reference-based image Super-Resolution with deformable attention transformer // Lecture Notes in Computer Science. 2022. V. 13678. P. 325–342. https://doi.org/10.1007/978-3-031-19797-0_19

Заключение

В работе были рассмотрены достоинства и недостатки двух существующих классов методов увеличения разрешения изображений: сверточных нейронных сетей с использованием референсных изображений и диффузионных моделей. В то время как первые страдают от низкого качества областей, где отсутствует референсная информация, вторые генерируют ложные детали на изображении.

Была предложена гибридная модель, построенная на базе LDM, включающая в себя специальный модуль слияния, добавляющий референсную информацию. По результатам визуальной оценки, а также количественного сравнения предложенного метода с чистой диффузионной моделью и чистой сверточной сетью с использованием референсных изображений, было показано преимущество в перцептивном качестве предложенного метода, а также уменьшение метрики LPIPS на минимум 20 % и значительное улучшение по CLIP-IQA и FID по сравнению с обоими рассматриваемыми методами.

Предложенный метод может применяться для увеличения разрешения любых изображений с использованием референсной информации.

References

- Dong C., Loy C.C., He K., Tang X. Image Super-Resolution using deep convolutional networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2016, vol. 38, no. 2, pp. 295–307. https://doi. org/10.1109/TPAMI.2015.2439281
- Kim J., Lee J.K., Lee K.M. Accurate image Super-Resolution using very deep convolutional networks. *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 1646– 1654. https://doi.org/10.1109/CVPR.2016.182
- Lim B., Son S., Kim H., Nah S., Lee K.M. Enhanced deep residual networks for single image Super-Resolution. Proc. of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2017, pp. 1132–1140. https://doi.org/10.1109/ CVPRW.2017.151
- Ledig C., Theis L., Huszár F., Caballero J., Cunningham A., Acosta A., Aitken A., Tejani A., Totz J., Wang Z., Shi W. Photo-realistic single image Super-Resolution using a generative adversarial network. *Proc.* of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, pp. 105–114. https://doi.org/10.1109/10.1109/ CVPR.2017.19
- Wang X., Xie L., Dong C., Shan Y. Real-ESRGAN: training real-world blind Super-Resolution with pure synthetic data. *Proc. of the IEEE/ CVF International Conference on Computer Vision Workshops* (*ICCVW*), 2021, pp. 1905–1914. https://doi.org/10.1109/ ICCVW54120.2021.00217
- Zhang Z., Wang Z., Lin Z., Qi H. Image Super-Resolution by neural texture transfer. Proc. of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019, pp. 7974–7983. https:// doi.org/10.1109/CVPR.2019.00817
- Jiang Y., Chan K.C.K., Wang X., Loy C.C., Liu Z. Robust Referencebased Super-Resolution via C2-Matching. *Proc. of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2021, pp. 2103–2112. https://doi.org/10.1109/CVPR46437.2021.00214
- Cao J., Liang J., Zhang K., Li Y., Zhang Y., Wang W., Van Gool L. Reference-based image Super-Resolution with deformable attention transformer. *Lecture Notes in Computer Science*, 2022, vol. 13678, pp. 325–342. https://doi.org/10.1007/978-3-031-19797-0_19

- Zhang L., Li X., He D., Li F., Ding E., Zhang Z. LMR: a large-scale multi-reference dataset for Reference-based Super-Resolution // Proc. of the IEEE/CVF International Conference on Computer Vision (ICCV). 2023. P. 13072–13081. https://doi.org/10.1109/ ICCV51070.2023.01206
- Li G., Xing W., Zhao L., Lan Z., Sun J., Zhang Z., Zhang Q., Lin H., Lin Z. Self-Reference image Super-Resolution via pre-trained diffusion large model and window adjustable transformer // Proc. of the 31st ACM International Conference on Multimedia. 2023. P. 7981–7992. https://doi.org/10.1145/3581783.3611866
- Ho J., Jain A., Abbeel P. Denoising diffusion probabilistic models // arXiv. 2020. arXiv:2006.11239. https://doi.org/10.48550/ arXiv.2006.11239
- Song J., Meng C., Ermon S. Denoising diffusion implicit models // arXiv. 2020. arXiv:2010.02502. https://doi.org/10.48550/ arXiv.2010.02502
- Rombach R., Blattmann A., Lorenz D., Esser P., Ommer B. High-Resolution image synthesis with latent diffusion models // Proc. of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2022. P. 10674–10685. https://doi.org/10.1109/ CVPR52688.2022.01042
- Li H., Yang Y., Chang M., Chen S., Feng H., Xu Z., Li Q., Chen Y. SRDiff: Single Image Super-Resolution with diffusion probabilistic models // Neurocomputing. 2022. V. 479. P. 47–59. https://doi. org/10.1016/j.neucom.2022.01.029
- Yu F., Gu J., Li Z., Liu J., Kong X., Wang X., He J., Qiao Y., Dong C. Scaling Up to Excellence: practicing model scaling for photo-realistic image restoration in the wild // Proc. of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2024. P. 25669–25680. https://doi.org/10.1109/CVPR52733.2024.02425
- Zhang R., Isola P., Efros A.A., Shechtman E., Wang O. The unreasonable effectiveness of deep features as a perceptual metric // Proc. of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2018. P. 586–595. https://doi.org/10.1109/ CVPR.2018.00068
- Wang J., Chan K.C.K., Loy C.C. Exploring CLIP for assessing the look and feel of images // Proc. of the 37th AAAI Conference on Artificial Intelligence. 2023. V. 37. N 2. P. 2555–2563. https://doi. org/10.1609/aaai.v37i2.25353
- Heusel M., Ramsauer H., Unterthiner T., Nessler B., Hochreiter S. GANs trained by a two time-scale update rule converge to a local nash equilibrium // Proc. of the 31st International Conference on Neural Information Processing Systems (NIPS '17). 2017. P. 6629–6640.

Авторы

Денисов Алексей Константинович — ассистент, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, sc 57210698353, https://orcid.org/0000-0001-8135-1723, denisov@ itmo.ru

Быковский Сергей Вячеславович — кандидат технических наук, доцент, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, вс 57216469537, https://orcid.org/0000-0003-4163-9743, sergei_bykovskii@itmo.ru

Кустарев Павел Валерьевич — кандидат технических наук, декан, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, sc 35317916600, https://orcid.org/0000-0001-9326-0837, kustarev@itmo.ru

Статья поступила в редакцию 26.11.2024 Одобрена после рецензирования 05.02.2025 Принята к печати 19.03.2025

- Zhang L., Li X., He D., Li F., Ding E., Zhang Z. LMR: a large-scale multi-reference dataset for Reference-based Super-Resolution. *Proc.* of the IEEE/CVF International Conference on Computer Vision (ICCV), 2023, pp. 13072–13081. https://doi.org/10.1109/ ICCV51070.2023.01206
- Li G., Xing W., Zhao L., Lan Z., Sun J., Zhang Z., Zhang Q., Lin H., Lin Z. Self-Reference image Super-Resolution via pre-trained diffusion large model and window adjustable transformer. *Proc. of the 31st ACM International Conference on Multimedia*, 2023, pp. 7981–7992. https://doi.org/10.1145/3581783.3611866
- Ho J., Jain A., Abbeel P. Denoising diffusion probabilistic models. arXiv, 2020, arXiv:2006.11239. https://doi.org/10.48550/ arXiv.2006.11239
- Song J., Meng C., Ermon S. Denoising diffusion implicit models. arXiv, 2020, arXiv:2010.02502. https://doi.org/10.48550/ arXiv.2010.02502
- Rombach R., Blattmann A., Lorenz D., Esser P., Ommer B. High-Resolution image synthesis with latent diffusion models. *Proc. of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (CVPR), 2022, pp. 10674–10685. https://doi.org/10.1109/ CVPR52688.2022.01042
- Li H., Yang Y., Chang M., Chen S., Feng H., Xu Z., Li Q., Chen Y. SRDiff: Single Image Super-Resolution with diffusion probabilistic models. *Neurocomputing*, 2022, vol. 479, pp. 47–59. https://doi. org/10.1016/j.neucom.2022.01.029
- Yu F., Gu J., Li Z., Liu J., Kong X., Wang X., He J., Qiao Y., Dong C. Scaling Up to Excellence: practicing model scaling for photorealistic image restoration in the wild. *Proc. of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2024, pp. 25669–25680. https://doi.org/10.1109/ CVPR52733.2024.02425
- Zhang R., Isola P., Efros A.A., Shechtman E., Wang O. The unreasonable effectiveness of deep features as a perceptual metric. *Proc. of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018, pp. 586–595. https://doi.org/10.1109/ CVPR.2018.00068
- Wang J., Chan K.C.K., Loy C.C. Exploring CLIP for assessing the look and feel of images. *Proc. of the 37th AAAI Conference on Artificial Intelligence*, 2023, vol. 37, no. 2. pp. 2555–2563. https:// doi.org/10.1609/aaai.v37i2.25353
- Heusel M., Ramsauer H., Unterthiner T., Nessler B., Hochreiter S. GANs trained by a two time-scale update rule converge to a local nash equilibrium. Proc. of the 31st International Conference on Neural Information Processing Systems (NIPS '17), 2017, pp. 6629–6640.

Authors

Aleksei K. Denisov — Assistant, ITMO University, Saint Petersburg, 197101, Russian Federation, sc 57210698353, https://orcid.org/0000-0001-8135-1723, denisov@itmo.ru

Sergei V. Bykovskii — PhD, Associate Professor, ITMO University, Saint Petersburg, 197101, Russian Federation, sc 57216469537, https://orcid. org/0000-0003-4163-9743, sergei_bykovskii@itmo.ru

Pavel V. Kustarev — PhD, Dean, ITMO University, Saint Petersburg, 197101, Russian Federation, SC 35317916600, https://orcid.org/0000-0001-9326-0837, kustarev@itmo.ru

Received 26.11.2024 Approved after reviewing 05.02.2025 Accepted 19.03.2025



Работа доступна по лицензии Creative Commons «Attribution-NonCommercial» **I/İTMO**

НАУЧНО-ТЕХНИЧЕСКИЙ ВЕСТНИК ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ, МЕХАНИКИ И ОПТИКИ март–апрель 2025 Том 25 № 2 http://ntv.ifmo.ru/ SCIENTIFIC AND TECHNICAL JOURNAL OF INFORMATION TECHNOLOGIES, MECHANICS AND OPTICS March–April 2025 Vol. 25 № 2 http://ntv.ifmo.ru/en/ ISSN 2226-1494 (print) ISSN 2500-0373 (online)

ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ, МЕХАНИКИ И ОПТИКИ

:

doi: 10.17586/2226-1494-2025-25-2-328-338

Verification of the formal requirements for the system behavior based on automaton objects

Fedor A. Novikov¹, Irina V. Afanasieva², Ludmila N. Fedorchenko^{3⊠}, Taisia A. Kharisova⁴

¹ Peter the Great St. Petersburg Polytechnic University, Saint Petersburg, 195251, Russian Federation

² Special Astrophysical Observatory of the Russian Academy of Sciences (SAO RAS), Nizhny Arkhyz, 369167, Russian Federation

³ St. Petersburg Federal Research Center of the Russian Academy of Sciences, Saint Petersburg, 199178, Russian Federation

⁴ Ioffe Institute, Saint Petersburg, 194021, Russian Federation

¹ fedornovikov51@gmail.com, https://orcid.org/0000-0003-4450-0173

² riv615@gmail.com, https://orcid.org/0000-0003-4225-4124

³ lnf@iias.spb.su^{\overline{1}}, https://orcid.org/0000-0002-4008-9316

⁴ tais.harisova@mail.ru, https://orcid.org/0009-0002-3456-0471

Abstract

The paper studies the problems of automatic verification of the behavior of a reactive system in accordance with formalized requirements, i.e., automatic verification. The reactive system is described by interconnected automaton objects. Formalized requirements are written as conditioned regular expressions. In this case mathematically reliable verification of the system is possible. The proposed solution is based on the use of the CIAO (Cooperative Interaction of Automaton Objects) language to specify the interaction of automaton objects. This paper considers the third version of the language, CIAO v.3, which defines the means of describing automaton classes, the means of instantiating automaton objects, and linking these objects to the system using a link diagram. The requirements to be verified are specified using the so-called conditioned regular expressions constructed over a set of elementary actions and conditions defined in the system. A software tool has been developed that, using a link diagram, builds a semantic graph, i.e., a directed graph, where all paths from the initial nodes represent the execution protocols of the automata-based program, thereby specifying the semantics of the reactive system. Then, it is checked whether the automata-based program complies with the requirement defined by the conditioned regular expression. If a discrepancy is detected, the tool shows the place where the semantic graph exactly does not comply with the requirement. Algorithms have been developed that allow automatic verification of reactive systems with respect to the formalized requirements of a certain class. An example of a program is given that demonstrates elevator control in the CIAO v.3 language, and the constructed reactive system is verified for compliance with formally specified requirements. The purpose of the article is to demonstrate software implementation of the tool for automatic verification of a program in the CIAO v.3 language.

Keywords

automatic verification, conditioned regular expressions, behavior model, automata-based programming, state transition graph, UML, state machine diagram, concurrent behavior, software architecture, reactive system

For citation: Novikov F.A., Afanasieva I.V., Fedorchenko L.N., Kharisova T.A. Verification of the formal requirements for the system behavior based on automaton objects. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 2025, vol. 25, no. 2, pp. 328–338. doi: 10.17586/2226-1494-2025-25-2-328-338

© Novikov F.A., Afanasieva I.V., Fedorchenko L.N., Kharisova T.A., 2025

УДК 004.415.52, 004.434

Проверка соответствия поведения системы на основе автоматных объектов формальным требованиям

Федор Александрович Новиков¹, Ирина Викторовна Афанасьева², Людмила Николаевна Федорченко³⊠, Таисия Анваровна Харисова⁴

¹ Санкт-Петербургский политехнический университет Петра Великого, Санкт-Петербург, 195251, Российская Федерация

² Специальная астрофизическая обсерватория Российской академии наук, Нижний Архыз, 369167, Российская Федерация

³ Санкт-Петербургский Федеральный исследовательский центр Российской академии наук, Санкт-Петербург, 199178, Российская Федерация

⁴ Физико-технический институт им. А.Ф. Иоффе Российской академии наук, Санкт-Петербург, 194021, Российская Федерация

¹ fedornovikov51@gmail.com, https://orcid.org/0000-0003-4450-0173

² riv615@gmail.com, https://orcid.org/0000-0003-4225-4124

³ lnf@iias.spb.su^{\equiv}, https://orcid.org/0000-0002-4008-9316

⁴ tais.harisova@mail.ru, https://orcid.org/0009-0002-3456-0471

Аннотация

Введение. В работе исследованы проблемы автоматической проверки соответствия поведения реагирующей системы формализованным требованиям, а именно, автоматической верификации. Реагирующая система описывается в форме взаимосвязанных автоматных объектов. Формализованные требования записываются в виде обусловленных регулярных выражений. Показано, что в этом случае возможна математически достоверная верификация системы. Метод. Предложенное решение основано на применении языка спецификации взаимодействия автоматных объектов Cooperative Interaction of Automaton Objects (CIAO). В данной работе рассматривается третья версия языка, CIAO v.3, в котором определены средства описания автоматных классов, средства инстанциации автоматных объектов и связывания этих объектов в систему с помощью схемы связей. Верифицируемые требования заданы с помощью так называемых обусловленных регулярных выражений, построенных над множеством элементарных действий и условий, определенных в системе. Разработан программный инструмент, который, используя схему связей, строит семантический граф — ориентированный граф, в котором все пути из начальных узлов представляют протоколы выполнения автоматной программы, задавая тем самым семантику реагирующей системы. Выполнена проверка соответствия автоматной программы требованию, определяемому обусловленным регулярным выражением. В случае обнаружения несоответствия инструмент показывает место, где именно семантический граф не соответствует требованию. Основные результаты. Разработаны алгоритмы, позволяющие проводить автоматическую верификацию реагирующих систем относительно формализованных требований определенного класса. Рассмотрен пример программы, демонстрирующий управление работой лифта на языке CIAO v.3 и проведена верификация построенной реагирующей системы на соответствие формально заданным требованиям. Обсуждение. Продемонстрирована программная реализации инструмента автоматической верификации программы на языке CIAO v.3.

Ключевые слова

автоматическая верификация, обусловленные регулярные выражения, модель поведения, автоматное программирование, граф переходов состояний, унифицированный язык моделирования UML, диаграмма конечного автомата, параллельное поведение, архитектура программного обеспечения, реагирующая система

Ссылка для цитирования: Новиков Ф.А., Афанасьева И.В., Федорченко Л.Н., Харисова Т.А. Проверка соответствия поведения системы на основе автоматных объектов формальным требованиям // Научнотехнический вестник информационных технологий, механики и оптики. 2025. Т. 25, № 2. С. 328–338 (на англ. яз.). doi: 10.17586/2226-1494-2025-25-2-328-338

Introduction and relationship to previous studies

The article discusses a special case of the general concept of verification and defines an automaton object, in comparison with other cases of using the word "automaton".

The process of checking the compliance of the system behavior with the requirements imposed on it is referred to as *verification*¹. Typically, such verification is performed by conducting a finite number of tests, which is called *testing*. Testing cannot ensure absolute reliability because, for most complex software systems, the set of possible behavior scenarios is infinite, and exhaustive testing is impossible. Reliability is desirable in all cases; however, in some cases, unreliability is unacceptable. Software with unacceptable failures is usually referred to as *responsible* (life-critical). Among the various methods for ensuring the reliability of responsible systems, *formal* verification methods for software are more promising than traditional testing methods. Verification, as a check for compliance with requirements, allows for various explanations and interpretations [1]. If the requirements are formalized and the system behavior is specified algorithmically, *automatic* verification is possible. We share the concepts outlined in [2] and believe that responsible software requires the use

¹ GOST R ISO/IEC 25010-2015. Information technology. Systems and software engineering. Systems and software Quality Requirements and Evaluation (SQuaRE). System and software quality models. Moscow, Standartinform, 2015, 36 p. (in Russian)

of mathematical methods to verify software conformity to formal specifications.

Numerous formal verification methods coexist and are used that allow automation, and for the implementation of automatic verification; the choice of software design tools and software requirement specification tools is critically important.

In the most general case, when a general purpose programming language is used to construct programs, and a first-order predicate calculus language is used to define formal specifications, formal automatic verification of complex systems appears to be an insurmountably difficult or even practically insoluble task [3]. In addition, a remarkable study [4] presented examples of manual (non-automatic) verification of fairly complex algorithms. The constructions in this book are perfect in the sense that the postconditions considered there are complete, i.e., exhaustive specifications of programs. Therefore, programs verified in this way are indeed valid by construction. However, it is unlikely that Dijkstra's method can be fully automated at present — the "poles" of verification are too far apart: declarative specification and imperative program.

The next step towards verification automation was the Model Checking method. In this method, the requirement (specification) being verified is defined by a temporal logic formula, and it is not the program itself that is verified, but its model, the so-called Kripke automaton (Kripke structure) [5]. At present, this method is sufficiently advanced, supported by industrial tools and widely used, including in the discussed area of reactive systems [6]. Note that the success of the Model Checking method, in our opinion, is to some extent due to the convergence of the verification "poles": the Kripke structure is much more formalized than the program text in a programming language, and temporal logic formulas have significant expressive power when specifying processes that develop over time. The Model Checking method itself is impeccable, but two questions remain outside the scope of the method: how to ensure the completeness of specifications by means of temporal logic and how to ensure the adequacy of the program and its Kripke model.

The question of how to ensure the completeness of formal specifications remains open and nothing better than private ad hoc methods have been proposed so far. The question of the correspondence between the verified model and original program has received a satisfactory solution within the framework of the automata-based programming paradigm [7, 8]. The fact is that the automata-based programming paradigm, along with other advantages, has the property that the description of behavior in the form of interacting automata is unambiguously and formally translated, on the one hand, into general-purpose programming languages and, on the other hand, translated into the Kripke structure. Thus, if we begin developing a system by fixing the description of its behavior in the automata-based programming paradigm, the question of the adequacy of the model used and the program being verified is removed. Taking this circumstance into account, our constructions are based on the automata-based programming paradigm.

Another approach to ensuring the feasibility of automatic verification methods is to use a language (or more precisely, a family of languages) for constructing reactive systems, which we call Cooperative Interaction of Automata Objects (CIAO). The history of the language has undergone several stages of development. In particular, the first version of the CIAO language was used to create an automata-based method for determining domainspecific languages [9–11]. The second version of the CIAO language [12, 13] was used to implement a critical system for collecting and processing astrometric data [14]. In the process of using the language, a number of improvements were made, as a result of which it became possible to use it to verify low-level communication protocols [15] and to solve other problems [16]. The results obtained showed that the CIAO language can be successfully used for the automatic verification of device control programs [17, 18]. Subsequent research in this area led us to the need to clarify the behavior model that underlies the language and to improve the graphical notations.

As a result, the CIAO v.3 language [19] was developed for which automatic verification algorithms were implemented. Note that the CIAO v.3 language is designed in such a way that, based on a system of interacting automaton objects, it is possible to uniquely and efficiently (with linear complexity by the total number of transitions in the system) construct a semantic graph [17] in which all paths from the initial states define all possible system execution protocols. The so-called Conditioned Regular Expressions (CREs) [20] are used as a means of specifying formal specifications. In fact, CREs specify a pattern that must match the paths in the semantic graph. Checking that the specified elements in a given sequence occur in a certain relative order is a well-known task in syntactic analysis (pattern matching). Note that a sequence pattern is a weaker specification tool compared to temporal logic. Of course, specifying permissible sequences of actions is only a partial, not a complete description of system requirements. However, such requirements can be effectively verified automatically. Thus, it was possible to bring the "poles" of verification even closer together in comparison with the Model Checking method. The next important concept is the automaton object. We have taken this term from the lexicon of automata-based programming [19], but it should not be identified with the concept of an "automated object" [7]. Although the concept of an automaton object used here was developed on the basis of the concept of an automated object, the differences are significant, especially from the point of view of verification. There are three most important differences:

- firstly, automaton objects are endowed with internal memory in the form of local variables where it is possible to store information obtained in the parameters of events (input actions), calculate new information and transfer information to the outside in the parameters of effects (output actions). Thus, automaton objects fall out of the class of finite automata which have only finite internal memory;
- secondly, automaton objects are instances of automaton classes in the sense that objects are instances of classes in object-oriented programming. Automaton objects

receive the initial values of all variables, including the initial state during instantiation, and interact exclusively by accessing the provided interfaces. The fact that the internal behavior of an automaton object is specified by a state transition graph is not essential for verification and is not even mandatory. We prefer the automatonbased programming style described in the fundamental book [21], but we do not insist on its use;

 thirdly, the behavior of automaton objects is nondeterministic since they coexist in parallel, interact asynchronously, and are completely equal in rights. No central control automaton is assumed.

Cooperative interaction is the "zest" that ensures the feasibility of automatic verification in this case. Automaton objects are absolutely equal and interact with each other in accordance with the connection scheme [19] which connects them in the provided and required interfaces coordinated by types and kinds. An elementary act of interaction begins with the fact that in one automaton object (or in the external environment) an action (effect) is called which is transmitted through a connected pair of interfaces to another automaton object (or to the external environment) where the transmitted action is perceived as an event that triggers a transition. During the transition, the automaton object goes to another state (or remains in the same state) and, possibly, causes an effect. Calling an effect starts a new elementary interaction, and so on, the process continues. If there is no effect in the automaton object or if there is no transition in this state for the incoming event, then the event is lost and the process (thread) is terminated. In the proposed approach, the behavior of the system is a sequence (more precisely, a partially ordered set) of such elementary interactions. The proposed verification method allows us to check whether the behavior of the system corresponds to a given pattern or not.

Many researchers have considered the issues of verification of automata-based programs, which is reflected in various publications, from pioneering articles [22] to textbooks [23]. A natural question arises: why do we need another study and how does our work differ from the previous ones? A comparison is necessary. For comparison, we chose the article [24] in which the verification issue is covered as thoroughly and strictly as possible. In the approach [24], reactive systems defined by a hierarchy of finite automata are subject to verification. In the approach to verification that we are developing, the class of systems under consideration is the same: reactive systems. The methods for describing reactive systems differ somewhat in notation and the constructions used, but both methods are equally powerful since they are Turing complete [16]. The semantic graph that we introduced is also a state transition model, like the Kripke structure, but does not have a load in the form of temporal logic formulas. Conditioned regular expressions, which we use as a specification tool, are logically a weaker formalism than all temporal calculi. Thus, the approach we develop never goes beyond the limits outlined in the article [24]. The difference has another nature, and this difference is much more important than the differences in notation. The difference is in the motivation and goals of the study. In the case of verification of automata-based programs using the method [24], a

method is chosen (model checking), an object is specified (an automata-based program), and a study is conducted to determine what can be obtained from this object using this method. Since the article [24] obtained the maximum of what can, in principle, be obtained, we can conclude that the study has achieved its goal, the problem is solved, and further manipulations are inappropriate. In our case, everything is the other way around. The goal is to achieve the ability to automatically guarantee error-free behavior of the system, the object is pointwise designated (a set of examples of reactive systems), and it is necessary to find a minimum of means that ensure the achievement of the goal. Note that the side effect of the minimality of the means used is, as a rule (in this case, this is true), the efficiency of the resulting algorithms and the scalability of the approach. In such a formulation of the problem, the study should not be declared complete, since new examples and new interpretations of infallibility appear in life — there is no limit to perfection.

This study demonstrated the capability of automatic verification of conformity to formal program specifications using the CIAO v.3 language.

Using the CIAO v.3 specification language

Let us consider an example of constructing a behavior specification for a fragment of an elevator control system which we started to consider in the article [16] and continued in the articles [17–20]. This example demonstrates the expressive power of the CIAO v.3 language and the benefits of introducing the concept of automata classes and a connection scheme. The fragment of the system under consideration consists of a control device and several actuators: a cabin, doors, and lighting. Technical details are omitted; our focus is on describing the interactions between the system components.

The behavior of the control device (the **Controller** class) is shown in Fig. 1, on the left.

Let us assume that there are four possible types of influence from the external environment on the system:

- 1. The user may want to use the system (press the door handle if the door has a handle, or press the door opening button, or show a QR code, etc). The system processes any such event as a provided command (in event).
- 2. The user may enter the elevator cabin if the doors are open and the light is on (enter event).
- 3. The user may want to stop using the system (press the door handle or press the door opening button from the inside, etc.). Such an event is designated as **out**.
- 4. The user may exit the elevator via open doors (exit event).

The actions performed during transitions from one state to another determine the behavior of the system. The reasons for transitions in the CIAO v.3 language are limited to two cases: a transition by a call event which is designated by calling the provided interface with the "command" stereotype, and a transition by a timer event which is designated by the **after** keyword (Fig. 1). We do not use changing or signaling events that are provided for in UML. In addition, the CIAO v.3 language does not use unmotivated (spontaneous) transitions upon completion [21].



Fig. 1. Example description of an elevator control system fragment in the CIAO v.3 language

From the perspective of interaction description, it can be seen that the behavior of the actuators is very similar. In other words, each device is either 'on' and performs a function or is 'off' and does not perform a function. In addition, all devices can report what state they are 'in' at the moment: 'on' or 'off'. Therefore, in the CIAO v.3 language, it is sufficient to describe one automata class (the **Device** class, Fig. 1) which implements the behavior of the trigger [22], and then use three objects (the **door**, **light**, and **cabin** objects) in the connection scheme, linking their provided **on** and **off** interfaces with the required interfaces of the control system, as shown in Fig. 1.

In this example, we demonstrate the convenience and flexibility of the proposed connection scheme in the CIAO v.3 language. The proposed system involves only one instance of the **Controller** class and three instances of the **Device** class (Fig. 1). In such cases, the CIAO v.3 language allows us to treat a class with one instance as a "singleton" class [17], and not explicitly create a separate instance by specifying the instantiation parameters directly in the class description (initial state **Idle**).

As stated in a previous article [19], the CIAO v.3 language exists in two forms: graphical (shown above in Fig. 1) and textual (Listing 1). When using the developed software tools for automatic verification, the textual form is still used, and the graphical form is an illustration of human perception.

Listing 1. Description of the elevator control system in the CIAO v.3 language

```
01 ciao Elevator // program Elevator
in CIAO language
02 class Controller // description
of the Controller class
03 events // provided commands
04 in
```

```
05
         out
06
         enter
07
         exit
08
       effects // required commands
09
         open door
10
         close door
11
         switch on
12
         switch off
13
         occupy
14
         free
15
       conditions // required queries
16
         occupied
17
       states
         Idle -> in[occupied] -> Idle
18
[else] / switch on; open door -> I2B
         I2B -> after(t1) / close
19
door; switch off -> Idle
20
         I2B -> enter / close door;
occupy -> Busy
21
         Busy -> out / open door ->
B2I
22
         B2I -> after(t2) / close door
-> Busy
23
         B2I -> exit / close door;
switch off; free -> Idle
     class Device // description of
24
the Device class
25
       events
26
         off
27
         on
28
       assertions // provided queries
29
         is0n
30
       states // states and
transitions
31
         Off -> on -> On
32
         Off -> off -> Off
33
         On -> on -> On
```

Научно-технический вестник информационных технологий, механики и оптики, 2025, том 25, № 2 Scientific and Technical Journal of Information Technologies, Mechanics and Optics, 2025, vol. 25, no 2

```
34
         On -> off -> Off
35
     scheme // connection scheme
36
       objects // instantiation of
automaton objects
37
         controller = new
Controller (Idle)
         door = new Device(Off)
38
         light = new Device(Off)
39
40
         cabin = new Device(Off)
       links // linked interfaces
41
42
         controller.open door <- door.
on
43
         controller.close door <-
door.off
44
         controller.switch on <-
light.on
45
         controller.switch off <-</pre>
light.off
46
         controller.occupy <- cabin.on</pre>
47
         controller.free <- cabin.off</pre>
48
         controller.occupied <- cabin.
isOn
       public // public interfaces
49
50
         controller.in
51
         controller.out
52
         controller.enter
53
         controller.exit
54 . // end of program
```

The example presented in this paper covers only a small part of elevator control problems. In particular, the reliability of the elevator control system can be increased by providing for the handling of exceptional situations or failures. For example, it is easy to make the doors not open if the light fails to turn on (the bulb burns out), and instead escalate the malfunction to the appropriate level (for example, send an emergency signal to the dispatcher of the building in which the elevator is installed). For simplicity, we exclude all such subtleties. The purpose of this example was to demonstrate the sufficiency of the CIAO v.3 language tools in the automatic verification of a reactive system.

Construction of a semantic graph based on the connection scheme

Having a connection scheme, it is possible to create a *semantic graph*, i.e., a directed graph in which all paths from the initial nodes represent all possible protocols for executing an automaton program; thus, it defines semantics.

There are various ways to create a semantic graph. In particular, in [17, 18], we placed the actions to be performed in the graph nodes and the conditions to be checked on the arcs. The software implementation showed that it is more convenient to place actions with the conditions on graph arcs, with one action or condition per arc. This slightly increases the number of nodes but simplifies the machineprocessing algorithm of the graph.

In such agreements, the algorithm for constructing a semantic graph identifies initial events based on the analysis

of the connection scheme, i.e., events whose arrival from an outside world can trigger the operation of the system, and then tracing all possible paths in the state transition graphs of automaton objects, using the connection diagram to move from one state transition graph to another. In fact, constructing a semantic graph is a *symbolic execution* of an automaton program [22–27].

In this case, the analysis of the connection diagram yields an unambiguous result. The system has four external interfaces (**in**, **out**, **enter**, **exit**, the remaining interfaces are connected), and of these, only the **in** event can be processed in the initial state; thus, there is a single initial node (it is indicated by the black circle with an arrow in Fig. 2, a, bottom), from which a transition is possible by the **in** event.

Next, in accordance with the state transition graph (Fig. 2, *a*, top), the guard condition [**occupied**] is checked. If the elevator is occupied, then service is impossible and a return to the initial state occurs; otherwise, actions **switch** on and open door are performed (Fig. 2, *a*, bottom).

Then, two possible scenarios are possible: either the passenger enters the elevator cabin (enter event) or the passenger does not dare to do so. Then, after time t1, the doors are closed, the light is released, and the system returns to the initial state. All other transitions in Fig. 1 are processed similarly, and the semantic graph is built automatically by tracing the paths in the state transition graphs and moving from one automaton object to another according to the connection scheme. The final result is shown in Fig. 2, b.

Conditioned regular expressions as requirement formalization

All possible sequences of actions performed during the operation of a reactive system determine the overall behavior of the entire system. Each sequence of actions is called an execution occurrence protocol. In this case, protocols may include both acceptable sequences satisfying the system requirements and unacceptable sequences that must be classified as erroneous. Our view is that formalized requirements (specifications) must be provided in the form of a description of acceptable sequences of events/actions. Regular expressions with guard conditions (we refer to them as CREs) can be used to mathematically describe such specifications.

Various versions of regular expression languages are currently in use and co-exist [28]. We used the regular expression language with the operations of union (notation A|B), concatenation (notation AB), and Tseitin iteration (notation $\{A\#B\}$) [29]. In addition to the binary operation of Tseitin iteration, defined by the Kleene iteration as $A\#B = A(BA)^*$, it is also possible to use the unary Kleene operation where the expression A# is equivalent to the expression AA^* , and the expression #A is equivalent to the expression A^* . Here A and B are arbitrary regular expressions.

A special feature of the CRE is the composition of the elementary languages used. Firstly, these are the names of the events and effects that appear on transitions in the semantic graph (Fig. 2, b). Secondly, it is the guard conditions which are written in square brackets. Thirdly, it is the symbol 'A' (from the word any) which denotes any



Fig. 2. Fragment of the transformation of the program state graph into a semantic graph (*a*) and a complete semantic graph constructed automatically (*b*)

event, effect, or guard condition other than those explicitly used in the given CRE.

The syntax rules of the CRE language are written in the same conventions as those used in the article [19]. The terminal symbols in the rules are underlined. Metasymbols and nonterminals are highlighted in bold. CRE syntax diagrams are presented in the Table.

CRE : (SYMBOL | UNION | ITERATION) # ,. UNION : (CRE #]). ITERATION : $\{ \# CRE \} | \{ CRE \# \} | \{ CRE \# CRE \} .$ SYMBOL : name | CONDITION | A . CONDITION : [! name | name] . A : \underline{A} . For example, CRE

 $\{\#A\}$, switch_on, $\{\#A\}$, open_door, $\{\#A\}$, close_door, $\{\#A\}$, switch_off, $\{\#A\}$

describes the set of all sequences in which actions **switch_on**, **open_door**, **close_door**, and **switch_off** occur exactly once and strictly in the specified order.

Verification of the Automaton Objects Behavior

Verification, i.e., checking that the behavior of an automaton program corresponds to formal requirements, is performed automatically. However, the formalization of requirements, i.e., composing conditioned regular

CRE SYMBOL SYMBOL CONDITION CONDITION ITERATION ITERATION UNION 비극 **● #** } - CRE **₩**} H (CRE CRE CRE

Table. CRE syntax diagram

Научно-технический вестник информационных технологий, механики и оптики, 2025, том 25, № 2 Scientific and Technical Journal of Information Technologies, Mechanics and Optics, 2025, vol. 25, no 2 expressions describing acceptable sequences of actions and statements, is not performed automatically. This is a creative, trial-and-error method.

Here, we provide an example illustrating the capabilities of the proposed verification method. Let the elevator control system satisfy the natural informal requirement given below.

Informal requirement. When a passenger enters the elevator, i. e., he/she is in the elevator cabin, and exits the elevator, the light *must* be switched on. At all other times the light *may* be switched off.

This requirement must be *formalized*, i.e., the permissible sequences of actions performed to turn on and off the light, open, and close the doors must be written down as a regular expression. As an example, we provide a possible scenario for formalizing the requirement using the "successive approximation" method.

The first approximation is as follows: to enter the elevator and exit the elevator, the doors must be open. This is how the first version of formalization is obtained.

Formal requirement (option 1). The actions of turning on/off the light and opening/closing the doors must be performed once in the established order.

This requirement is input to the verification program as expressed in the sequence

{#*A*}, switch_on, {#*A*}, open_door, {#*A*}, close_ door, {#*A*}, switch_off, {#*A*} **Result:** False, i.e., the requirement is not met (Fig. 3, *a*).

The door-opening action **open_door** is encountered after the door-closing action **close_door**. The transition along the arc **open_door** and all subsequent transitions are considered erroneous and are marked red in the semantic graph; the transition along the arc **switch_off** also does not satisfy the following requirement: the arc closes the cycle, and in this requirement all actions are performed once outside the cycle.

Second approximation: it is necessary to consider the cyclic nature of the process.

Formal requirement (option 2). The actions of turning on/off the light and opening/closing the doors must be performed cyclically in the established order:

Result: False (Fig. 3, b). The door-opening action **open_door** is encountered after the door-closing action **close_door**; the transition along the arc **open_door** and all subsequent transitions are considered erroneous and are marked red in the semantic graph.

In the third attempt to formalize **formal requirement** (**option 1**), the shortcomings of the first two were eliminated.

Formal requirement (option 3). The actions of turning on/off the light and opening/closing the doors must



Fig. 3. Verification result of requirement: option 1 (a) and option 2 (b)

be performed cyclically in the established order, while between the actions of turning on/off the light there may be an arbitrary number of actions of opening/closing the doors (but not less than one).

```
{#({#A },switch_on,{#A },{({#A },open_door,{#A },
close door,{#A })#},{#A },switch off,{#A })}
```

Result: True.

In other words, the fulfillment of the formal requirement is verified: all sequences defined by the semantic graph satisfy the regular expression.

How adequate is the formalization? This is a legitimate and a difficult question. Informal inductive reasoning shows that if a passenger was outside the elevator cabin at the initial moment, he/she would enter the elevator, stay in the cabin, and exit the elevator only when the light is on, and the immutable requirement "the light must be on" is fulfilled.

A passenger may act in bad faith: call the elevator but do not enter it and leave for another business. In this case, the light is turned on for some time (namely, **t1**) in vain. After the timeout, the light is turned off; therefore, the optional requirement "the light may be turned off" is fulfilled.

Discussion and prospects for further research

The proposed automatic verification method was tested on several examples and demonstrated good applicability. In future research, we plan to expand the scope of the proposed method and the CIAO v.3 language in at least the following five directions.

- 1. The concept of a connection scheme between interacting automaton objects allows for both statically and *dynamic* interactions between automaton objects. This capability is especially important for reactive systems whose behavior changes according to environmental conditions.
- 2. Automaton objects allow for unlimited structural *decomposition* while maintaining object-oriented encapsulation. The concept of private interfaces allows us to detail both the connection scheme and

References

- 1. Kulyamin V.V. *Software Verification Methods*. Moscow, ISP RAS, 2008, 111 p. (in Russian)
- Shalyto A.A. Validation of state machine specifications. Scientific and Technical Journal of Information Technologies, Mechanics and Optics, 2023, vol. 23, no. 2, pp. 436–438. (in Russian). https://doi. org/10.17586/2226-1494-2023-23-2-436-438
- Lavrov S.S. Programming. Mathematical Foundations, Tools, Theory. St. Petersburg, BHV Petersburg Publ., 2001, 320 p. (in Russian)
- 4. Dijkstra E.W. A Discipline of Programming. Prentice Hall, 1976, 217 p.
- 5. Karpov Iu.G. *Model Checking. Verification of Parallel and Distributed Software Systems.* St. Petersburg, BHV-Petersburg, 2010, 560 p. (in Russian)
- Vinogradov R.A., Kuz'min E.V., Sokolov V.A. Verification of automata programs using CPN/Tools. *Modelirovanie i analiz informacionnyh* sistem, 2006, vol. 13, no. 2, pp. 4–15. (in Russian)
- Shalyto A. Automata-based programming paradigm. Scientific and Technical Journal of Information Technologies, Mechanics and Optics, 2008, vol. 53, pp. 3–23. (in Russian)

the formal specification of behavior at nested levels. This capability is necessary for verifying complex reactive systems which often consist of many relatively independent subsystems.

- 3. The provided queries can be generalized so that they deliver results not only of Boolean types, but also of any other enumerated types. In terms of conventional programming languages, this means using not only branching by condition (the "if" operator) but also branching by value (the "switch" operator). This feature is very useful in various discrete control systems. It is so useful that it carries the main semantic load of the book [21].
- 4. The capabilities of the link diagram can be expanded so that the provided command interface of one automaton object can be linked to the required command interfaces of several other automaton objects at once. This feature allows us to exploit the competitive advantages of the client-server architecture within the framework of automaton programming.
- 5. It is also possible to link the required command interface of a single automaton object with the provided command interfaces of other automaton objects. This implies the concurrent launch of several interacting processes. It is difficult to overestimate this feature if the execution of the responsive system is assumed to be on multiprocessor hardware.

These priority improvements listed above do not exhaust the potential capabilities of the proposed model. The provided interfaces can be linked with the required interfaces of the same state machine, and the automaton object sends events to itself. Note that the capabilities of the model are not limited. The proposed behavior model is essentially asynchronous, parallel, and nondeterministic. It is quite easy to obtain a conventional sequential deterministic computation scheme: simply pull the automaton objects into a chain, linking the required interfaces of the previous object with the provided interfaces of the next object. However, if sequential execution is not required, the proposed model enables the construction of natural parallel schemes that are effective on modern platforms.

Литература

- Кулямин В.В. Методы верификации программного обеспечения. М.: ИСП РАН, 2008. 111 с.
- 2. Шалыто А.А. Валидация автоматных спецификаций // Научнотехнический вестник информационных технологий, механики и оптики. 2023. Т. 23. № 2. С. 436–438. https://doi.org/10.17586/2226-1494-2023-23-2-436-438
- Лавров С.С. Программирование. Математические основы, средства, теория. СПб.: БХВ-Петербург, 2001. 320 с.
- Dijkstra E.W. A Discipline of Programming. Prentice Hall, 1976. 217 р.
 Карпов Ю.Г. Model Checking. Верификация параллельных и рас-
- пределенных систем. СПб.: БХВ-Петербург, 2010. 560 с.
- Виноградов Р.А., Кузьмин Е.В., Соколов В.А. Верификация автоматных программ средствами CPN/Tools // Моделирование и анализ информационных систем. 2006. Т. 13. № 2. С. 4–15.
- Шалыто А.А. Парадигма автоматного программирования // Научно-технический вестник Санкт-Петербургского государственного университета информационных технологий, механики и оптики. 2008. № 53. С. 3–23.

- Polikarpova N.I., Shalyto A.A. Automata-Based Programming. St. Petersburg, Piter Publ., 2011, 176 p. (in Russian)
- Novikov F.A., Tikhonova U.N. An Automata Based Method for Domain Specific Languages Definition. Part 1. *Information and Control Systems*, 2009, no. 6 (43), pp. 34–40. (in Russian)
- Novikov F.A., Tikhonova U.N. An Automata Based Method for Domain Specific Languages Definition. Part 2. *Information and Control Systems*, 2010, no. 2 (45), pp. 31–37. (in Russian)
- Novikov F.A., Tikhonova U.N. An Automata Based Method for Domain Specific Languages Definition. Part 3. *Information and Control Systems*, 2010, no. 3 (46), pp. 29–37. (in Russian)
- Novikov F., Fedorchenko L., Vorobiev V., Fatkieva R., Levonevskiy D. Attribute-based approach of defining the secure behavior of automata objects. *Proc. of the 10th International Conference on Security of Information and Networks (SIN'17)*, 2017, pp. 67–72. https://doi.org/10.1145/3136825.3136887
- Novikov F.A., Afanasyeva I.V. Cooperative interaction of automata objects. *Information and control systems*, 2016, no. 6 (85), pp. 50–64. (in Russian). https://doi.org/10.15217/issn1684-8853.2016.6.50
- Afanasieva I.V. Data acquisition and control system for highperformance large-area CCD Systems. *Astrophysical Bulletin*, 2015, vol. 70, no. 2, pp. 232–237. https://doi.org/10.1134/ S1990341315020108
- Levonevskiy D., Novikov F., Fedorchenko L., Afanasieva I. Verification of internet protocol properties using cooperating automaton objects. *Proc. of the 12th International Conference on Security of Information and Networks (SIN'19)*, 2019, pp. 1–4. https:// doi.org/10.1145/3357613.3357639
- Afanasyeva I.V., Novikov F.A., Fedorchenko L.N. Methodology for constructing event-driven software systems using the CIAO specification language. *SPIIRAS Proceedings*, 2020, vol. 19, no. 3, pp. 481–514. (in Russian). https://doi.org/10.15622/sp.2020.19.3.1
- Afanasieva I.V., Novikov F.A., Fedorchenko L.N. Verification of event-driven software systems using the specification language of cooperating automata objects. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 2023, vol. 23, no. 4, pp. 750–756. https://doi.org/10.17586/2226-1494-2023-23-4-750-756
- Novikov F.A., Afanasyeva I.V., Fedorchenko L.N., Kharisova T.A. Application of conditional regular expressions in problems of verification of control automata programs. 14th All-Russian conference on control problems. Collection of Scientific Papers. Moscow, IPU RAS, 2024, pp. 2651–2655. (in Russian)
- Novikov F.A., Afanasieva I.V., Fedorchenko L.N., Kharisova T.A. Specification language for automatabased objects cooperation. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 2024, vol. 24, no. 6, pp. 1035–1043. https:// doi.org/10.17586/2226-1494-2024-24-6-1035-1043
- Novikov F.A., Afanasyeva I.V., Fedorchenko L.N., Kharisova T.A. Application of Conditional Regular Expressions in Verification Problems of Control Automata Programs. *Proc. of Materials of the All-Russian Scientific Conference "PhysMech Science Week"*, St. Petersburg, POLYTECH-PRESS, 2024, pp. 167–170. (in Russian)
- Novikov F.A., Ivanov D.Iu. UML Modeling. Theory, Practice, Video Course. St. Petersburg, Professional'naja literature Publ., 2010, 640 p. (in Russian)
- Shalyto A.A. Switch-Technology. Algorithmization and Programming of the Logical Control Problems. St. Petersburg, Nauka Publ., 1998, 617 p. (in Russian)
- Velder S.E., Shalyto A.A. Methods of verification of models of automata-based programs. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 2008, vol. 53, pp. 123–137. (in Russian)
- Velder S.E., Lukin M.A., Shalyto A.A., Iaminov B.R. Verification of Automata-Based Programs. St. Petersburg, Nauka Publ., 2011, 244 p. (in Russian)
- Kuzmin E.V., Sokolov V.A. Modeling, specification, and verification of automaton programs. *Programming and Computer Software*, 2008, vol. 34, no. 1, pp. 27–43. https://doi.org/10.1134/s0361768808010040
- Gamma E., Helm R., Johnson R., Vlissides J. Design Patterns: Elements of Reusable Object-Oriented Software. Pearson Education, 1994, 416 p.
- Gerasimov A.Y. Directed dynamic symbolic execution for static analysis warnings confirmation. *Programming and Computer Software*, 2018, vol. 44, no. 5, pp. 316–323. https://doi.org/10.1134/ S036176881805002X

- Поликарпова Н.И., Шалыто А.А. Автоматное программирование. СПб.: Питер, 2010. 176 с.
- 9. Новиков Ф.А., Тихонова У.Н. Автоматный метод определения проблемно-ориентированных языков. Ч. 1 // Информационноуправляющие системы. 2009. № 6 (43). С. 34–40.
- Новиков Ф.А., Тихонова У.Н. Автоматный метод определения проблемно-ориентированных языков. Ч. 2 // Информационноуправляющие системы. 2010. № 2 (45). С. 31–37.
- Новиков Ф.А., Тихонова У.Н. Автоматный метод определения проблемно-ориентированных языков. Ч. 3 // Информационноуправляющие системы. 2010. № 3(46). С. 29–37.
- Novikov F., Fedorchenko L., Vorobiev V., Fatkieva R., Levonevskiy D. Attribute-based approach of defining the secure behavior of automata objects // Proc. of the 10th International Conference on Security of Information and Networks (SIN'17). 2017. P. 67–72. https://doi.org/10.1145/3136825.3136887
- Новиков Ф.А., Афанасьева И.В. Кооперативное взаимодействие автоматных объектов // Информационно-управляющие системы. 2016. № 6 (85). С. 50–64. https://doi.org/10.15217/issn1684-8853.2016.6.50
- Afanasieva I.V. Data acquisition and control system for highperformance large-area CCD Systems // Astrophysical Bulletin. 2015.
 V. 70. N 2. P. 232–237. https://doi.org/10.1134/S1990341315020108
- Levonevskiy D., Novikov F., Fedorchenko L., Afanasieva I. Verification of internet protocol properties using cooperating automaton objects // Proc. of the 12th International Conference on Security of Information and Networks (SIN'19). 2019. P. 1–4. https:// doi.org/10.1145/3357613.3357639
- Афанасьева И.В., Новиков Ф.А., Федорченко Л.Н. Методика построения событийно-управляемых программных систем с использованием языка спецификации СІАО // Труды СПИИРАН. 2020. Т. 19. № 3. С. 481–514. https://doi.org/10.15622/sp.2020.19.3.1
- Afanasieva I.V., Novikov F.A., Fedorchenko L.N. Verification of event-driven software systems using the specification language of cooperating automata objects // Scientific and Technical Journal of Information Technologies, Mechanics and Optics. 2023. V. 23. N 4. P. 750–756. https://doi.org/10.17586/2226-1494-2023-23-4-750-756
- Новиков Ф.А., Афанасьева И.В., Федорченко Л.Н., Харисова Т.А. Применение условных регулярных выражений в задачах верификации управляющих автоматных программ. Сборник научных трудов XIV Всероссийского совещания по проблемам управления. М.: ИПУ РАН, 2024. С. 2651–2655.
- Novikov F.A., Afanasieva I.V., Fedorchenko L.N., Kharisova T.A. Specification language for automatabased objects cooperation // Scientific and Technical Journal of Information Technologies, Mechanics and Optics. 2024. V. 24. N 6. P. 1035–1043. https://doi. org/10.17586/2226-1494-2024-24-6-1035-1043
- Новиков Ф.А., Афанасьева И.В., Федорченко Л.Н., Харисова Т.А. Применение условных регулярных выражений в задачах верификации управляющих автоматных программ. Сборник материалов Всероссийской научной конференции "Неделя науки ФизМех". СПб.: ПОЛИТЕХ-ПРЕСС, 2024. С. 167–170.
- Новиков Ф.А., Иванов Д.Ю. Моделирование на UML. Теория, практика, видеокурс. СПб.: Профессиональная литература, 2010. 640 с.
- Шалыто А.А. Switch-технология. Алгоритмизация и программирование задач логического управления. СПб.: Наука, 1998. 617 с.
- Вельдер С.Э., Шалыто А.А. Методы верификации моделей автоматных программ // Научно-технический вестник Санкт-Петербургского государственного университета информационных технологий, механики и оптики. 2008. № 53. С. 123–137.
- 24. Вельдер С.Э., Лукин М.А., Шалыто А.А., Яминов Б.Р. Верификация автоматных программ. СПб.: Наука, 2011. 244 с.
- Кузьмин Е.В., Соколов В.А. Моделирование, спецификация и верификация «автоматных» программ // Программирование. 2008. Т. 34. № 1. С. 38–60.
- Гамма Э., Хелм Р., Джонсон Р., Влиссидес Д. Паттерны объектно-ориентированного проектирования. СПб.: Питер, 2020. 448 с.
- 27. Герасимов А.Ю. Направленное динамическое символьное исполнение программ для подтверждения ошибок в программах // Программирование. 2018. № 5. С. 31–42. https://doi.org/10.31857/ S013234740001213-9
- Friedl J.E.F. Mastering Regular Expressions: Understand Your Data and Be More. O'Reilly Media, Inc., 2006. 544 p.
- 29. Fedorchenko L., Baranov S. Equivalent transformations and regularization in context-free grammars // Cybernetics and

- Friedl J.E.F. Mastering Regular Expressions: Understand Your Data and Be More. O'Reilly Media, Inc., 2006, 544 p.
- Fedorchenko L., Baranov S. Equivalent transformations and regularization in context-free grammars. *Cybernetics and Information Technologies*, 2015, vol. 14, no. 4, pp. 29–44. https://doi.org/10.1515/ cait-2014-0003

Authors

Fedor A. Novikov — D.Sc., Senior Researcher, Professor, Peter the Great St. Petersburg Polytechnic University (SPbPU), Saint Petersburg, 195251, Russian Federation; sc 16441904500, https://orcid.org/0000-0003-4450-0173, fedornovikov51@gmail.com

Irina V. Afanasieva — PhD, Head of Laboratory, Special Astrophysical Observatory of the Russian Academy of Sciences (SAO RAS), Nizhny Arkhyz, 369167, Russian Federation, sc 57210431774, https://orcid. org/0000-0003-4225-4124, riv615@gmail.com

Ludmila N. Fedorchenko — PhD, Senior Researcher, St. Petersburg Federal Research Center of the Russian Academy of Sciences, Saint Petersburg, 199178, Russian Federation, sc 36561350100, https://orcid. org/0000-0002-4008-9316, lnf@iias.spb.su

Taisia A. Kharisova — Engineer, Ioffe Institute, Saint Petersburg, 194021, Russian Federation, https://orcid.org/0009-0002-3456-0471, tais.harisova@mail.ru

Information Technologies. 2015. V. 14. N 4. P. 29–44. https://doi.org/10.1515/cait-2014-0003

Авторы

Новиков Федор Александрович — доктор технических наук, старший научный сотрудник, профессор, Санкт-Петербургский политехнический университет Петра Великого, Санкт-Петербург, 195251, Российская Федерация, sc 16441904500, https://orcid.org/0000-0003-4450-0173, fedornovikov51@gmail.com

Афанасьева Ирина Викторовна — кандидат технических наук, заведующий лабораторией, Специальная астрофизическая обсерватория Российской академии наук, Нижний Архыз, 369167, Российская Федерация, SC 57210431774, https://orcid.org/0000-0003-4225-4124, riv615@gmail.com

Федорченко Людмила Николаевна — кандидат технических наук, старший научный сотрудник, Санкт-Петербургский Федеральный исследовательский центр Российской академии наук, Санкт-Петербург, 199178, Российская Федерация, se 36561350100, https://orcid.org/0000-0002-4008-9316, lnf@iias.spb.su

Харисова Таисия Анваровна — инженер, Физико-технический институт им. А.Ф. Иоффе Российской академии наук, Санкт-Петербург, 194021, Российская Федерация, https://orcid.org/0009-0002-3456-0471, tais.harisova@mail.ru

Received 25.12.2024 Approved after reviewing 27.02.2025 Accepted 30.03.2025 Статья поступила в редакцию 25.12.2024 Одобрена после рецензирования 27.02.2025 Принята к печати 30.03.2025



Работа доступна по лицензии Creative Commons «Attribution-NonCommercial» **I/İTMO**

НАУЧНО-ТЕХНИЧЕСКИЙ ВЕСТНИК ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ, МЕХАНИКИ И ОПТИКИ март–апрель 2025 Том 25 № 2 http://ntv.ifmo.ru/ SCIENTIFIC AND TECHNICAL JOURNAL OF INFORMATION TECHNOLOGIES, MECHANICS AND OPTICS March–April 2025 Vol. 25 № 2 http://ntv.ifmo.ru/en/ ISSN 2226-1494 (print) ISSN 2500-0373 (online)

ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ, МЕХАНИКИ И ОПТИК

doi: 10.17586/2226-1494-2025-25-2-339-344

Vector search using method of clustering using ensemble of oblivious trees Nikita A. Tomilov¹, Vladimir P. Turov^{2⊠}, Alexander A. Babayants³, Alexev V. Platonov⁴

^{1,2,3,4} ITMO University, Saint Petersburg, 197101, Russian Federation

¹ programmer174@icloud.com, https://orcid.org/0000-0001-9325-0356

² firemoon@icloud.com^{\Box}, https://orcid.org/0009-0009-1470-7633

³ babayants.alexander@gmail.com, https://orcid.org/0009-0004-6367-6398

⁴ avplatonov@itmo.ru, https://orcid.org/0000-0002-8485-1296

Abstract

Information retrieval using machine learning algorithms is based on transforming the original multimodal documents into vector representations. These vectors are then indexed, and the search is performed within this index. A popular method for indexing is vector clustering such as with *k*-nearest neighbors. We propose a clustering method based on an ensemble of Oblivious Decision Trees and introduce a vector search algorithm built on this method. The proposed clustering method is deterministic and supports parameter serialization for the ensemble. The essence of the method involves training an ensemble of binary or ternary Oblivious Trees. This ensemble is then used to compute a hash for each of the original vectors. Vectors with the same hash are considered to belong to the same cluster. For searching, several clusters are selected whose centroids are closest to the vector representation of the search query followed by a full search of the vector representations within the selected clusters. The proposed method demonstrates search quality comparable to widely used industry-standard vector search libraries, such as Faiss, Annoy, and HNSWlib. For datasets with an angular distance metric, the proposed search method achieves accuracy equal to or better than existing solutions. For datasets with a Euclidean distance metric, the search quality is on par with existing solutions. The developed method can be applied to improve search quality in the development of multimodal search systems. The ability to serialize enables clustering data on one computational node and transferring ensemble parameters to another, allowing the proposed algorithm to be utilized in distributed systems.

Keywords

vector representations, vector search, embeddings, oblivious decision tree

For citation: Tomilov N.A., Turov V.P., Babayants A.A., Platonov A.V. Vector search using method of clustering using ensemble of oblivious trees. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 2025, vol. 25, no. 2, pp. 339–344. doi: 10.17586/2226-1494-2025-25-2-339-344

УДК 004.021

Векторный поиск методом кластеризации с помощью ансамбля небрежных решающих деревьев Никита Андреевич Томилов¹, Владимир Павлович Туров²,

Александр Амаякович Бабаянц³, Алексей Владимирович Платонов⁴

1,2,3,4 Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация

¹ programmer174@icloud.com, https://orcid.org/0000-0001-9325-0356

² firemoon@icloud.com^{\Box}, https://orcid.org/0009-0009-1470-7633

³ babayants.alexander@gmail.com, https://orcid.org/0009-0004-6367-6398

⁴ avplatonov@itmo.ru, https://orcid.org/0000-0002-8485-1296

Аннотация

Введение. Информационный поиск с использованием алгоритмов машинного обучения основывается на преобразовании исходных мультимодальных документов в векторные представления, далее строится индекс векторных представлений и производится поиск внутри индекса. Популярным способом построения индекса

© Tomilov N.A., Turov V.P., Babayants A.A., Platonov A.V., 2025

является кластеризация векторных представлений, например с помощью k-ближайших соседей. В работе предложен метод кластеризации с помощью ансамбля небрежных решающих деревьев, а также алгоритм векторного поиска на основе этого метода. Разработанный метод кластеризации является детерминированным и предоставляет возможность сериализации параметров ансамбля. Метод. Сущность метода состоит в обучении ансамбля двоичных или троичных небрежных решающих деревьев. Этот ансамбль используется для вычисления хэша для каждого из исходных векторных представлений. Векторные представления, имеющие одинаковый хэш, считаются принадлежащими к одному кластеру. Для поиска выбирается несколько кластеров, центроиды которых наиболее приближены к векторному представлению поискового запроса, после чего производится полный перебор векторных представлений выбранных кластеров. Основные результаты. Представленный метод показывает качество поиска, сравнимое с широко используемыми в индустрии библиотеками векторного поиска Faiss, Annoy и HNSWlib. Для протестированного набора данных с евклидовой метрикой расстояния предложенный метод поиска медленнее, чем существующие решения, но для протестированного набора данных с угловой метрикой расстояния результат сравним или лучше. Обсуждение. Разработанный метод может быть применен для улучшения качества поиска при создании мультимодальных поисковых систем. Возможность сериализации позволяет кластеризовать данные на одном вычислительном узле и передавать параметры ансамбля на другой вычислительный узел, что дает возможность использовать предложенный алгоритм в распределенных системах.

Ключевые слова

векторные представления, векторный поиск, эмбеддинг, небрежное решающее дерево

Ссылка для цитирования: Томилов Н.А., Туров В.П., Бабаянц А.А., Платонов А.В. Векторный поиск методом кластеризации с помощью ансамбля небрежных решающих деревьев // Научно-технический вестник информационных технологий, механики и оптики. 2025. Т. 25, № 2. С. 339–344 (на англ. яз.). doi: 10.17586/2226-1494-2025-25-2-339-344

Introduction

One of the methods for organizing search in large databases of heterogeneous data and documents involves a machine learning-based approach. This approach transforms data into representations that reflect the semantic structure of both textual [1] and multimodal documents in the form of vector embeddings being sequences of floatingpoint numbers. In this framework, efficient search can be achieved by constructing an index of these embeddings, transforming the search query into a vector embedding, and performing a nearest neighbor search using the constructed index [2]. Clustering, or grouping, vector embeddings is among the most popular indexing methods. In this method, the search index consists of cluster centroids. To perform a search, the system selects one or more centroids closest to the query and then performs an exhaustive search among the vector embeddings within the corresponding clusters. This indexing approach reduces the number of vector embeddings considered during the search to those within a single cluster [3]. A common clustering method for this purpose is the *k*-Nearest Neighbor (*k*NN) algorithm [4].

The most significant drawback of this method is its computational complexity. To determine which cluster a given vector embedding belongs to, it is necessary to calculate the distance from the embedding to all centroids of all existing clusters. It might also be necessary to, after assigning the embedding to the nearest cluster, recalculate the centroid of that cluster. Performing these computations requires not only access to the centroids of all clusters but also access to all vectors associated with the target centroid, which complicates the use of this algorithm in distributed systems. Less critical drawbacks include the use of a pseudorandom number generator during the clustering process, which may lead to different outputs for identical input data, as well as the limited number of hyper parameters available in the algorithm. In this paper we propose a clustering method using an ensemble of Oblivious Decision Trees (ODT) [5] as well as an approximate search method leveraging this clustering. This clustering assigns a hash to each vector embedding which serves as a cluster identifier and can be used to implement a Bloom filter for vector embeddings [6]. The rules for computing this hash are deterministic and independent of the vector data, allowing them to be serialized and used separately from the vector data. Our hypothesis is that this clustering method overcomes the drawbacks of the *k*NN method while providing comparable search accuracy.

Description of proposed clustering method

In this study, we propose an approximate vector search method based on clustering using an ensemble of binary or ternary ODT. Since training trees require significant computational effort, we build the ensemble by randomly selecting a training sample of size $N = N_{dataset} \times TrainRatio$, where $N_{dataset}$ is the amount of vectors in the dataset, $0.1 \leq TrainRatio \leq 1$. We transform each vector embedding of dimension d from the sample into M vector embeddings of a smaller dimension d_m , created by randomly selecting components from the original embedding. We also store a mapping table that links the indices of the components in the generated embeddings to their indices in the original embedding. This approach allows some components of the original embedding to remain unused or to appear multiple times, including multiple occurrences within the same generated embedding. After this transformation, we train a tree for each of the *M* sets of generated embeddings. These *M* trained trees form the ensemble.

The essence of the algorithm for constructing a ternary ODT of *depth* for vector embeddings of dimension *d* lies in selecting a set of elements $(K_j; X_j)$, where $1 \le j \le depth$, $1 \le K_j \le d_m$, at each tree level, such that the set of vector embeddings *V* is divided into three non-overlapping

subsets based on whether the K_j -th component of a vector embedding from V is less than X_j -delta, greater than X_j + delta, or approximately equal to X_j within the delta tolerance. This delta parameter defines the size of a group and is a configurable parameter of the structure. The partitioning rule is a hyper parameter, and in this study, we use a partitioning approach that minimizes the variance of distances between each vector embedding and the average vector embedding for that group.

The process is iterative and continues until the required depth is reached. At each step, the next division is performed for each of the three groups formed by the previous division. Thus, after the first division, the original set of vector representations V is split into three parts, V_1 , V_2 , and V_3 . On the next stage (the second level), each of the groups V_1 , V_2 , and V_3 is divided into three parts again, resulting in a total of nine groups, and so on. A distinctive feature of ODT is that the division for all groups at level *j* is done using the same value X_i [7]. Consequently, when dividing V, a situation may arise where the K_i -th component of all vector representations in V is strictly greater than or strictly less than X_i and thus all vector representations end up in one resulting group V_1 , while the groups V_2 , and V_3 are empty. This is a normal situation, but as a result, the final number of non-empty groups may be less than $\log_3(depth).$

After performing the calculations, the tree is described by an array of tuples $(K_j; X_j)$ and the *delta* parameter. Each node at depth *j* represents the operation "is the K_j -th value of the vector representation less than X_j -*delta*, greater than $X_j + delta$, or approximately equal to X_j within the tolerance of *delta*".

The process of individual tree construction is repeated for each of the M sets of vector representations to form the ensemble. After constructing the ensemble of trees, at the final stage, it is necessary to use the previously saved correspondence table to restore the K_i index for each element, associating the index of the element to be compared, within the M subvector with the d_m , dimensions, with the index of this element in the original vector representation of size d.

For each vector from the original dataset, calculations can be performed using the obtained tree, encoding the comparison result at each tree level *j* as '0' if the K_j -th component of the vector is less than X_j -delta, '1' if it is greater than $X_j + delta$, or 'E' in the case of approximate equality. This results in a ternary string of length depth which serves as the hash of the vector, and the length of this hash is equal to d_m . When creating an ensemble of *M* trees, *M* hashes of the length *nBits* are generated. We can call these individual hashes *SH*. These individual hashes can be combined into a single hash called *TH* of length *M nBits*. This is illustrated in Fig. 1. The exact order in which the combination occurs can be arbitrary, but it must be the same for each vector representation across the current set of vector representations.

In other words, by using M ODT, we can assign each vector from the original dataset a hash string TH. With this approach, similar vectors will have the same hash TH, meaning that vectors from the original dataset with identical hashes will belong to the same cluster.

A binary tree is a special case of a ternary tree where the value of delta equals zero. Consequently, during construction, each tree node splits into two parts: "greater" and "less than or equal". The resulting hash *TH* is then a binary string containing '0' and '1', thus it can be converted to an integer number.

Thanks to the fact that the tree is described by an array of tuples (K_j ; X_j), the tree can be serialized and transmitted separately from the data it was trained on. This allows, for example, the hash of a vector to be computed directly in client-side code when organizing client-server interactions, even without access to the data stored on the server, or in distributed systems without access to the data stored



Fig. 1. The process of transforming the original vector to its hash

on another cluster node. In addition, the large number of parameters and hyper parameters for selecting the appropriate partition allows fine-tuning of the ensemble for specific tasks. However, a significant drawback is the computational complexity of building the tree ensemble due to the exhaustive search of X_j values at each depth level. Even without considering the computation of distance variance, the number of iterations required for the search has a quadratic dependency on the number of input data, with additional increasing factors from the clustering parameters. Nevertheless, thanks to good parallelization, this drawback can be mitigated by using parallel or even heterogeneous computing (e.g., employing both CPU and GPU) [8].

In order to perform a search on data clustered using ODT, after clustering by hash value, it is necessary to compute and store the average vector for each cluster.

Description of proposed search method

When performing the kNN search for the vector embedding V_q obtained from the search query, we select N_q clusters, whose centroids are closest to V_q , from all clusters, where $N_q = \log_{10}N$, with N being the total number of clusters. Additionally, if present, we select the cluster whose hash matches the hash of V_q . Next, for each vector embedding in the selected clusters, we compute its distance to V_q . Then, for each cluster, we select k vector embeddings that are closest to V_q . The resulting vector embeddings are sorted in ascending order of distance to the k nearest embeddings, and the final result consists of the k embeddings closest to V_q .

The advantages of the algorithm include its versatility. The algorithm is independent of the distance metric and works with any method of storing vector embeddings, as long as the method allows storing metadata for each vector and/or supports grouping vector embeddings. However, the proposed algorithm does not completely eliminate the exhaustive search of vector embeddings within a group, which negatively impacts search time, especially in cases where parallel computations cannot be utilized.

With the exception of the step involving the selection of the cluster whose hash matches the hash of the query vector embedding, the proposed search algorithm is independent of the clustering algorithm. This allows for comparison of different clustering algorithms, as will be demonstrated in the comparison of the proposed clustering method using an ensemble of ODT with clustering using the kNN method.

Experiment setup

To validate the proposed vector search algorithm, we implemented it in the Kotlin programming language. For storage, we chose a previously implemented library for storing vector embeddings that group the embeddings into a single page file.

To test the proposed solution, we selected two datasets containing different vectors: the NYT-256-angular¹ and

¹ Newman D. Bag of Words // UCI Machine Learning Repository. 2008. doi: 10.24432/C5ZG6P Fashion-MNIST-784-euclidean [9] datasets taken from the popular ANN-Benchmarks [10] collection which serves as the industry standard for measuring the performance of approximate vector search algorithms. Specifically, Fashion-MNIST contains 50,000 vector embeddings of dimension 784, representing 32×32 pixel images where each component of the vector encodes the brightness of the corresponding pixel ranging from 0 to 255. The NYT dataset includes 290,000 vector embeddings of text articles from the New York Times, each containing 256 components. These two datasets differ in the semantics of the encoded data, value ranges, and distance metrics, allowing us to evaluate the algorithm performance on different types of input data. However, because the tree training implementation is time-intensive, we selected only the first 5,000 vector embeddings from each dataset to limit the tree training time to 12 hours.

The selected vector embeddings were transformed into file sets corresponding to the storage configurations listed below. The specific parameter values were chosen after numerous experiments, as they offer a balance between the algorithm accuracy, search speed, and the tree ensemble construction time.

As an alternative solution for comparison, we selected the traditional nearest neighbor clustering method which was previously implemented in the study on vector index compression [11]. In total, for the experiment, the original datasets were clustered and stored as follows:

- 1. Clustering vector embeddings using ternary ODT ensemble with the following parameters:
 - *delta* 0.15 and 0.35;
 - -M-2, 4, 8;
 - d_m 392 and 784 for the Fashion-MNIST dataset, 128 and 256 for the NYT dataset (in other words, 50 % and 100 % from the source dataset dimensionality);
 - *depth —* 2, 4, 8;
 - *TrainRatio* = 1.0 (in other words, all of the 5 thousand vectors were used to train the trees);
 - The partitioning rule: partition to minimize the variance of distances between each vector embedding and the average vector embedding for that group.
- 2. Clustering vector embeddings using ensemble of binary ODT, having the same parameters as ternary trees, except for *delta* which is 0;
- 3. Clustering vector embeddings using *k*NN clustering, maximum number of clusters being an arbitrary number between 15 and 15,000.

After obtaining the sets of clusters, we measured the search speed and accuracy using the following cluster variants:

- Clusters obtained by ternary tree ensembles for all parameter combinations.
- Clusters obtained by binary tree ensembles for all parameter combinations.
- Clusters obtained using the nearest neighbor method.

In addition to testing the performance of the proposed solution and the proposed benchmark, we also conducted performance testing of libraries that are industry standards



Fig. 2. Comparison results between kNN clustering and clustering using decision trees: Fashion-MNIST dataset (a); NYT dataset (b)

for approximate vector search, namely Annoy¹, Faiss [12], and HNSWLib [13]. The parameters of these storage systems were chosen to match those in ANN-Benchmarks [10], and the same 5,000 vector embeddings were selected for indexing. The test setup has the following specifications: AMD Ryzen 7 7700X 8C16T; 64GB RAM; NVMe WD SN850X 2TB; OS Ubuntu 22.04; OpenJDK 17; a tool for comparing vector search algorithms, developed in previous research [14], using Java Microbenchmark Harness [15].

Experiment results

The measurement results are presented in Fig. 2 where the best storage system is the one with the higher response and lower search time.

For the Fashion-MNIST dataset [9], the proposed clustering solution using trees cannot compete with kNN clustering under any combination of tree construction parameters. Since the vector embeddings in this dataset represent images of ten types of clothing, such data clusters well into a small number of clusters, no more than 1,000. Exhaustive search across these clusters provides high accuracy, and with such a small number of clusters and a small volume of data within each cluster, the search is fast. In contrast, clustering using ODT, with different input parameter sets, produces between 16 and 20,000 clusters. When there are too few clusters, leading to a large number of vector embeddings within each cluster, the search within the cluster becomes too slow. On the other hand, with too many clusters, the number of clusters to be searched becomes too small, resulting in lower accuracy. Acceptable accuracy is only achieved for parameter sets that produce

a clustering with fewer than 1,000 clusters, but with worse data grouping, leading to lower search accuracy.

For the NYT dataset, the proposed clustering solution using trees can compete with kNN clustering for certain parameter combinations. Clustering with trees still produces a large number of clusters ranging from 15 to 246,000; however, in this case, clustering also generates up to 15,000 clusters, which reduces the speed advantage caused by a small number of clusters. This increase compared to the previous dataset is due to the nature of the data which consists of dense embeddings of text documents. For parameter sets producing up to 100 clusters, clustering with trees provides a higher response compared to kNN clustering, at the cost of a slight increase in search time.

For both datasets, comparison with industrial standards for vector search shows that the proposed search method can compete in terms of both speed and accuracy with Annoy and Faiss, and for the NYT dataset, it can compete in accuracy with HNSW at the cost of longer search time.

Conclusion

For the selected datasets, the proposed search method in clusters generated by Oblivious Decision Trees (ODT) achieves search quality comparable to or better than that of the *k*NN method or libraries like Faiss, Annoy, and HNSW. However, for the Fashion-MNIST dataset, clustering using ODT demonstrates significantly slower search performance at the same quality metric values.

These results suggest that the proposed clustering and search method is comparable in search quality to industry standards. This makes further exploration of ODT for vector search worthwhile, specifically in accelerating the ensemble construction process and reducing the number of trees in the ensemble. The former would enable faster index construction, while the latter would speed up the search itself.

¹ GitHub — spotify/annoy: Approximate Nearest Neighbors in C++/Python optimized for memory usage and loading/saving to disk. Available at: https://github.com/ spotify/annoy, unrestricted. Language: English. (accessed: 14.06.2024).

References

- Grbovic M., Cheng H. Real-time personalization using embeddings for search ranking at Airbnb. Proc. of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, 2018, pp. 311–320. https://doi.org/10.1145/3219819.3219885
- Berry M.W., Drmac Z., Jessup E.R. Matrices, vector spaces, and information retrieval. *SIAM Review*, 1999, vol. 41, no. 2, pp. 335– 362. https://doi.org/10.1137/S0036144598347035
- Korn F., Sidiropoulos N., Faloutsos C., Siegel E., Protopapas Z Fast nearest neighbor search in medical image databases. *Proc. of the 22th International Conference on Very Large Data Bases (VLDB '96)*, 1996, pp. 215–226. https://doi.org/10.5555/645922.673493
- Seidl T., Kriegel H.-P. Optimal multi-step k-nearest neighbor search. ACM SIGMOD Record, 1998, vol. 27, no. 2, pp. 154–165. https://doi. org/10.1145/276305.276319
- Lou Y., Obukhov M. BDT: Gradient Boosted Decision Tables for high accuracy and scoring efficiency. *Proc. of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (KDD '17), 2017, pp. 1893–1901. https://doi. org/10.1145/3097983.3098175
- Chakrabarti S. Efficient image retrieval using multi neural hash codes and bloom filters. *Proc. of the IEEE International Conference for Innovation in Technology (INOCON)*, 2020, pp. 1–6. https://doi. org/10.1109/INOCON50539.2020.9298228
- Rokach L., Maimon O. *Data Mining with Decision Trees: Theory and Applications*. 2nd ed. World Scientific Publishing, 2014. 328 p.
- Fumero J., Papadimitriou M., Zakkak F.S., Xekalaki M., Clarkson J., Kotselidis C. Dynamic application reconfiguration on heterogeneous hardware. *Proc. of the 15th ACM SIGPLAN/SIGOPS International Conference on Virtual Execution Environments*, 2019, pp. 165–178. https://doi.org/10.1145/3313808.3313819
- Xiao H., Rasul K., Vollgraf R. Fashion-mnist: a novel image dataset for benchmarking machine learning algorithms. *arXiv*, 2017, arXiv:1708.07747. https://doi.org/10.48550/arXiv.1708.07747
- Aumüller M., Bernhardsson E., Faithfull A. Ann-benchmarks: A benchmarking tool for approximate nearest neighbor algorithms. *Lecture Notes in Computer Science*, 2017, vol. 10609, pp. 34–49. https://doi.org/10.1007/978-3-319-68474-1_3
- Tomilov N.A., Turov V.P., Babayants A.A., Platonov A.V. A method of storing vector data in compressed form using clustering. *Scientific* and Technical Journal of Information Technologies, Mechanics and Optics, 2024, vol. 24, no. 1, pp. 112–117. (in Russian). https://doi. org/10.17586/2226-1494-2024-24-1-112-117
- Douze M., Guzhva A., Deng C., Johnson J., Szilvasy G., Mazaré P.-E., Lomeli M., Hosseini L., Jégou H. The Faiss library. arXiv, 2024, arXiv.2401.08281. https://doi.org/10.48550/arXiv.2401.08281
- Malkov Y., Yashunin D. Efficient and robust approximate nearest neighbor search using Hierarchical Navigable Small World graphs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2020, vol. 42, no. 4. pp. 824–836. https://doi.org/10.1109/ TPAMI.2018.2889473
- Tomilov N.A., Turov V.P., Babayants A.A. Developing a tool to compare vector search algorithms. *Proc. of the 11th Congress of Young Scientists*, 2022, vol. 1. pp. 446–450. (in Russian).
- Laaber C., Leitner P. An evaluation of open-source software microbenchmark suites for continuous performance assessment. *Proc.* of the 15th International Conference on Mining Software Repositories (MSR '18), 2018, pp. 119–130. https://doi.org/10.1145/3196398.3196407

Authors

Nikita A. Tomilov — PhD Student, ITMO University, Saint Petersburg, 197101, Russian Federation, SC 57225127284, https://orcid.org/0000-0001-9325-0356, programmer174@icloud.com

Vladimir P. Turov — PhD Student, ITMO University, Saint Petersburg, 197101, Russian Federation, sc 58910796700, https://orcid.org/0009-0009-1470-7633, firemoon@icloud.com

Alexander A. Babayants — PhD Student, ITMO University, Saint Petersburg, 197101, Russian Federation, SC 58910002300, https://orcid. org/0009-0004-6367-6398, babayants.alexander@gmail.com

Alexey V. Platonov — PhD, Associate Professor, Associate Professor, ITMO University, Saint Petersburg, 197101, Russian Federation, sc 57197736275, https://orcid.org/0000-0002-8485-1296, avplatonov@ itmo.ru

Received 08.11.2024 Approved after reviewing 23.01.2025 Accepted 19.03.2025

Литература

- Grbovic M., Cheng H. Real-time personalization using embeddings for search ranking at airbnb // Proc. of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. 2018. P. 311–320. https://doi.org/10.1145/3219819.3219885
- Berry M.W., Drmac Z., Jessup E.R. Matrices, vector spaces, and information retrieval // SIAM Review. 1999. V. 41. N 2. P. 335–362. https://doi.org/10.1137/S0036144598347035
- Korn F., Sidiropoulos N., Faloutsos C., Siegel E., Protopapas Z Fast nearest neighbor search in medical image databases // Proc. of the 22th International Conference on Very Large Data Bases (VLDB '96). 1996. P. 215–226. https://doi.org/10.5555/645922.673493
- Seidl T., Kriegel H.-P. Optimal multi-step k-nearest neighbor search // ACM SIGMOD Record. 1998. V. 27. N 2. P. 154–165. https://doi. org/10.1145/276305.276319
- Lou Y., Obukhov M. BDT: Gradient Boosted Decision Tables for high accuracy and scoring efficiency // Proc. of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '17).2017.P. 1893–1901.https://doi.org/10.1145/3097983.3098175
- Chakrabarti S. Efficient image retrieval using multi neural hash codes and bloom filters // Proc. of the IEEE International Conference for Innovation in Technology (INOCON). 2020. P. 1–6. https://doi. org/10.1109/INOCON50539.2020.9298228
- Rokach L., Maimon O. Data Mining with Decision Trees: Theory and Applications / 2nd ed. World Scientific Publishing, 2014. 328 p.
- Fumero J., Papadimitriou M., Zakkak F.S., Xekalaki M., Clarkson J., Kotselidis C. Dynamic application reconfiguration on heterogeneous hardware // Proc. of the 15th ACM SIGPLAN/SIGOPS International Conference on Virtual Execution Environments. 2019. P. 165–178. https://doi.org/10.1145/3313808.3313819
- Xiao H., Rasul K., Vollgraf R. Fashion-mnist: a novel image dataset for benchmarking machine learning algorithms // arXiv. 2017. arXiv:1708.07747. https://doi.org/10.48550/arXiv.1708.07747
- Aumüller M., Bernhardsson E., Faithfull A. Ann-benchmarks: A benchmarking tool for approximate nearest neighbor algorithms // Lecture Notes in Computer Science. 2017. V. 10609. P. 34–49. https:// doi.org/10.1007/978-3-319-68474-1_3
- Томилов Н.А., Туров В.П., Бабаянц А.А., Платонов А.В. Метод хранения векторных представлений в сжатом виде с применением кластеризации // Научно-технический вестник информационных технологий, механики и оптики. 2024. Т. 24, № 1. С. 112–117. https://doi.org/10.17586/2226-1494-2024-24-1-112-117
- Douze M., Guzhva A., Deng C., Johnson J., Szilvasy G., Mazaré P.-E., Lomeli M., Hosseini L., Jégou H. The Faiss library // arXiv. 2024. arXiv.2401.08281. https://doi.org/10.48550/arXiv.2401.08281
- Malkov Y., Yashunin D. Efficient and robust approximate nearest neighbor search using Hierarchical Navigable Small World graphs // IEEE Transactions on Pattern Analysis and Machine Intelligence. 2020. V. 42. N 4. P. 824–836. https://doi.org/10.1109/ TPAMI.2018.2889473
- 14. Туров В.П., Томилов Н.А., Бабаянц А.А. Разработка инструмента сравнения алгоритмов векторного поиска // XI Конгресс молодых учёных: Сборник научных трудов. СПб.: Национальный исследовательский университет ИТМО, 2022. Т. 1. С. 446–450.
- Laaber C., Leitner P. An evaluation of open-source software microbenchmark suites for continuous performance assessment // Proc. of the 15th International Conference on Mining Software Repositories (MSR '18). 2018. P. 119–130. https://doi.org/10.1145/3196398.3196407

Авторы

Томилов Никита Андреевич — аспирант, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, вс 57225127284, https://orcid.org/0000-0001-9325-0356, programmer174@icloud.com Туров Владимир Павлович — аспирант, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, вс 58910796700, https://orcid.org/0009-0009-1470-7633, firemoon@icloud.com

Бабаянц Александр Амаякович — аспирант, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, вс 58910002300, https://orcid.org/0009-0004-6367-6398, babayants.alexander@gmail.com Платонов Алексей Владимирович — кандидат технических наук, доцент, доцент, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, вс 57197736275, https://orcid.org/0000-0002-8485-1296, avplatonov@itmo.ru

Статья поступила в редакцию 08.11.2024 Одобрена после рецензирования 23.01.2025 Принята к печати 19.03.2025 **I/İTMO**

НАУЧНО-ТЕХНИЧЕСКИЙ ВЕСТНИК ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ, МЕХАНИКИ И ОПТИКИ март–апрель 2025 Том 25 № 2 http://ntv.ifmo.ru/ SCIENTIFIC AND TECHNICAL JOURNAL OF INFORMATION TECHNOLOGIES, MECHANICS AND OPTICS March–April 2025 Vol. 25 No 2 http://ntv.ifmo.ru/en/ ISSN 2226-1494 (print) ISSN 2500-0373 (online)

ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ, МЕХАНИКИ И ОПТИКІ

ΜΑΤΕΜΑΤИЧЕСКОЕ И КОМПЬЮТЕРНОЕ МОДЕЛИРОВАНИЕ MODELING AND SIMULATION

10.17586/2226-1494-2025-25-2-345-353 УДК 621.3.01; 512.62

Полиномы комплексной мощности для решения задачи идентификации параметров линейной цепи Николай Владимирович Коровкин¹, Алексей Юрьевич Гришенцев²

¹ Санкт-Петербургский политехнический университет Петра Великого, Санкт-Петербург, 195251, Российская Федерация

² Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация

¹ Nikolay.Korovkin@gmail.com, https://orcid.org/0000-0002-1173-8727

² AGrishentsev@yandex.ru^{\overlinewy}, https://orcid.org/0000-0003-1373-0670

Аннотация

Введение. Идентификация параметров линейных электрических цепей относится к задачам анализа и обратным задачам физики. Современные исследования в этой области преимущественно сводятся к определению импульсной характеристики и/или передаточной функции электрической цепи. Задачи идентификации параметров линейных электрических цепей обычно ограничиваются выявлением двух-трех последовательно и/или параллельно соединенных линейных компонентов или Т- и П-образных четырехполюсников. Идентификация параметров в этом случае происходит путем исследования схемы в режимах короткого замыкания и холостого хода. Существуют несколько частных решений задачи идентификации на одной заданной частоте. В работе предложено решение задачи идентификации параметров линейных электрических цепей с использованием полиномов, выражающих комплексную мощность и дополнительных уравнений. Метод. Для решения задачи идентификации параметров пассивного линейного двухполюсника разработан метод синтеза основных уравнений с использованием разности комплексной мощности, вычисленной аналитически и полученной в результате приборных измерений, и представленной интерполяционным полиномом. Разработан метод синтеза дополнительных уравнений на основании производных по частоте от комплексных сопротивления и проводимости. Верхняя граница числа независимых уравнений оценивается как мощность множества степеней при круговой частоте, входящей в состав исследуемого полинома. Оценка наибольшего числа независимых уравнений позволяет сделать вывод о разрешимости задачи с применением основных уравнений, а также необходимости формирования дополнительных уравнений. Решения задачи идентификации параметров линейных пассивных электрических цепей реализуются численно методами компьютерной алгебры. Основные результаты. Практическое применение разработанных методов показано на примере численного моделирования. В качестве примера практического применения предлагаемых методов показано решение задачи определения параметров компонентов реальной схемы. Моделирование реализовано в программной среде Wolfram Mathematica. Обсуждение. Предложенное решение, в отличие от известных подходов, позволяет определять параметры компонентов линейных пассивных электрических цепей (двухполюсников). Разработан метод синтеза уравнений на основании разностей комплексной мощности, полученной в результате аналитического расчета и приборных измерений. Представлен метод синтеза дополнительных уравнений путем дифференцирования по частоте комплексных сопротивления и проводимости. Внедрение метода позволяет получить относительно простые формы уравнений. Уравнения можно синтезировать заранее для наиболее распространенных типовых схем

Ключевые слова

электротехника, радиоэлектроника, радиофизика, теория автоматического управления, линейные электрические цепи, идентификация параметров, анализ цепей, математическое обеспечение САПР

Благодарности

Работа выполнена при поддержке Министерства науки и высшего образования Российской Федерации, проект тематики научных исследований № 2019-0898.

Ссылка для цитирования: Коровкин Н.В., Гришенцев А.Ю. Полиномы комплексной мощности для решения задачи идентификации параметров линейной цепи // Научно-технический вестник информационных технологий, механики и оптики. 2025. Т. 25, № 2. С. 345–353. doi: 10.17586/2226-1494-2025-25-2-345-353

© Коровкин Н.В., Гришенцев А.Ю., 2025

Научно-технический вестник информационных технологий, механики и оптики, 2025, том 25, № 2 Scientific and Technical Journal of Information Technologies, Mechanics and Optics, 2025, vol. 25, no 2

Complex power polynomials for solving the problem of identifying linear circuit parameters

Nikolay V. Korovkin¹, Alexey Yu. Grishentsev²

¹ Peter the Great Saint-Petersburg Polytechnic University, Saint Petersburg, 195251, Russian Federation

² ITMO University, Saint Petersburg, 197101, Russian Federation

¹ Nikolay.Korovkin@gmail.com, https://orcid.org/0000-0002-1173-8727

² AGrishentsev@yandex.ru^{\overlinewy}, https://orcid.org/0000-0003-1373-0670

Abstract

The identification of parameters of linear electrical circuits refers to the problems of analysis and inverse problems of electrical engineering. Modern research in this field mainly boils down to determining the impulse response and/or the transfer function of an electrical circuit. The tasks of identifying the parameters of linear electrical circuits are usually limited to identifying two or three linear components connected in series and/or in parallel, or as T- and U-shaped fourpoles. In this case, the parameters are identified by examining the circuit in short-circuit and idle modes. There are also several particular solutions to the problem of identification at a given frequency. In this paper, we propose a solution to the problem of identifying the parameters of linear electrical circuits using polynomials expressing complex power and additional equations. To solve the problem of identifying the parameters of a passive linear bipolar, a method has been developed for synthesizing basic equations using the difference in complex power, calculated analytically and calculated as a result of instrument measurements and represented by an interpolation polynomial. A method for synthesizing additional equations based on the frequency derivatives of complex resistance and complex conductivity has been developed. The upper bound of the number of independent equations is estimated as the power of the set of degrees at a circular frequency, which is part of the polynomial under study. Estimating the largest number of independent equations will allow us to conclude that the problem is solvable using the basic equations, as well as the need to form additional equations. The solutions to the problem of identifying the parameters of linear passive electrical circuits are implemented numerically using computer algebra. The practical application of the developed methods is shown by the numeric example. As an example of the application of the proposed methods, the solution of the problem of determining the parameters of all elements of an electrical circuit is shown. The simulation is implemented in the Wolfram Mathematica software environment. The proposed solution, unlike the known approaches, allows us to determine the parameters of the components of linear passive electrical circuits (bipolar). For the first time, a method for synthesizing equations based on differences in complex power obtained as a result of analytical calculations and instrument measurements is proposed. The method of synthesizing additional equations by differentiating complex resistances and conductivities by frequency makes it possible to obtain relatively simple forms of equations. Such equations can be synthesized in advance for the most common typical schemes.

Keywords

electrical engineering, radio electronics, radiophysics, automatic control theory, linear electrical circuits, parameter identification, circuit analysis, CAD

Acknowledgements

This work was supported by the Ministry of Science and Higher Education of the Russian Federation, State Assignment no. 2019-0898.

For citation: Korovkin N.V., Grishentsev A.Yu. Complex power polynomials for solving the problem of identifying linear circuit parameters. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 2025, vol. 25, no. 2, pp. 345–353 (in Russian). doi: 10.17586/2226-1494-2025-25-2-345-353

Введение

Многие задачи радиофизики, радиоэлектроники и электротехники, а также смежных направлений (геофизики, автоматического управления и других), сводятся к идентификации параметров линейной электрической цепи (ИПЛЭЦ). Решение задачи ИПЛЭЦ актуально для определения параметров групп датчиков, расположенных в активных средах или труднодоступных частях устройств и сооружений [1], для определения параметров биофизических схем замещения [2] при моделировании биологических и геофизических объектов, наладке аппаратуры с применением методов идентификации и пр.

С другой стороны, разработка методов идентификации параметров линейных систем в общем, и линейных электрических цепей [3] в частности, способствует развитию фундаментальных дисциплин. Например, таких дисциплин как теоретическая электротехника [4], теория автоматического управления, теория систем автоматизированного проектирования (САПР) в части математического обеспечения [5–7].

Значительный вклад в развитие теории идентификации в области электротехники внесли следующие отечественные и зарубежные исследователи: Г. Крон, П.А. Бутырин, В.В. Филаретов, С.А. Курганов и др. [8–11].

Достаточно большое число публикаций по теории идентификации представлено в теории автоматического управления и смежных областях [12–14].

Задача идентификации линейных электрических цепей может быть интерпретирована и решена различными методами. Если рассматривать линейную цепь как «черный ящик», то может понадобиться решение задачи идентификации топологии схемы, подобные задачи могут быть актуальны в обратных задачах геофизики и других приложениях. Задача топологической идентификации нетривиальна и пока общего решения не имеет. В случае, если линейная цепь представлена «черным ящиком», то задача идентификации импульсной характеристики, во временной области и передаточной функции в частотной области, в настоящее время решена и используется в электротехнике и в теории автоматического управления. Если линейная схема рассматривается как «серый ящик» (Grey Box [15]), например, в случае известной топологии и неизвестных параметрах схемы, то решение такой задачи имеется только для некоторых частных случаев, например, для случая Т- или П-образного четырехполюсника. В работе [1] рассмотрено частное решение, позволяющее производить идентификацию параметров компонентов «серого ящика» с известной топологией. Отметим, что задача идентификации параметров компонентов линейной цепи, представимой в виде «серого ящика» с известной топологией, в общем виде не решена. В настоящей работе предлагается вариант общего решения, которое является основным результатом исследований.

Предлагаемые в работе методы можно использовать в качестве математической основы построения приборных систем ИПЛЭЦ, а значит, их разработка способствует развитию отрасли приборостроения.

В ряде обратных задач идентификации необходимо знать значения параметров элементов цепи, а не только передаточную функцию. Примерами таких задач являются: идентификация нагрузки трансформатора подстанции; идентификация параметров групп датчиков, размещенных в недоступных для непосредственного измерения отдельного компонента конструкциях и сооружениях; идентификация элементов RLC контуров антенно-фидерных устройств, контуров гетеродинных приемников, фильтров и других устройств; идентификация параметров для выявления неисправных компонентов или компонентов с параметрами, не соответствующими определенным в рамках заданной конструкции; идентификации компонентов в биологических и геофизических схемах замещения при известной топологии схемы и других. Спектр задач, в которых необходима идентификация параметров компонентов линейных цепей, достаточно широк.

Таким образом, актуальность работы обоснована, а результаты востребованы и способствуют развитию теории идентификации параметров линейных систем и практики построения приборных систем идентификации.

Задача исследования: разработка и совершенствование методов идентификации параметров линейной цепи на основе полинома комплексной мощности.

Исследования, приведенные в данной работе, основаны на положениях, принятых в теоретической электротехнике, в частности использованы идеальные RLCкомпоненты (катушка индуктивности, конденсатор, резистор). Такие идеальные компоненты не «шумят» (у них нет теплового шума) и обладают только одними сосредоточенными в пространстве параметрами: индуктивностью, емкостью и активным сопротивлением, причем параметры неизменны от времени, частоты, температуры, тока, напряжения и т. д. Способ использования идеальных компонентов для моделирования линейных цепей надежен и вместе с законами Г.Р. Кирхгофа, элементами теории графов, дифференциальными уравнениями во временной области и их операторными преобразованиями является основой построения теории линейных электрических цепей.

В работе не исследуется влияние шумов измерения на точность идентификации, а также обусловленность и устойчивость решения, что актуально для настоящих и будущих исследований. Следует отметить, что в работе [16] приведены исследования обусловленности получаемых уравнений, а также влияния возмущения входных данных задачи идентификации на ошибку решения. В настоящей работе внимание авторов сосредоточено на предлагаемых методах формирования основных уравнений на базе полиномов комплексной мощности и дополнительных уравнений за счет вычисления производных от комплексных сопротивления и проводимости. Влияния шумов и методы противодействия шумам, т. е. извлечения полезной информации на фоне шумов, достаточно хорошо исследованы в курсе статистической радиотехники [17, 18] и могут быть использованы для развития предлагаемой в работе теории при разработке практических методов реализации.

Постановка задачи и метод составления базовых уравнений

В качестве основы постановки задачи рассмотрим двухполюсник (рис. 1) с внешними полюсными зажимами *l* и *2*, к которым подключен источник электродвижущей силы (ЭДС) $\dot{E}(\omega)$, при этом через ветвь, содержащую ЭДС, протекает комплексный ток $\dot{I}(\omega)$, где $\omega = 2\pi f$ — круговая частота. Схема линейного RLC двухполюсника известна, но номинальные значения некоторых или всех параметров RLC-компонентов схемы неизвестны. Обозначим компоненты с неизвестными параметрами x_k , $k = \overline{1, K}$; компоненты с известными параметрами $-h_p$, p = 1, P. Тогда M = (P + K)— полное число компонентов схемы. Задача сводится к расчету численных величин элементов множества $\{x_k\}$, т. е. ИПЛЭЦ.

Используя закон Ома, запишем следующие уравнения:

$$Z(\omega) = \frac{\dot{E}(\omega)}{\dot{I}(\omega)}$$

$$Y(\omega) = \frac{\dot{I}(\omega)}{\dot{E}(\omega)}$$
(1)

Учитывая, что схема двухполюсника известна, выразим комплексные сопротивление $Z(\omega)$ и проводимость $Y(\omega)$ через отношение полиномов, которые



Puc. 1. Двухполюсник *Fig. 1.* Two-pole circuit

несложно получить, например, методом контурных токов [4]:

$$\begin{bmatrix} Z(\omega) = \frac{P_U(\{x_k\}, \{h_p\}, \omega)}{P_I(\{x_k\}, \{h_p\}, \omega)} \\ Y(\omega) = \frac{P_I(\{x_k\}, \{h_p\}, \omega)}{P_U(\{x_k\}, \{h_p\}, \omega)}, \end{bmatrix}$$
(2)

где $P_U(\{x_k\}, \{h_p\}, \omega)$ — полином, выражающий напряжение на зажимах l и 2; $P_I(\{x_k\}, \{h_p\}, \omega)$ – полином, выражающий ток в ветви, содержащей зажимы l и 2.

Известно, что комплексную мощность в цепи переменного тока можно рассчитать с помощью выражения: $S(\omega) = \dot{U}(\omega)\dot{I}(\omega)$, где $\dot{U}(\omega)$ и $\dot{I}(\omega)$ — комплексные напряжение и ток, подводимой цепи; $\ddot{I}(\omega)$ — величина комплексно-сопряженная с $\dot{I}(\omega)$.

На основании уравнений (1) и (2), запишем выражение комплексной мощности для исследуемого двухполюсника:

$$P_U(\{x_k\}, \{h_p\}, \omega) P_I^*(\{x_k\}, \{h_p\}, \omega) = \dot{E}(\omega) \overset{*}{I}(\omega).$$
(3)

Пусть в частном случае все компоненты двухполюсника (рис. 1) неизвестны, т. е. $\{h_p\} = \emptyset$, P = 0 и x_{k} , $k = \overline{1, M}$, тогда выражение (3) можно записать в виде:

$$P_U(\{x_k\}, \omega) P_I^*(\{x_k\}, \omega) = \dot{E}(\omega) \overset{*}{I}(\omega).$$
(4)

Причем $\dot{E}(\omega)\hat{I}(\omega)$ — произведение измеренных величин. Для этого произведения получим интерполяционный полином:

$$\dot{E}(\omega)\ddot{I}(\omega) = \sum_{\nu=0}^{V} r_{\nu}\omega^{\nu},$$
(5)

где <u> r_v </u> — полиномиальные коэффициенты при ω^v , v = 0, V — целочисленный индекс и показатель степени. Для получения интерполяционного полинома возможно использовать различные методы, обеспечивающие точное решение при известных точных значениях функции и аргумента в узлах интерполяции или минимальную ошибку при приближенных значениях функции и аргумента в узлах интерполяции. В настоящей работе и в [16], учитывая использование идеальных компонентов и отсутствия шумов измерения, применен метод интерполяции на основе схемы Горнера, позволяющий при известных точных значениях функции и аргумента в узлах интерполяции получить точное выражение интерполяционного полинома [19, 20]. При наличии шумов и приближенных значениях функции в узлах интерполяции оценить ошибку можно с помощью остаточного члена и методов оценки погрешности известных из вычислительной математики [19, 20].

Используя (5) и сгруппировав правую и левую части (4) по степеням при ω^{ν} , $\nu = \overline{0, V}$ запишем следующее выражение:

$$\sum_{\nu=0}^{V} \left(\sum_{l_1^{\nu} l_2^{\nu} \cdots \ l_M^{\nu}} G_{l_1^{\nu} l_2^{\nu} \cdots \ l_M^{\nu}} x_1^{l_1^{\nu}} x_2^{l_2^{\nu}} \dots x_2^{l_M^{\nu}} - r_{\nu} \right) \omega^{\nu} = 0, \quad (6)$$

где $l_1^{\nu} l_2^{\nu} \dots l_M^{\nu}$ — индексы суммы, а также индексы коэффициентов $G_{l_1^{\nu} l_2^{\nu} \dots l_M^{\nu}}$ полинома, и показатели степени различных комбинаций произведений $x_{l_1}^{l_1^{\nu}} x_{2}^{l_2^{\nu}} \dots x_{M}^{l_M^{\nu}}$ неизвестных $\{x_k\}$. Выражение суммы $\sum_{\substack{l_1^{\nu} l_2^{\nu} \dots l_M^{\nu}} G_{l_1^{\nu} l_2^{\nu} \dots l_M^{\nu}}$ $x_{l_1}^{l_1^{\nu}} x_{2}^{l_2^{\nu}} \dots x_{M}^{l_M^{\nu}}$ означает частичную сумму полинома (4), включающее сумму всех членов в виде произведения коэффициентов $G_{l_1^{\nu} l_2^{\nu} \dots l_M^{\nu}}$ на произведение неизвестных $G_{l_1^{\nu} l_2^{\nu} \dots l_M^{\nu}}$ при заданном значении ν степени при ω .

Выражение $\sum_{l_1^{l}l_2^{l}\cdots l_M^{v}} G_{l_1^{l}l_2^{l}\cdots l_M^{v}} x_1^{l_1^{v}} x_2^{l_2^{v}} \dots x_M^{l_M^{v}}$ может быть записано иначе: $\sum_{l_1^{v}} \sum_{l_2^{v}} \dots \sum_{l_M^{v}} G_{l_1^{l}l_2^{v}\cdots l_M^{v}} x_1^{l_1^{v}} x_2^{l_2^{v}} \dots x_M^{l_M^{v}}$. Далее в разделе «Пример численного моделирования» приведен частный случай полинома (4).

Если все параметры схемы известны (вычислены или измерены), т. е. $\{x_k\} = \emptyset$ и $h_p, p = \overline{1, M}$, то выражение (6) будет иметь вид:

$$\sum_{\nu=0}^{V} \left(\sum_{l_{1}^{\nu} l_{2}^{\nu} \cdots \ l_{M}^{\nu}} G_{l_{1}^{\nu} l_{2}^{\nu} \cdots \ l_{M}^{\nu}} h_{1}^{l_{1}^{\nu}} h_{2}^{l_{2}^{\nu}} \dots x_{M}^{l_{M}^{\nu}} - r_{\nu} \right) \omega^{\nu} = 0.$$
(7)

Теорема. Все частичные суммы $\begin{pmatrix} \sum \\ l_1^{\gamma} l_2^{\nu} \cdots l_M^{\nu} h_1^{l_1^{\nu}} h_2^{l_2^{\nu}} \cdots x_M^{l_M^{\nu}} - r_{\nu} \end{pmatrix}$ полинома (7) равны нулю.

Доказательство. В силу функционального равенства $P_U(\{h_p\}, \omega)P_I^*(\{h_p\}, \omega)$ и $\dot{E}(\omega)I(\omega)$, а значит $P_U(\{h_p\}, \omega)P_I^*(\{h_k\} \ u \sum_{\nu=0}^{V} r_\nu \omega^\nu$, и общности формы в смысле упорядоченности по степеням $\nu = \overline{0, V}$ при ω , каждая частичная сумма $\left(\sum_{I_1^* I_2^* \cdots I_M^\nu} G_{I_1^* I_2^*} \cdots I_M^{I_M^\nu} - r_\nu\right)$, при нулевой погрешности измерений ЭДС $\dot{E}(\omega)$ и тока $\tilde{I}(\omega)$ равна нулю, т. е. $\left(\sum_{I_1^* I_2^* \cdots I_M^\nu} G_{I_1^* I_2^*} \cdots I_M^\nu h_1^{I_1^\nu} h_2^{I_2^*} \cdots h_M^{I_M^\nu} - r_\nu\right) = 0$, что требовалось доказать.

Следствие 1. Каждая частичная сумма из (6) образует уравнение

$$\left(\sum_{l_1^{\nu} l_2^{\nu} \cdots \ l_M^{\nu}} G_{l_1^{\nu} l_2^{\nu} \cdots \ l_M^{\nu}} x_1^{l_1^{\nu}} x_2^{l_2^{\nu}} \dots x_M^{l_M^{\nu}} - r_{\nu}\right) = 0.$$

Замечание 1. На практике встречаются полиномы вида (6) которые могут содержать нулевые значения $\sum_{l_1^{V}l_2^{V}\cdots l_M^{V}} G_{l_1^{V}l_2^{V}\cdots l_M^{V}} r_1^{l_X^{V}} r_2^{l_Z^{V}} \dots r_M^{l_M^{V}} = 0$ и $r_v = 0$, для некоторых степеней v, v < V при круговой частоте ω . Обозначим буквой $\Lambda = (0, 1, ..., V)$ — упорядоченное множество всевозможных степеней, включая нулевую, при ω ; Ω — подмножество множества Λ , т. е. $\Omega \subseteq \Lambda$, которое содержит только те степени при ω , для которых

 $\sum_{l_1^{\nu} l_2^{\nu} \cdots l_M^{\nu}} G_{l_1^{\nu} l_2^{\nu} \cdots l_M^{\nu}} x_1^{l_1^{\nu}} x_2^{l_2^{\nu}} \dots x_M^{l_M^{\nu}} \neq 0$ и $r_{\nu} \neq 0$. Тогда число уравнений вида $\left(\sum_{l_1^{\nu} l_2^{\nu} \cdots l_M^{\nu}} G_{l_1^{\nu} l_2^{\nu} \cdots l_M^{\nu}} x_1^{l_1^{\nu}} x_2^{l_2^{\nu}} \dots x_M^{l_M^{\nu}} - r_{\nu}\right) = 0$ которые можно получить из (7), будет card(Ω) — мощность множества Ω .

Следствие 2. Число независимых уравнений, которые можно получить из (6) с помощью разделения на ча-

стичные суммы $\left(\sum_{l_1^{\nu}l_2^{\nu}\cdots l_M^{\nu}} G_{l_1^{\nu}l_2^{\nu}\cdots l_M^{\nu}} x_1^{l_1^{\nu}} x_2^{l_2^{\nu}} \dots x_M^{l_M^{\nu}} - r_{\nu}\right) = 0,$ не превышает card(Ω).

Замечание 2. Не превосходящее card(
$$\Omega$$
)
максимальное число уравнений вида
 $\left(\sum_{l_1^{\nu} l_2^{\nu} \cdots l_M^{\nu}} G_{l_1^{\nu} l_2^{\nu} \cdots l_M^{\nu}} x_1^{l_1^{\nu}} x_2^{l_2^{\nu}} \cdots x_M^{l_M^{\nu}} - r_{\nu}\right) = 0$ можно получить

из (4), т.е. для случая, когда параметры всех компонентов схемы неизвестны $\{h_p\} = \emptyset$.

Замечание 3. Следует отметить, что рассмотренный способ синтеза уравнений основан на разности комплексных мощностей $S(\omega)$ полученных в аналитической форме и с помощью измерений. С точки зрения электротехники такой подход обоснован и имеет физический смысл, но для составления уравнений и решения задачи достаточно алгебраического смысла. При этом выражение (3) может иметь вид:

$$P_{U}(\{x_{k}\}, \{h_{p}\}, \omega)P_{I}(\{x_{k}\}, \{h_{p}\}, \omega) = \dot{E}(\omega)\dot{I}(\omega), \quad (8)$$

т. е. не содержать комплексного сопряжения.

В результате для уравнения (8) будет справедлива теорема и следствия 1 и 2.

Замечание 4. Параметры схемы имеют вещественные значения, т. е. $x_k, h_p \in \mathbb{R}$ и $x_k > 0, h_p > 0$ для $k = \overline{1, K}, p = \overline{1, P}$ и M = P + K.

Следствие 3. Условия $x_k \in \mathbb{R}$ & $x_k > 0$, $k = \overline{1, K}$ необходимо использовать для отделения посторонних корней при решении (6) или (8).

Синтез дополнительных уравнений. При решении практических задач могут возникать ситуации, когда число неизвестных x_k , $k = \overline{1, K}$, больше числа уравнений $K > \operatorname{card}(\Omega)$, в таком случае для вычисления x_k , $k = \overline{1, K}$ необходимы дополнительные уравнения. Рассмотрим метод синтеза дополнительных уравнений.

Метод синтеза основан на взятии интегралов или производных (в настоящей работе рассмотрены производные) от комплексных сопротивления $Z(\omega)$ и проводимости $Y(\omega)$ по круговой частоте ω . Так, например, первые производные для выражений (1), (2) будут иметь следующий вид:

$$\begin{cases} Z'(\omega) = \frac{\dot{E}'(\omega)\dot{I}(\omega) - \dot{E}(\omega)\dot{I}'(\omega)}{\dot{I}^{2}(\omega)} \\ Y'(\omega) = \frac{\dot{I}'(\omega)\dot{E}(\omega) - \dot{I}(\omega)\dot{E}'(\omega)}{\dot{E}^{2}(\omega)}, \end{cases}$$
(9)

И

$$\begin{pmatrix}
E^{-}(\omega) \\
P_{U}^{\prime}(\{x_{k}\}, \{h_{p}\}, \omega)P_{I}(\{x_{k}\}, \{h_{p}\}, \omega) - \\
P_{U}^{\prime}(\{x_{k}\}, \{h_{p}\}, \omega)P_{I}^{\prime}(\{x_{k}\}, \{h_{p}\}, \omega) - \\
P_{I}^{2}(\{x_{k}\}, \{h_{p}\}, \omega)P_{I}^{\prime}(\{x_{k}\}, \{h_{p}\}, \omega) \\
P_{I}^{\prime}(\{x_{k}\}, \{h_{p}\}, \omega)P_{U}(\{x_{k}\}, \{h_{p}\}, \omega) - \\
P_{I}^{\prime}(\{x_{k}\}, \{h_{p}\}, \omega)P_{U}^{\prime}(\{x_{k}\}, \{h_{p}\}, \omega) - \\
P_{U}^{\prime}(\{x_{k}\}, \{h_{p}\}, \omega)P_{U}^{\prime}(\{x_{k}\}, \{h_{p}\}, \omega) - \\
P_{U}^{\prime}(\{x_{k}\}, \{h_{p}\}, \omega)P_{U}^{\prime}(\{x_{k}\}, \{h_{p}\}, \omega) - \\
P_{U}^{\prime}(\{x_{k}\}, в к выражениям (9) и (10) аналогичные преобразования (3)–(8), получим следующие уравнения:

$$\begin{cases} (P'_{U}(\{x_{k}\},\{h_{p}\},\omega)P_{I}(\{x_{k}\},\{h_{p}\},\omega) - \\ -P_{U}(\{x_{k}\},\{h_{p}\},\omega)P'_{I}(\{x_{k}\},\{h_{p}\},\omega)) \times \\ \times P_{I}^{2}(\{x_{k}\},\{h_{p}\},\omega) = \\ = (\dot{E}'(\omega)\dot{I}(\omega) - \dot{E}(\omega)\dot{I}'(\omega))\dot{I}^{2}(\omega) \\ (P'_{I}(\{x_{k}\},\{h_{p}\},\omega)P_{U}(\{x_{k}\},\{h_{p}\},\omega) - \\ -P_{I}(\{x_{k}\},\{h_{p}\},\omega)P'_{U}(\{x_{k}\},\{h_{p}\},\omega)) \times \\ \times P_{U}^{2}(\{x_{k}\},\{h_{p}\},\omega) = \\ = (\dot{I}'(\omega)\dot{E}(\omega) - \dot{I}(\omega)\dot{E}'(\omega))\dot{E}^{2}(\omega) \end{cases}$$
(11)

Уравнения (11) являются полиномами, степень которых и характеристика card(Ω) может быть больше аналогичной характеристики для (3). При использовании интегралов от комплексных сопротивления $Z(\omega)$ и проводимости $Y(\omega)$, получаемые выражения не являются полиномами, поэтому их анализ требует отдельного исследования, выходящего за рамки данной работы.

Заметим, что уравнения (11) не будут иметь смысл комплексной мощности даже при использовании комплексного сопряжения, поэтому если комплексное сопряжение не позволяет дополнительно упростить выражение (11), применение комплексного сопряжения для уравнений (11) использовать не имеет смысла.

Отметим, что производные от комплексных сопротивлений $Z'(\omega)$ и проводимости $Y'(\omega)$ имеют различные формы и, как следствие, позволяют получить различные формы уравнений. При этом эксперимент показывает, что уравнения (9) и (10) являются независимыми как между собой, так и по отношению к уравнениям, получаемым из (6). Данный факт позволяет решать задачи, в которых число неизвестных превышает card(Ω) для (6).

Подчеркнем, что для полученных полиномов (11) справедлива доказанная теорема и ее следствия.

Измерения и преобразования сигналов. Измерение напряжения и тока на входе двухполюсника удобно производить с помощью запоминающего осциллографа, обеспечивающего достаточную точность как за счет разрешения по амплитуде, так и за счет числа отсчетов. В качестве альтернативного способа измерений можно использовать векторный анализатор, который сразу позволяет получить значения измеряемых величин в частотной области. При измерениях с помощью осциллографа во временной области преобразование в частотную область несложно выполнить за счет дискретного преобразования Фурье. Отдельного внимания

требует вычисление производных тока
$$\frac{d^{(n)}l(\omega)}{d\omega^{(n)}}$$
 и напряжения $\frac{d^{(n)}\dot{U}(\omega)}{d\omega^{(n)}}$ по частоте на основании измерений тока

i(t) и напряжения u(t), сделанных во временной области, для этого можно воспользоваться дискретными аналогами известных [21, 22] формул:

$$\begin{cases} \frac{d^{(n)}\dot{I}(\omega)}{d\omega^{(n)}} = (-j)^n \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} i(t)t^n e^{-j\omega t} dt \\ \frac{d^{(n)}\dot{U}(\omega)}{d\omega^{(n)}} = (-j)^n \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} u(t)t^n e^{-j\omega t} dt \end{cases}, \quad (12)$$

где i(t) — ток и u(t) — напряжение во временной области; t — время; n — порядок производной.

Другой способ вычисления производных тока $\frac{d^{(n)}\dot{I}(\omega)}{d\omega^{(n)}}$ и напряжения $\frac{d^{(n)}\dot{U}(\omega)}{d\omega^{(n)}}$ по частоте можно реализовать методом конечных разностей, используя вычисленные с помощью преобразования Фурье значения комплексных тока $\dot{I}(\omega)$ и напряжения $\dot{U}(\omega)$.

Как известно вычисление производных с помощью правых или левых конечных разностей дает смещение (в данном случае по частоте), поэтому предпочтительнее использовать метод вычисления производной за счет свойств гармонических функций и преобразования Фурье с помощью выражения (12).

Пример численного моделирования

В качестве примера практического применения предлагаемых методов рассмотрим решение следующей задачи. Пусть необходимо определить параметры всех компонентов схемы (рис. 2) при известном токе $\dot{I}(\omega)$ в ветви, содержащей зажимы l и 2, и напряжении на зажимах $\dot{E}(\omega)$. Моделирование будем производить в программной среде Wolfram Mathematica¹.

С помощью метода контурных токов при заданном значении ЭДС $\dot{E}(\omega)$ вычислим функции комплексных сопротивления $Z(\omega)$ и проводимости $Y(\omega)$:

$$R_{1}R_{2} + R_{1}R_{3} + R_{2}R_{3} + j\omega R_{1}L_{2} + + j\omega R_{2}L_{1} + j\omega R_{3}L_{1} + Z(\omega) = \frac{\dot{E}(\omega)}{\dot{I}(\omega)} = \frac{+ j\omega R_{3}L_{2} - \omega^{2}L_{1}L_{2}}{R_{2} + R_{3} + j\omega L_{2}},$$
(13)
$$Y(\omega) = \frac{\dot{I}(\omega)}{\dot{E}(\omega)} = \frac{R_{2} + R_{3} + j\omega L_{2}}{R_{1}R_{2} + R_{1}R_{3} + R_{2}R_{3} + j\omega R_{1}L_{2} + + j\omega R_{2}L_{1} + j\omega R_{3}L_{1} + + j\omega R_{3}L_{2} - \omega^{2}L_{1}L_{2}},$$

где $j = \sqrt{-1}$; L_1 и L_2 — индукции.

Полагая, что все компоненты схемы неизвестны, введем следующие обозначения $R_1 = x_1, R_2 = x_2, R_3 = x_3, L_1 = x_4$ и $L_2 = x_5$. Используя (13), запишем полином в форме (4) в виде:

$$x_{1}x_{2}^{2} + 2x_{1}x_{2}x_{3} + x_{2}^{2}x_{3} + x_{1}x_{3}^{2} + x_{2}x_{3}^{2} - -j\omega(x_{2}^{2}x_{4} + 2x_{2}x_{3}x_{4} + x_{3}^{2}x_{4} + 2x_{1}x_{2}x_{5} + 2x_{1}x_{3}x_{5} + 2x_{2}x_{3}x_{5} + x_{3}^{2}x_{5}) - \dots$$
(14)
$$-\omega^{2}(x_{1}x_{5}^{2} + x_{3}x_{5}^{2} + 2x_{2}x_{4}x_{5} + 2x_{3}x_{4}x_{5}) + +j\omega^{3}x_{4}x_{5}^{2} = \dot{E}(\omega)\dot{I}(\omega)$$

Значение ЭДС не зависит от частоты, положим, что $\dot{E}(\omega) = 1$ В, для выполнения математического моделирования тока $\dot{I}(\omega)$, симулирующего численные измерения, зададим определенные номинальные значения компонентов: $R_1 = 1$ Ом, $R_2 = 2$ Ом, $R_3 = 3$ Ом, $L_1 = 4$ Гн и $L_2 = 5$ Гн.





При выборе номинальных значений параметров компонентов будем руководствоваться при решении соображениями получения относительно компактных форм выражений. С помощью моделирования вычислим численные значения тока $\dot{I}(\omega)$, заменяющие измерения, и методом Горнера [19, 20] синтезируем интерполяционный полином:

$$\dot{E}(\omega)\ddot{I}(\omega) = 55 - 255j\omega - 300\omega^2 + 100j\omega^3.$$
 (15)

Используя полученные выражения (14) и (15), на основании сформулированной и доказанной теоремы и следствия 1, синтезируем систему уравнений:

$$x_{4}x_{5}^{2} = 100$$

$$x_{1}x_{5}^{2} + x_{3}x_{5}^{2} + 2x_{2}x_{4}x_{5} + 2x_{3}x_{4}x_{5} = 300$$

$$x_{2}^{2}x_{4} + 2x_{2}x_{3}x_{4} + x_{3}^{2}x_{4} + 2x_{1}x_{2}x_{5} + 2x_{1}x_{3}x_{5} + . \quad (16)$$

$$+ 2x_{2}x_{3}x_{5} + x_{3}^{2}x_{5} = 255$$

$$x_{1}x_{2}^{2} + 2x_{1}x_{2}x_{3} + x_{2}^{2}x_{3} + x_{1}x_{3}^{2} + x_{2}x_{3}^{2} = 55$$

Очевидно, что при общем числе неизвестных — 5, число уравнений четыре, так как для (16) card(Ω) = 4. Следовательно, уравнений для решения задачи недостаточно. Синтезируем дополнительные к (16) уравнения на основании вычисления производных от комплексных сопротивления $Z(\omega)$ и проводимости $Y(\omega)$ по круговой частоте ω . Из полученных производных выберем ту, которая позволяет получить недостающее уравнение в более компактной форме. Производная от комплексного сопротивления $Z(\omega)$ позволяет получить 5 дополнительных уравнений, производная от проводимости $Y(\omega)$ — 7 дополнительных уравнений. Отметим, что наиболее простая форма дополнительного уравнения получена для $Z(\omega)$, поэтому и уравнения будем синтезировать на основе $Z'(\omega)$:

$$Z'(\omega) = \frac{j(x_2^2 x_4 + 2x_2 x_3 x_4 + x_3^2 x_4 + x_3^2 x_5) - \frac{j(x_2^2 x_4 x_5 + x_3 x_4 x_5) - j(x_2 x_4 x_5 + x_3 x_4 + x_3 x_4 + x_3 x_4 + x_3 x_4 + x_3 x_5) - j(x_2 x_4 x_5 + x_3 x_4 + x_3 x_4 + x_3 x_4 + x_3 x_4 + x_3 x_5) - j(x_4 x_5 + x_5 x_4 + x_5 x_5) - j(x_4 x_5 +$$

Запишем выражение для полинома (17), используя уравнения (11):

¹ Wolfram Mathematica. URL: https://www.wolfram.com/ mathematica/ (дата обращения: 03.11.2024).

$$j(x_{2}^{4}x_{4} + 4x_{2}^{3}x_{3}x_{4} + 6x_{2}^{2}x_{3}^{2}x_{4} + 4x_{2}x_{3}^{3}x_{4} + x_{3}^{4}x_{4} + x_{2}^{2}x_{3}^{2}x_{5} + 2x_{2}x_{3}^{3}x_{5} + x_{3}^{4}x_{5}) - -\omega(4x_{2}^{3}x_{4}x_{5} + 12x_{2}^{2}x_{3}x_{4}x_{5} + 12x_{2}x_{3}^{2}x_{4}x_{5} + 4x_{3}^{3}x_{4}x_{5} + 2x_{2}x_{3}^{2}x_{5}^{2} + 2x_{3}^{3}x_{5}^{2}) - \ldots$$
(18)
$$-j\omega^{2}(6x_{2}^{2}x_{4}x_{5}^{2} + 12x_{2}x_{3}x_{4}x_{5}^{2} + 6x_{3}^{2}x_{4}x_{5}^{2} + x_{3}^{2}x_{5}^{2}) + \omega^{3}(4x_{2}x_{4}x_{5}^{3} + 4x_{3}x_{4}x_{5}^{3}) + j\omega^{4}x_{4}x_{5}^{4} = (\dot{E}'(\omega)\dot{I}(\omega) - \dot{E}(\omega)\dot{I}'(\omega))\dot{I}^{2}(\omega)$$

Для (18) число card(Ω) = 5, что позволяет получить 5 дополнительных уравнений к полученным ранее четырем уравнениям (16). Вычислим интерполяционный полином пятой степени для правой части (18):

$$(\dot{E}'(\omega)\dot{I}(\omega) - \dot{E}(\omega)\dot{I}'(\omega))\dot{I}^{2}(\omega) =$$

$$= 3625i - 12\ 250\omega - 16\ 125j\omega^{2} +.$$
(19)
$$+10\ 000\omega^{3} + 2500j\omega^{4}$$

Используя (18), (19), запишем 5 дополнительных к (16) уравнений:

$$x_{2}^{4}x_{4} + 4x_{2}^{3}x_{3}x_{4} + 6x_{2}^{2}x_{3}^{2}x_{4} + 4x_{2}x_{3}^{3}x_{4} + x_{3}^{4}x_{4} + x_{2}^{2}x_{3}^{2}x_{5} + 2x_{2}x_{3}^{3}x_{5} + x_{3}^{4}x_{5} = 3625$$

$$4x_{2}^{3}x_{4}x_{5} + 12x_{2}^{2}x_{3}x_{4}x_{5} + 12x_{2}x_{3}^{2}x_{4}x_{5} + 4x_{3}^{3}x_{4}x_{5} + 2x_{2}x_{3}^{2}x_{5}^{2} + 2x_{3}^{3}x_{5}^{2} = 12\ 250 \qquad . (20)$$

$$6x_{2}^{2}x_{4}x_{5}^{2} + 12x_{2}x_{3}x_{4}x_{5}^{2} + 6x_{3}^{2}x_{4}x_{5}^{2} + x_{3}^{2}x_{5}^{2} = 16\ 125$$

$$4x_{2}x_{4}x_{5}^{2} + 4x_{3}x_{4}x_{5}^{2} = 10\ 000$$

$$x_{4}x_{5}^{4} = 2500$$

Теперь, имея достаточное число уравнений (16) и (20), можно найти решение поставленной задачи. Для решения будем использовать все уравнения из системы (16) и самое простое (нижнее) из системы (20).

В результате вычислений получаем 11 групп корней, исключив посторонние на основании условия $x_k \in \mathbb{R} \& x_k > 0, k = \overline{1, K}$, получим искомое решение: $\{x_1 = 1, x_2 = 2, x_3 = 3, x_4 = 4, x_5 = 5\}.$

Обсуждение

Анализ текущего состояния проблемы идентификации параметров линейных электрических цепей показывает, что с одной стороны результаты исследований востребованы в значительном числе практических приложений, а с другой стороны, что теория и практика идентификации параметров линейных электрических цепей разработана недостаточно. Преимущественно исследования сводятся к идентификации импульсной характеристики и/или передаточной функции, а задачи идентификации параметров линейных электрических цепей обычно ограничиваются идентификацией пары последовательно или параллельно соединенных линейных компонентов или Т- и П-образными четырехполюсниками, идентификация параметров которых происходит за счет исследования схемы в режиме короткого замыкания и холостого хода. Имеется несколько частных решений задачи идентификации параметров на одной частоте, например, в работе [1]. В настоящей работе изложена часть исследований, производимых авторами, отличающаяся от известных как принципом синтеза основных уравнений, на основе полиномов комплексной мощности, так и предлагаемым методом синтеза дополнительных уравнений за счет вычисления производных от комплексных сопротивления $Z(\omega)$ и проводимости Y(ω) по частоте. Следует отметить, что сама идея вычисления производных по частоте от $Z(\omega)$ и $Y(\omega)$ является новой и в известных работах не обнаружена, при этом исследование таких производных показывает значительные перспективы в решении задач идентификации параметров. Разработанные методы идентификации, могут быть использованы не только в электротехнике, но, например, в задачах идентификации параметров систем автоматического управления, а доказанная теорема способствует развитию алгебры многочленов. В качестве развития методов предполагается в дальнейшем провести следующие исследования: обусловленности уравнений и устойчивости решения в зависимости от возмущения входных данных; выявление и систематизация топологических признаков схем, приводящих к вырожденным уравнениям; проведение натурных экспериментов.

Заключение

В работе предложены методы синтеза основных и дополнительных уравнений для идентификации параметров линейных электрических двухполюсников. В ходе исследований сформулирована и доказана теорема о верхней границе числа независимых уравнений. Разработан метод синтеза уравнений на основании разностей комплексной мощности, полученной в результате аналитического расчета и в результате приборных измерений. Разработан метод синтеза дополнительных уравнений за счет дифференцирования по частоте комплексных сопротивления $Z(\omega)$ и проводимости $Y(\omega)$.

Разработанные методы позволяют получить относительно простые формы уравнений, причем такие формы можно синтезировать заранее для наиболее распространенных типовых схем, что удобно, например, в случае массовых измерений.

Теоретические исследования снабжены практическим примером, демонстрирующим способ применения теории в решении практических задач.

Предложенные методы способствуют развитию теоретической электротехники, радиотехнике, теории автоматического управления и могут быть математической основой построения приборов идентификации параметров линейных электрических цепей.

Литература

- Коровкин Н.В., Миневич Т.Г., Соловьева Е.Б. Определение электромагнитных параметров группы датчиков, предназначенных для измерений в активных средах или труднодоступных частях устройств и сооружений // Электротехника. 2023. № 3. С. 40–44. https://doi.org/10.53891/00135860_2023_3_40
- Зуев А.Л., Мишланов В.Ю., Судаков А.И., Шакиров Н.В., Фролов А.В. Эквивалентные электрические модели биологических объектов // Российский журнал биомеханики. 2012. Т. 16. № 1. С. 110–120.
- Коровкин Н.В., Гришенцев А.Ю. Определение параметров элементов линейного N-полюсника по входным частотным характеристикам // Электричество. 2024. № 6. С. 48–57. https://doi. org/10.24160/0013-5380-2024-6-48-57
- Демирчян К.С., Нейман Л.Р., Коровкин Н.В., Чечурин В.Л. Теоретические основы электротехники. Т. 1. СПб: Питер, 2006. 463 с.
- Гридин В.Н., Анисимов В.И., Алмаасали С.А. Применение метода диакоптики для моделирования и расчета больших систем // Проблемы управления. 2014. № 4. С. 9–13.
- Герасимов И.В., Кузьмин С.А., Лозовой Л.Н., Никитин А.В. Основания технологии комплементарного проектирования наукоемких изделий. С-Пб.: Изд-во СПбГЭТУ «ЛЭТИ» 2010. 196 с.
- Гришенцев А.Ю., Гурьянов А.В., Кузнецова О.В., Шукалов А.В., Коробейников А.Г. Математическое обеспечение в системах автоматизированного проектирования. СПб.: Университет ИТМО, 2017. 88 с.
- Крон Г. Исследование сложных систем по частям диакоптика. М.: Наука, 1972. 544 с.
- Бутырин П.А., Гришкевич А.А. Минимальные структуры математических моделей электрических цепей // Электричество. 1992. № 2. С. 11–21.
- Бутырин П.А., Васьковская Т.А. Принципы декомпозиции сложных электрических цепей при их диагностике по частям // Электричество. 2001. № 6. С. 41–48.
- Курганов С.А., Филаретов В.В. Неявный принцип наложения воздействий в линейных электрических цепях // Электричество. 2005. № 1. С. 45–60.
- Tang R., Zhu W., He Y. Research on anomaly identification of PV power station operation data based on fuzzy association rules // IEEE Access. 2024. V. 12. P. 156662–156672. https://doi.org/10.1109/ ACCESS.2024.3485082
- Uyanik I., Saranli U., Ankarali M.M., Cowan N.J., Morgül Ö. Frequency-Domain subspace identification of Linear Time-Periodic (LTP) systems // IEEE Transactions on Automatic Control. 2019. V. 64. N 6. P. 2529–2536. https://doi.org/10.1109/TAC.2018.2867360
- Tachibana H., Tanaka N., Maeda Y., Iwasaki M. Comparisons of Frequency Response Function identification methods using single motion data: time- and frequency-domain approaches // Proc. of the IEEE International Conference on Mechatronics (ICM). 2019. P. 498– 503. https://doi.org/10.1109/ICMECH.2019.8722867
- Tepljakov A., Petlenkov E., Belikov J. Grey box identification of fractional-order system models from frequency domain data // Proc. of the 41st International Conference on Telecommunications and Signal Processing (TSP). 2018. P. 1–4. https://doi.org/10.1109/ TSP.2018.8441247
- 16. Коровкин Н.В., Гришенцев А.Ю., Миневич Т.Г. Исследование обусловленности задачи определения параметров линейного N-полюсника по входным частотным характеристикам // Журнал радиоэлектроники. 2024. № 10. С. 16. https://doi. org/10.30898/1684-1719.2024.10.12
- 17. Тихонов В.И. Статистическая радиотехника. М.: Радио и связь, 1982. 624 с.
- Левин Б.Р. Теоретические основы статистической радиотехники. М.: Радио и связь, 1989. 656 с.
- Демидович Б.П., Марон И.А. Основы вычислительной математики. М.: Наука, 1969. 664 с.
- 20. Волков Е.А. Вычисление значений многочлена. М.: Наука, 1987. 248 с.
- 21. Дёч Г. Руководство к практическому применению преобразования Лапласа и Z-преобразования. М.: Наука, 1971. 288 с.
- 22. Ильин В.А., Поздняк Э.Г. Основы математического анализа. Ч. 2. М.: Физматлит, 2001. 464 с.

References

- . Korovkin N.V., Minevich T.G., Solovyeva E.B. Determining the electromagnetic parameters of a group of sensors intended for measuring in reactive media or in hard-to-reach parts of equipment. *Russian Electrical Engineering*, 2023, vol. 94, no. 3, pp. 181–185. https://doi.org/10.3103/S1068371223030070
- Zuev A.L., Mishlanov V.Y., Sudakov A.I., Shakirov N.V., Frolov A.V. Equivalent electrical models of biological objects. *Russian Journal* of *Biomechanics*, 2012, vol. 16, no. 1 (55), pp. 98–107.
- Korovkin N.V., Grishentsev A.Yu. Determination of the parameters of a linear N-Pole's elements from the input frequency responses. *Elektrichestvo*, 2024, no. 6, pp. 48–57. (in Russian). https://doi. org/10.24160/0013-5380-2024-6-48-57
- Demirchian K.S., Neiman L.R., Korovkin N.V., Chechurin V.L. *Theoretical Foundations of Electrical Engineering*. vol. 1. St. Petersburg, Piter Publ., 2006, 463 p. (in Russian)
- Gridin V.N., Anisimov V.I., Almaasali S.A. Application of diakoptics method for modeling and calculation of large systems. *Control Sciences*, 2014, no. 4, pp. 9–13. (in Russian)
- Gerasimov I.V., Kuz'min S.A., Lozovoj L.N., Nikitin A.V. Fundamentals of Technology for Complementary Design of High-Tech Products. St. Petersburg, ETU "LETI" Publ., 2010, 196 p. (in Russian)
- Grishentcev A.Iu., Gurianov A.V., Kuznetcova O.V., Shukalov A.V., Korobeinikov A.G. *Mathematical Support in Computer-Aided Design Systems*. St. Petersburg, ITMO University, 2017, 88 p. (in Russian)
- Kron G. Diakoptics: The Piecewise Solution of Large-Scale Systems. Macdonald, 1963, 166 p.
- Butyrin P.A., Grishkevich A.A. Minimal structures of mathematical models of electrical circuits. *Elektrichestvo*, 1992, no. 2, pp. 11–21. (in Russian)
- Butyrin P.A., Vas'kovskaya T.A. Decomposition principles for the complex electric circuits when diagnosing them by parts. *Elektrichestvo*, 2001, no. 6, pp. 41–48. (in Russian)
- Kurganov S.A., Filaretov V.V. An implicit superposition principle for actions in linear electric circuits. *Elektrichestvo*, 2005, no. 1, pp. 32– 43. (in Russian)
- Tang R., Zhu W., He Y. Research on anomaly identification of PV power station operation data based on fuzzy association rules. *IEEE Access*, 2024, vol. 12, pp. 156662–156672. https://doi.org/10.1109/ ACCESS.2024.3485082
- Uyanik I., Saranli U., Ankarali M.M., Cowan N.J., Morgül Ö. Frequency-Domain subspace identification of Linear Time-Periodic (LTP) systems. *IEEE Transactions on Automatic Control*, 2019, vol. 64, no. 6, pp. 2529–2536. https://doi.org/10.1109/ TAC.2018.2867360
- Tachibana H., Tanaka N., Maeda Y., Iwasaki M. Comparisons of Frequency Response Function identification methods using single motion data: time- and frequency-domain approaches. *Proc. of the IEEE International Conference on Mechatronics (ICM)*, 2019, pp. 498–503. https://doi.org/10.1109/ICMECH.2019.8722867
- Tepljakov A., Petlenkov E., Belikov J. Grey box identification of fractional-order system models from frequency domain data. *Proc. of the 41st International Conference on Telecommunications and Signal Processing (TSP)*, 2018, pp. 1–4. https://doi.org/10.1109/ TSP.2018.8441247
- Korovkin N.V., Grishentsev A.Y., Minevich T.G. Investigation of the conditionality of the problem of determining the parameters of a linear N-pole by input frequency characteristics. *Journal of Radio Electronics*, 2024, no. 10, pp. 16. (in Russian). https://doi. org/10.30898/1684-1719.2024.10.12
- Tikhonov V.I. Statistical Radio Engineering. Moscow, Radio i svjaz', 1982, 624 p. (in Russian)
- Levin B.R. Theoretical Foundations of Statistical Radio Engineering. Moscow, Radio i svjaz', 1989, 656 p. (in Russian)
- Demidovich B.P., Maron I.A. Basics of Computational Mathematics. Moscow, Nauka, 1969, 664 p. (in Russian)
- 20. Volkov E.A. *Calculating the Values of a Polynomial*. Moscow, Nauka, 1987, 248 p. (in Russian)
- Doetsch G. Anleitung zum Praktischen Gebrauch der Laplace-Transformation und der Z-Transformation. R. Oldenbourg, 1967, 256 p. (in German)
- 22. Ilin V.A., Pozdniak E.G. Basics of Mathematical Analysis. Part 2. Moscow, Fizmatlit, 2001, 464 p. (in Russian)

Авторы

Коровкин Николай Владимирович — доктор технических наук, профессор, профессор, Санкт-Петербургский политехнический университет Петра Великого, Санкт-Петербург, 195251, Российская Федерация, вс 6601971248, https://orcid.org/0000-0002-1173-8727, Nikolay.Korovkin@gmail.com

Гришенцев Алексей Юрьевич — доктор технических наук, доцент, доцент, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, sc 56321138400, https://orcid.org/0000-0003-1373-0670, AGrishentsev@yandex.ru

Статья поступила в редакцию 13.10.2024 Одобрена после рецензирования 21.01.2025 Принята к печати 27.03.2025

Authors

Alexey Yu. Grishentsev — D.Sc., Associate Professor, Associate Professor, ITMO University, Saint Petersburg, 197101, Russian Federation, se 56321138400, https://orcid.org/0000-0003-1373-0670, AGrishentsev@yandex.ru

Received 13.10.2024 Approved after reviewing 21.01.2025 Accepted 27.03.2025



Работа доступна по лицензии Creative Commons «Attribution-NonCommercial» I/İTMO

НАУЧНО-ТЕХНИЧЕСКИЙ ВЕСТНИК ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ, МЕХАНИКИ И ОПТИКИ март–апрель 2025 Том 25 № 2 http://ntv.ifmo.ru/ SCIENTIFIC AND TECHNICAL JOURNAL OF INFORMATION TECHNOLOGIES, MECHANICS AND OPTICS March–April 2025 Vol. 25 No 2 http://ntv.ifmo.ru/en/ ISSN 2226-1494 (print) ISSN 2500-0373 (online)

ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ, МЕХАНИКИ И ОПТИКИ

университет итмо

doi: 10.17586/2226-1494-2025-25-2-354-361 УДК 004.056.53

Компьютерное моделирование защищенности помещения от утечки информации по виброакустическому каналу Сергей Васильевич Скрыль¹, Михаил Владимирович Мурашов², Екатерина Сергеевна Голубцова^{3⊠}, Юлия Викторовна Грачёва⁴, Мария Алексеевна Паротькина⁵

1.2.3.4.5 Московский государственный технический университет им. Н.Э. Баумана, Москва, 105005, Российская Федерация

¹ skryl@bmstu.ru, https://orcid.org/0000-0002-4309-6255

² murashov@bmstu.ru, https://orcid.org/0000-0001-7564-0158

³ golubeka@bmstu.ru^{\box}, https://orcid.org/0000-0003-4395-2135

⁴ uvgracheva@bmstu.ru, https://orcid.org/0000-0003-0884-5617

⁵ masha20022811@gmail.com, https://orcid.org/0000-0001-5418-3111

Аннотация

Введение. Качество применения активных средств защиты информации от утечки по виброакустическому каналу традиционно оценивается по измеряемому значению соотношения уровня сигнал/шум. В этом случае для проверки защищенности выполняются измерения в нескольких контрольных точках, при этом размер выборки сравнительно мал, а процесс измерений является трудоемким. В настоящей работе учтены ранее полученные результаты исследований и моделирования распространения акустических волн в защищаемом помещении. Предлагается вместе с измеряемыми значениями соотношения сигнал/шум на поверхностях ограждающих конструкций рассчитывать речевую разборчивость в каждой точке внешней ограждающей поверхности. Заключение о защищенности помещения от утечки информации по виброакустическому каналу делается в результате сравнения полученных значений со шкалой оценки качества перехваченного речевого сообщения. Метод. Предложенный метод оценки защищенности помещения основан на компьютерном моделировании распространения звука. Рассмотрено взаимодействие акустических колебаний воздуха со стеной как одного из базовых ограждающих элементов помещения. Представленный подход применим и для анализа других элементов конструкции помещения. Конечно-элементная модель стены реализована в программе ANSYS. Модель предусматривает закрепление виброакустического излучателя на поверхности стены с одной стороны, а источника звука на некотором расстоянии от ее поверхности. В каждой точке с противоположной стороны поверхности стены вычисляется соотношение сигнал/шум. Уровню сигнала соответствуют значения колебаний поверхности стены под действием источника звука, а уровню шума — под действием виброизлучателя, прикрепленного к стене с внутренней стороны. Основные результаты. На основе экспериментального моделирования дана оценка полученных значений разборчивости речи на всей поверхности стены путем их сопоставления с принятой в практике защиты шкалой оценки качества перехваченного речевого сообщения. Обсуждение. Разработанный метод позволяет определять наименее защищенные места исследуемого помещения путем составления карт распределения разборчивости речи на поверхностях стен и других ограждающих конструкций. Исследование может найти применение при создании системы защиты информации от утечки по виброакустическому каналу. Проверка защищенности ограждающих конструкций с применением предложенного метода компьютерного моделирования позволяет выбрать рациональное расположение средств защиты, их количество, а также уточнить защищенность помещений, оцениваемую с использованием традиционных метолов.

Ключевые слова

акустика, защищенность, разборчивость речи, метод конечных элементов, технический канал утечки информации, виброакустический излучатель

Ссылка для цитирования: Скрыль С.В., Мурашов М.В., Голубцова Е.С., Грачёва Ю.В., Паротькина М.А. Компьютерное моделирование защищенности помещения от утечки информации по виброакустическому каналу // Научно-технический вестник информационных технологий, механики и оптики. 2025. Т. 25, № 2. С. 354–361. doi: 10.17586/2226-1494-2025-25-2-354-361

© Скрыль С.В., Мурашов М.В., Голубцова Е.С., Грачёва Ю.В., Паротькина М.А., 2025

Computer simulation of the protection of a room from information leakage through a vibroacoustic channel Sergey V. Skryl¹, Mikhail V. Murashov², Ekaterina S. Golubtsova^{3⊠}, Julia V. Gracheva⁴, Maria A. Parotkina⁵

1,2,3,4,5 Bauman Moscow State Technical University, Moscow, 105005, Russian Federation

¹ skryl@bmstu.ru, https://orcid.org/0000-0002-4309-6255

² murashov@bmstu.ru, https://orcid.org/0000-0001-7564-0158

³ golubeka@bmstu.ru^{\equiv}, https://orcid.org/0000-0003-4395-2135

⁴ uvgracheva@bmstu.ru, https://orcid.org/0000-0003-0884-5617

⁵ masha20022811@gmail.com, https://orcid.org/0000-0001-5418-3111

Abstract

The quality of application of active means of information protection against leakage via vibroacoustic channel is traditionally assessed by the measured value of signal-to-noise ratio. In this case, measurements are taken at several control points to check the security, while the sample size is relatively small, and the measurement process is laborintensive. Based on the studies previously conducted by the authors of the article and modeling of acoustic wave propagation in the protected room, this paper proposes to calculate speech intelligibility at each point of the external enclosing surface along with the measured values of the signal-to-noise ratio on the surfaces of enclosing structures. A conclusion on the security of the room from information leakage via vibroacoustic channel is made by correlating the obtained values with the scale for assessing the quality of the intercepted speech message. The proposed method for assessing the security of a room is based on computer modeling of sound propagation. The interaction of acoustic air vibrations with a wall as one of the basic enclosing elements of the room is considered. The proposed approach is also applicable to the analysis of other structural elements of the room. The finite element model of the wall is implemented in the ANSYS program. The model provides for fixing a vibroacoustic emitter on the wall surface on one side of the wall, and a sound source at some distance from this surface. At each point on the opposite side of the wall surface, the signal-to-noise ratio is calculated. The signal level corresponds to the values of wall surface vibrations under the action of the sound source, and the noise level corresponds to the values of wall surface vibrations under the action of a vibroemitter attached to the wall from the inside. Based on experimental modeling, an assessment is made of the obtained values of speech intelligibility on the entire wall surface by comparing them with the scale for assessing the quality of intercepted speech messages adopted in security practice. The method proposed in the paper allows one to determine the least protected places in the studied premises by compiling maps of speech intelligibility distribution on the surfaces of walls and other enclosing structures. The study can find application in creating a system for protecting information from leakage through a vibroacoustic channel. Checking the security of enclosing structures using the proposed computer modeling method allows choosing a rational arrangement of protective equipment, their quantity, and also clarifying the security of premises assessed using traditional methods.

Keywords

acoustics, security, speech intelligibility, finite element method, technical channel of information leakage, vibroacoustic emitter

For citation: Skryl S.V., Murashov M.V., Golubtsova E.S., Gracheva Ju.V., Parotkina M.A. Computer simulation of the protection of a room from information leakage through a vibroacoustic channel. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 2025, vol. 25, no. 2, pp. 354–361 (in Russian). doi: 10.17586/2226-1494-2025-25-2-354-361

Введение

Акустические сигналы в помещении распространяются в нескольких средах — воздушной и через конструкции здания (стены, пол, потолок). Для перехвата сигналов используют высокочувствительные микрофоны и стетоскопы. В зависимости от среды распространения сигнал при его перехвате имеет разное информационное содержание [1, 2]. В качестве активных средств защиты акустической информации от утечки через ограждающие конструкции часто используются устанавливаемые на конструкциях здания виброизлучатели, создающие шумовую помеху [3]. Оценка защищенности помещения производится при помощи измерения в контрольных точках соотношения сигнал/шум [4]. При этом процесс измерения является достаточно трудоемким, а выборка контрольных точек обычно мала [5].

Компьютерное моделирование широко распространено во многих областях исследований, однако его применение в задачах акустического и виброакустического расчетов недостаточно оценено [6]. Компьютерное моделирование дает возможность получить данные в любой точке модели, что исключает необходимость многочисленных измерений в контрольных точках. Эту особенность можно использовать при защите помещений от утечки информации по виброакустическому каналу. Например, в [7] рассмотрено применение метода конечных элементов для расчета поля давлений в помещении. В [8] предложен метод оценки защищенности помещения, основанный на конечно-элементном моделировании распространения акустических колебаний через панорамное окно, рассмотрены вопросы отражения и переотражения акустических колебаний от стен, пола и потолка. Проведенные расчеты позволяют выявить незащищенные места на поверхности стекла, а значит, потенциально опасные для утечки информации и требующие дополнительной защиты. Но предложенный метод не показывает степень защищенности мест возможного съема информации, а отмечает лишь наличие незащищенных мест и какую площадь поверхности они занимают.

В настоящей работе в качестве критерия защищенности была принята словесная разборчивость речи. В зависимости от условий проведения измерения и вида шумовой помехи зависимость разборчивости речи от отношения сигнал/шум может быть представлена различными формулами. Однако она достаточно точно аппроксимируется согласно исследованию [9].

Предложенный метод основан на конечно-элементном моделировании распространения звука и дает возможность оценить количественно и качественно защищенность помещения. В качестве примера рассмотрено распространение вибрационных колебаний в бетонной стене защищаемого помещения, возникающих под действием находящегося внутри помещения источника звука. Соотношение сигнал/шум на всей внешней поверхности стены получено из сопоставления значений амплитуды колебаний от источника звука с величиной вибрационных колебаний в стене, возникающих под действием только виброизлучателя. Качество защиты может быть охарактеризовано расчетом разборчивости речи на внешней стороне ограждающей конструкции, проводимом на основании ранее полученного соотношения сигнал/шум. Разработанный метод позволяет специалисту по защите информации, проектирующему систему защиты, понять, в какой степени возможна утечка информации, и определить уязвимые места защищаемого помещения.

Целью исследования является разработка и описание основанного на компьютерном моделировании метода анализа защищенности помещения от утечки информации по виброакустическому каналу, а также составление карты распределения разборчивости речи на поверхности защищаемой ограждающей конструкции помещения.

Описание метода анализа защищенности помещения от утечки информации по виброакустическому каналу

Предлагаемый метод включает в себя следующие этапы.

Этап 1. Создание модели защищаемого помещения по заданным размерам и с учетом материалов конструкции, источников сигнала и активных средств защиты информации, расположенных согласно требованиям. Следует учесть, что в защищенных помещениях не предусматривается перестановка мебели и крупногабаритных конструкций.

Этап 2. Разбиение модели на конечные элементы.

Этап 3. Задание нагрузок для решения математической модели.

Математическая модель решаемой задачи состоит из следующих уравнений [10]:

уравнение звуковой волны в воздушной среде

$$\frac{1}{c^2}\frac{\partial^2 P}{\partial t^2} - \nabla^2 P = 0;$$

— уравнение колебаний в твердом теле

$$m\frac{\partial^2 \mathbf{x}}{\partial t^2} + b \frac{\partial \mathbf{x}}{\partial t} + k\mathbf{x} = 0;$$

условия на границе контакта газ-твердое тело

$$\mathbf{n}\nabla P|_{\mathrm{rp}} = -p_0 \mathbf{n} \frac{\partial^2 \mathbf{x}}{\partial t^2}\Big|_{\mathrm{rp}},$$

где c — скорость звука; P — акустическое давление; t — время; m — масса; b — коэффициент поглощения; k — коэффициент жесткости; \mathbf{x} — вектор амплитуды перемещения; \mathbf{n} — нормаль к границе газ-твердое тело; p_0 — плотность газа.

Этап 4. Расчет значений виброперемещений при включенном источнике акустического сигнала и выключенном виброизлучателе, а затем расчет значений виброперемещений при включенном виброизлучателе и выключенном источнике звука.

Этап 5. Чтобы исключить погрешности метода конечных элементов и получить более точный результат в программе ANSYS, на данном этапе необходимо проанализировать различные варианты и выбрать рациональное разбиение объемов на конечные элементы.

Этап 6. Соотношение сигнал/шум *Q* вычисляется по отношению амплитуд колебаний (в метрах) на внешней стороне ограждающей конструкции от воздействия отдельно источника звука и виброизлучателя по формуле

$$Q = 20 \lg \frac{A_s}{A_N},\tag{1}$$

где модули значений амплитуд виброперемещений на внешней стороне ограждающей конструкции: A_S — при включенном источнике звука и выключенном виброизлучателе; A_N — при включенном виброизлучателе и выключенном источнике звука.

Этап 7. Проведение расчета разборчивости речи на поверхности конструкции на основе полученных на этапе 6 значений соотношения (1). Так как значения виброперемещений не позволяют оценить защищенность, вычислим разборчивость речи W по значениям виброперемещений согласно формуле [11]

$$W = \frac{1}{1 + 0.21Q^{-1}}.$$
 (2)

Формула (2) носит общий характер, справедливый для интегральных значений разборчивости речи в полной полосе речевого сигнала. Также отметим, что формула (2) использована для построения модели в первом приближении, поскольку известно, что в практике возможны иные парциальные зависимости разборчивости речи от соотношения сигнал/шум в отдельных октавных полосах.

Этап 8. Сопоставление полученного значения разборчивости речи со шкалой оценки разборчивости. Составление заключения о защищенности помещения и, в случае необходимости, подготовка рекомендаций по улучшению качества системы защиты.

Численное моделирование элемента ограждающей конструкции

Описание численной модели. В качестве примера применения разработанного метода рассмотрим часть помещения, состоящую из стены, источника звука и генератора шума, состав которой соответствует типовой защищенной системе [12]. Моделью стены будет параллелепипед с прикрепленным к нему с внутренней стороны виброакустическим излучателем. Размеры построенной модели $4 \times 3 \times 1$ м (рис. 1). Отражения и переотражения акустических колебаний от стен, пола и потолка не рассматриваются.

Виброизлучатель и источник звука выполнены из стали. Источник звука, смоделированный в форме куба с длиной стороны 0,042 м, находится на расстоянии одного метра от стены и отделен слоем воздуха.

Внесение изменений в компьютерную модель не представляет больших трудностей. В частности, можно учесть перемещения источника звука, виброизлучателя, изменение размеров стены и других геометрических параметров, а также свойства материалов стены и неоднородность или многослойность конструкции.

В центре стены с внутренней стороны закреплен виброизлучатель. За его основу был принят широко распространенный в практике системы защиты речевой информации виброизлучатель модели CB-4Б — генератор вибраций производства НПО «АННА» с полосой воспроизводимых частот 175–11 200 Гц¹. Контакт виброизлучателя с поверхностью стены представляет собой неидеальный контакт двух шероховатых тел, конечно-элементное моделирование которого требует значительных вычислительных ресурсов [13, 14]. Для упрощения расчетов в первом приближении контакт виброизлучателя и стены соответствует соединению двух плоских поверхностей.

Установлено, что виброизлучатель и источник звука вызывают колебания, исходящие в перпендикулярном направлении от их поверхностей (нагрузки приложены к поверхностям источника звука и виброизлучателя по оси z). Чтобы оценить защищенность, нужно рассматривать случай наиболее благоприятный для злоумышленника. Зададим амплитуду колебаний на источнике звука, соответствующую верхней границе громкости человеческого голоса или шумного обсуждения в группе людей (80 дБ). Расчеты выполнены с использованием конечно-элементной программы ANSYS.

Проведен цикл расчетов для разных среднегеометрических октавных частот (500, 1000, 2000, 4000 и 8000 Гц). Так как численное моделирование широкополосных сигналов в данной задаче затруднительно из-за высокой вычислительной стоимости расчетов, предположим, что виброизлучатель и источник звука генерируют гармонический тональный сигнал на выбранных частотах. Для краткости представим результат для частоты 500 Гц. Ему соответствует амплитуда коллебаний источника звука 2,97 · 10⁻⁴ м. Амплитуда коле-



Puc. 1. Модель исследуемой стены *Fig. 1.* Model of the investigated wall

баний виброизлучателя на частоте 500 Гц составляет 1·10⁻⁴ м [15]. Результаты для других частот качественно не отличаются и соответствуют сделанным в работе выводам.

Исходные данные моделирования. Источник звука и виброизлучатель выполнены из легированной стали. Материал стены — бетон.

Размеры (длина × высота × толщина): стена $4 \times 3 \times 0,15$ м; воздушное пространство перед стеной $4 \times 3 \times 1$ м; источник звука $0,042 \times 0,042 \times 0,042$ м; виброизлучатель цилиндрической формы с радиусом 0,025 м и высотой 0,025 м.

Конечно-элементная модель сформирована в программе ANSYS. Стена и источник звука разбиты на конечные элементы типа SOLID186, а воздушное пространство — на элементы FLUID220 [6].

В целях снижения возможной погрешности аппроксимации был проведен анализ моделей с различной плотностью сетки конечных элементов. Принято, что по толщине стена должна быть разбита на два слоя. Увеличение количества элементов по толщине до трех дало выигрыш в точности на 6 %, но существенно уве-



Рис. 2. Разбиение модели на конечные элементы: воздушного пространства и стены (*a*); внутренней стороны стены и виброизлучателя (*b*)

Fig. 2. Splitting the model into finite elements: the air space and the wall (a); the inner side of the wall and the vibroemitter (b)

¹ Защита переговоров, видеонаблюдение, контроль доступа, охранно-пожарная безопасность [Электронный ресурс]. — URL: https://detsys.ru/catalog/vibroakusticheskaya_zashchita/ generator_vibroizluchatel_sv_4b/ (дата обращения: 23.01.2025).

личило время расчета модели. Уменьшение количества элементов до одного элемента по толщине дало уменьшение точности расчета на 30 %. Подобранное рациональное сеточное разбиение показано на рис. 2.

Результаты численного моделирования. Рассмотрим модель с включенным генератором шума и выключенным источником звука. Результаты расчета виброперемещений на стене (внешнем слое) представлены на рис. 3, *а.* Результаты расчета модели с включенным источником звука и выключенным виброизлучателем показаны на рис. 3, *b.*

Измерение уровня помех для речи является одним из широко используемых методов прогнозирования ее разборчивости и вычисляется как разность уровня



 $0,694 \cdot 10^{-10} \quad 0,245 \cdot 10^{-5} \quad 0,490 \cdot 10^{-5} \quad 0,735 \cdot 10^{-5} \quad 0,980 \cdot 10^{-5} \\ 0,122 \cdot 10^{-5} \quad 0,367 \cdot 10^{-5} \quad 0,612 \cdot 10^{-5} \quad 0,857 \cdot 10^{-5} \quad 0,110 \cdot 10^{-4}$



Рис. 3. Амплитуда виброперемещений на внешней стороне стены по оси *z*, м, при включенном виброизлучателе и выключенном источнике звука $(A_N)(a)$; при включенном источнике звука и выключенном виброизлучателе $(A_S)(b)$ (значения взяты по модулю)

Fig. 3. Amplitude of vibration displacements on the outer side of the wall along the *z* axis, m, with the vibration emitter turned on and the sound source turned off $(A_N)(a)$; with the sound source turned on and the vibration emitter turned off $(A_S)(b)$ (the values are taken by the absolute value)



звукового давления речи и уровня шумовых помех, измеренных в месте предполагаемого считывания информации¹. Уровень разборчивости речи является удовлетворительным, если уровень сигнала превышает уровень шумовых помех на величину более 10 дБ.

На рис. 4 показан уровень превышения сигнала над шумом. Он был рассчитан по формуле (1).

Сами значения виброперемещений не позволяют оценить защищенность, поэтому по формуле (2) была рассчитана разборчивость речи на поверхности конструкции на основе полученных значений соотношения (1).

На рис. 5 показаны значения разборчивости речи на внешней стороне стены, которые могут принимать значения от 0 до 1. Для каждой точки поверхности мож-





Fig. 5. The speech intelligibility values (*W*) obtained from the calculation on the outer side of the wall

¹ ГОСТ Р ИСО 9921-2013 Эргономика. Оценка речевой связи. Введен. 01.12.2014. М.: Стандартинформ, 2014. 24 с.


Рис. 6. Распределение разборчивости речи на поверхности стены (в области синего цвета невозможно установление факта наличия речи)

Fig. 6. Distribution of intelligibility on the surface of the wall (in the blue area it is impossible to establish the fact of the speech presence)

но увидеть и качественно оценить значение речевой разборчивости.

На рис. 6 различными цветами подробно показаны диапазоны в соответствии с возможной для перехвата информацией. Практический опыт показывает, что при показателе словесной разборчивости более 40 % возможно установление предмета ведущегося разговора и его содержание [4].

Обсуждение результатов численного моделирования

Размер пятен с максимальной степенью разборчивости достаточен для того, чтобы, не имея специальной аппаратуры для определения размеров поверхности, определить «на глаз» незащищенные места и закрепить акселерометр для съема информации. Синим обозначены места, где нет утечки информации. Для специалиста по защите информации очень важна такая карта, поскольку вероятность того, что измерения покажут незащищенные места в выбранных контрольных точках, крайне мала. Но для злоумышленника достаточно наличия одного незащищенного места, чтобы с него можно было снять информацию. На рис. 6 отчетливо видны эти пятна. Наиболее опасные места обозначены красным. Пятна занимают большую площадь и удалены от центра, где расположен виброизлучатель.

Полученное распределение соотношения сигнал/шум (рис. 4) показало достаточно логичную картину с минимальными величинами в центре стены, обусловленными действием шума от виброизлучателя. При этом картина распределения соотношения сигнал/ шум достаточно сложна даже при геометрически простой форме ограждающей конструкции. Полученное распределение разборчивости речи показало защиту от установления предмета ведущегося разговора в зоне крепления виброизлучателя. Вместе с этим получено наличие незащищенных зон, разбросанных у границ ограждающей конструкции, в виде небольших пятен размером порядка 5 см. Тем не менее, они достаточно крупные для того, чтобы злоумышленник, получив с помощью моделирования места их расположения и закрепив там датчик, мог вести запись разговора с подробным восстановлением его содержания. Без заранее известных сведений о форме и расположении этих зон их экспериментальное нахождение может оказаться затруднительно, так как доля суммарной площади пятен в площади стены невелика. Этот факт может также способствовать получению ложноположительного заключения о защищенности помещения при контрольных измерениях в случайно выбранных точках.

Заключение

В работе предложен основанный на компьютерном моделировании распространения акустических колебаний метод анализа защищенности помещения от утечки информации по виброакустическому каналу. Он позволяет определить уязвимые места защищенного помещения, составить карту распределения разборчивости речи на поверхности защищаемой ограждающей конструкции помещения. Данное исследование должно оказаться крайне полезным для специалиста по защите информации, который будет создавать систему защиты информации от утечки по виброакустическому каналу. Использование компьютерного моделирования позволяет вывести на новый уровень организацию системы защиты и повысить ее качество. Проверка с применением компьютерного моделирования сможет подтвердить или опровергнуть защищенность, в том числе для мест, незащищенность которых могла быть пропущена при проведении измерений классическим методом.

К основным ограничениям применения данного метода можно отнести сложность моделирования широкополосных сигналов. В качестве дальнейших исследований авторами запланировано проведение эксперимента для проверки полученных результатов.

Литература

- Wegener K., Bleicher F., Heisel U., Hoffmeisterd H.-W., Möhring H.-C. Noise and vibrations in machine tools // CIRP Annals. 2021. V. 70. N 2. P. 611–633. https://doi.org/10.1016/j. cirp.2021.05.010
- Насибуллаева Э.Ш. Моделирование акустического рассеяния от множества звукопроницаемых сфер в трехмерном пространстве // Вычислительные технологии. 2022. Т. 27. № 2. С. 19–36. https:// doi.org/10.25743/ICT.2022.27.2.003
- Yamamoto K., Masatoshi T., Nakagawa S. Privacy protection for speech signals // Procedia Social and Behavioral Sciences. 2010. V. 2. N 1. P. 153–160. https://doi.org/10.1016/j.sbspro.2010.01.029
- Хорев А.А. Защита информации от утечки по техническим каналам. Часть 1. Технические каналы утечки информации. М.: Гостехкомиссия России, 1998. 320 с.
- Бурлаков М.Е., Осипов М.Н. Акустические и виброакустические каналы утечки информации. Теоретические основы и базовый практикум. Самара: Издательство Самарского университета, 2021. 96 с.
- Howard C.Q., Cazzolato B.S. Acoustic Analyses Using MATLAB and ANSYS. CRC Press, 2014. 708 p.
- Мурашов М.В. Численное моделирование в ANSYS поля давлений при мероприятиях по обеспечению защиты информации от утечки по акустоэлектрическому каналу // Наука и образование: научное издание МГТУ им. Н.Э. Баумана. 2009. № 10. С. 2.
- Мурашов М.В., Голубцова Е.С., Капитонова Л.И., Федорова В.А., Савина Т.Т. Исследование защищенности помещения от утечки информации по виброакустическому каналу с помощью программы ANSYS // Защита информации. Инсайд. 2022. № 3 (105). С. 5–9.
- Horev A., Margarian T. Experimental studies of voice messages intelligibility // Proc. of the Conference of Russian Young Researchers in Electrical and Electronic Engineering (EIConRus). 2019. P. 1538– 1543. https://doi.org/10.1109/EIConRus.2019.8657302
- Щевьев Ю.П. Основы физической акустики. СПб.: Лань, 2021. 364 с.
- Герасименко В.Г., Лаврухин Ю.Н., Тупота В.И. Методы защиты акустической речевой информации от утечки по техническим каналам. М.: РЦИБ «Факел», 2008. 258 с.
- Lukmanova O., Horev A.A., Vorobeyko E., Volkova E.A. Research of the analog and digital noise generators characteristics for protection device // Proc. of the IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering, (EIConRus), 2020. P. 2093– 2096. https://doi.org/10.1109/EIConRus49466.2020.9039193
- Murashov M.V., Panin S.D. Numerical modelling of contact heat transfer problem with work hardened rough surfaces // International Journal of Heat and Mass Transfer. 2015. V. 90. P. 72–80. https://doi. org/10.1016/j.ijheatmasstransfer.2015.06.024
- 14. Мурашов М.В. Выбор параметров вычислительных алгоритмов при решении задачи контактного деформирования шероховатых тел в ANSYS // Вестник МГТУ им. Н.Э. Баумана. Серия Естественные науки. 2016. № 1 (64). С. 111–121. https://doi. org/10.18698/1812-3368-2016-1-111-121
- 15. Коробейников А.Г., Калинкина М.Е., Ткалич В.Л., Пирожникова О.И., Гришенцев А.Ю. Моделирование состояния поверхности мембраны при точечном воздействии // Научнотехнический вестник информационных технологий, механики и оптики. 2020. Т. 1. № 1. С. 155–161. https://doi.org/10.17586/2226-1494-2020-20-1-155-161

Авторы

Скрыль Сергей Васильевич — доктор технических наук, профессор, профессор, Московский государственный технический университет им. Н.Э. Баумана, Москва, 105005, Российская Федерация, 5 57211168577, https://orcid.org/0000-0002-4309-6255, skryl@bmstu.ru Мурашов Михаил Владимирович — доктор технических наук, профессор, доцент, Московский государственный технический университет им. Н.Э. Баумана, Москва, 105005, Российская Федерация, 5 55204837500, https://orcid.org/0000-0001-7564-0158, murashov@ bmstu.ru

Голубцова Екатерина Сергеевна — ассистент, Московский государственный технический университет им. Н.Э. Баумана, Москва, 105005, Российская Федерация, se 57566871100, https://orcid.org/0000-0003-4395-2135, golubeka@bmstu.ru

References

- Wegener K., Bleicher F., Heisel U., Hoffmeisterd H.-W., Möhring H.-C. Noise and vibrations in machine tools. *CIRP Annals*, 2021, vol. 70, no. 2, pp. 611–633. https://doi.org/10.1016/j. cirp.2021.05.010
- Nasibullaeva E.Sh. Simulation of acoustic scattering from a set of sound-permeable spheres in 3D space. *Computational Technologies*, 2022, vol. 27, no. 2, pp. 19–36. (in Russian). https://doi.org/10.25743/ ICT.2022.27.2.003
- Yamamoto K., Masatoshi T., Nakagawa S. Privacy protection for speech signals. *Procedia Social and Behavioral Sciences*, 2010, vol. 2, no. 1, pp. 153–160. https://doi.org/10.1016/j. sbspro.2010.01.029
- Khorev A.A. Protection of Information from Leakage via Technical Channels. Part 1. Technical Channels of Information Leakage. Moscow, Gostehkomissija Rossii Publ., 1998, 320 p. (in Russian)
- 5. Burlakov M.E., Osipov M.N. Acoustic and Vibroacoustic Channels of Information Leakage. Theoretical Foundations and Basic Practical Training. Samara, SSAU Publ., 2021. 96 p. (in Russian)
- 6. Howard C.Q., Cazzolato B.S. Acoustic Analyses Using MATLAB and ANSYS. CRC Press, 2014, 708 p.
- 7. Murashov M.V. Numerical modeling in ANSYS of the pressure field during measures to ensure information protection from leakage through an acoustoelectric channel. *Science and Education of the Bauman MSTU*, 2009, no. 10, pp. 2 (in Russian)
- Murashov M.V., Golubtsova E.S., Kapitonova L.I., Fedorova V.A., Savina T.T. Investigation of the room protection from information leakage through the vibroacoustic channel using the ANSYS program. *Zaŝita informacii. Inside*, 2022, no. 3 (105), pp. 5–9. (in Russian)
- Horev A., Margarian T. Experimental studies of voice messages intelligibility. Proc. of the Conference of Russian Young Researchers in Electrical and Electronic Engineering (EIConRus), 2019, pp. 1538–1543. https://doi.org/10.1109/EIConRus.2019.8657302
- Shchevev Iu.P. Fundamentals of Physical Acoustics. St. Petersburg, Lanbook Publ., 2021, 364 p. (in Russian)
- Gerasimenko V.G., Lavrukhin Iu.N., Tupota V.I. Methods of Protecting Acoustic Speech Information from Leakage via Technical Channels. Moscow, RCIB «Fakel», 2008, 258 p. (in Russian)
- Lukmanova O., Horev A.A., Vorobeyko E., Volkova E.A. Research of the analog and digital noise generators characteristics for protection device. *Proc. of the IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering, (EIConRus)*, 2020. pp. 2093-2096. https://doi.org/10.1109/ EIConRus49466.2020.9039193
- Murashov M.V., Panin S.D. Numerical modelling of contact heat transfer problem with work hardened rough surfaces. *International Journal of Heat and Mass Transfer*, 2015, vol. 90, pp. 72–80. https:// doi.org/10.1016/j.ijheatmasstransfer.2015.06.024
- 14. Murashov M.V. Selection of parameters for computational algorithms while solving the problem of rough bodies contact with ANSYS. *Herald of the Bauman Moscow State Technical University, Series Natural Sciences*, 2016, no. 1 (64), pp. 111–121. (in Russian). https:// doi.org/10.18698/1812-3368-2016-1-111-121
- Korobeynikov A.G., Kalinkina M.E., Tkalich V.L., Pirozhnikova O.I., Grishentcev A.Yu. Modeling of membrane surface state under point impact. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 2020, vol. 20, no. 1, pp. 155–161. (in Russian). https://doi.org/10.17586/2226-1494-2020-20-1-155-161

Authors

Sergey V. Skryl — D.Sc., Full Professor, Bauman Moscow State Technical University, Moscow, 105005, Russian Federation, 57211168577, https://orcid.org/0000-0002-4309-6255, skryl@bmstu.ru

Mikhail V. Murashov — D.Sc., Professor, Associate Professor, Bauman Moscow State Technical University, Moscow, 105005, Russian Federation, SC 55204837500, https://orcid.org/0000-0001-7564-0158, murashov@bmstu.ru

Ekaterina S. Golubtsova — Assistant, Bauman Moscow State Technical University, Moscow, 105005, Russian Federation, SC 57566871100, https://orcid.org/0000-0003-4395-2135, golubeka@bmstu.ru

Грачёва Юлия Викторовна — старший преподаватель, Московский государственный технический университет им. Н.Э. Баумана, Москва, 105005, Российская Федерация, https://orcid.org/0000-0003-0884-5617, uvgracheva@bmstu.ru

Паротькина Мария Алексеевна — лаборант, Московский государственный технический университет им. Н.Э. Баумана, Москва, 105005, Российская Федерация, **SS** 59397043700, https://orcid.org/0000-0001-5418-3111, masha20022811@gmail.com

Статья поступила в редакцию 26.12.2022 Одобрена после рецензирования 05.02.2025 Принята к печати 21.03.2025 Julia V. Gracheva — Senior Lecturer, Bauman Moscow State Technical University, Moscow, 105005, Russian Federation, https://orcid.org/0000-0003-0884-5617, uvgracheva@bmstu.ru

Maria A. Parotkina — Laboratory Assistant, Bauman Moscow State Technical University, Moscow, 105005, Russian Federation, SS 59397043700, https://orcid.org/0000-0001-5418-3111, masha20022811@gmail.com

Received 26.12.2022 Approved after reviewing 05.02.2025 Accepted 21.03.2025



Работа доступна по лицензии Creative Commons «Attribution-NonCommercial» **I/İTMO**

НАУЧНО-ТЕХНИЧЕСКИЙ ВЕСТНИК ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ, МЕХАНИКИ И ОПТИКИ март–апрель 2025 Том 25 № 2 http://ntv.ifmo.ru/ SCIENTIFIC AND TECHNICAL JOURNAL OF INFORMATION TECHNOLOGIES, MECHANICS AND OPTICS March–April 2025 Vol. 25 No 2 http://ntv.ifmo.ru/en/ ISSN 2226-1494 (print) ISSN 2500-0373 (online)

ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ, МЕХАНИКИ И ОПТИКИ

КРАТКИЕ СООБЩЕНИЯ BRIEF PAPERS

doi: 10.17586/2226-1494-2025-25-2-362-365 УДК 004.2

Платформа для киберфизических систем с многоуровневой организацией и смешанной критичностью вычислительных процессов Алексей Евгеньевич Платунов¹, Аркадий Олегович Ключев², Василий Юрьевич Пинкевич^{3⊠}, Владислав Аркадьевич Ключев⁴

1,2,3,4 Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация

1 aeplatunov@itmo.ru, https://orcid.org/0000-0003-3003-3949

² kluchev@itmo.ru, https://orcid.org/0000-0002-3892-8424

³ vupinkevich@itmo.ru^{\Box}, https://orcid.org/0000-0002-8635-5026

⁴ vakluchev@itmo.ru, https://orcid.org/0000-0002-3052-2200

Аннотация

Представлены результаты разработки программно-аппаратной конфигурируемой платформы, предназначенной для создания сложных киберфизических систем, например, таких как многооперационные технологические и лабораторные роботизированные установки, подразумевающие параллельное выполнение прикладных заданий с учетом ограничений реального времени. Разработанная архитектура платформы позволяет сочетать системные и прикладные вычислительные процессы со смешанной по характеру требований реального времени критичностью функций. В рамках создания платформы предложены специализированные модели организации вычислительного процесса с необходимым инструментарием и языками разработки, методы создания компонентов платформы, архитектурные шаблоны проектирования целевых систем на базе платформы. Представлены планы дальнейшего развития и использования платформы.

Ключевые слова

программно-аппаратная платформа, системы со смешанной критичностью, киберфизические системы, встраиваемые системы, системы реального времени

Ссылка для цитирования: Платунов А.Е., Ключев А.О., Пинкевич В.Ю., Ключев В.А. Платформа для киберфизических систем с многоуровневой организацией и смешанной критичностью вычислительных процессов // Научно-технический вестник информационных технологий, механики и оптики. 2025. Т. 25, № 2. С. 362–365. doi: 10.17586/2226-1494-2025-25-2-362-365

A platform for cyber-physical systems with multi-level organization and mixed critical computing processes

Alexey E. Platunov¹, Arkady O. Kluchev², Vasiliy Yu. Pinkevich³⊠, Vladislav A. Kluchev⁴

1,2,3,4 ITMO University, Saint Petersburg, 197101, Russian Federation

¹ aeplatunov@itmo.ru, https://orcid.org/0000-0003-3003-3949

² kluchev@itmo.ru, https://orcid.org/0000-0002-3892-8424

³ vupinkevich@itmo.ru^{\box}, https://orcid.org/0000-0002-8635-5026

⁴ vakluchev@itmo.ru, https://orcid.org/0000-0002-3052-2200

Abstract

The article presents results of the development of a software and hardware configurable platform designed to create complex cyber-physical systems, for example, such as multi-operational technological and laboratory robotic installations involving parallel execution of applied tasks with real-time constraints. The developed architecture of the platform allows combining system and application computing processes with a mixed nature of real-time requirements for critical functions. As part of the creation of the platform, we proposed specialized models of the computing process organization with the necessary tools and development languages, methods for creating platform components, and

© Платунов А.Е., Ключев А.О., Пинкевич В.Ю., Ключев В.А., 2025

architectural design patterns for target systems based on the platform. Plans for further development and use of the platform are presented.

Keywords

hardware and software platform, mixed-critical systems, cyber-physical systems, embedded systems, real-time systems **For citation:** Platunov A.E., Kluchev A.O., Pinkevich V.Yu., Kluchev V.A. A platform for cyber-physical systems with multilevel organization and mixed critical computing processes. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 2025, vol. 25, no. 2, pp. 362–365 (in Russian). doi: 10.17586/2226-1494-2025-25-2-362-365

Для сложных киберфизических систем характерно использование вычислительных компонентов, обеспечивающих выполнение широкого спектра функций, а также высокая степень интеграции вычислительных компонентов с оконечным технологическим оборудованием. На архитектуру и организацию рассматриваемых вычислительных систем существенно влияет такая особенность их функционального наполнения, как смешанная критичность — необходимость выполнять множество вычислительных задач с кардинально различающимися требованиями реального времени выполнения и устойчивости к сбоям [1]. Организация соблюдения всех необходимых временных ограничений в таких условиях при разумном уровне избыточности проектных решений является нетривиальной задачей [2, 3].

Авторы настоящей работы имеют многолетний опыт создания сложных встроенных и киберфизических систем с использованием оригинальной HLDметодологии (High Level Design Methodology) высокоуровневого проектирования встраиваемых систем [4–6]. По сравнению с традиционными маршрутами проектирования, она позволяет сокращать сроки и повышать качество создаваемых систем за счет того, что большее количество проектных решений принимаются на ранних этапах создания системы, работая с представлениями системы на абстрактных уровнях.

В работах [7, 8] сформулированы требования, рассмотрены проблемы и предложены основные архитектурные принципы создания конфигурируемой программно-аппаратной платформы систем со смешанной критичностью, предназначенной для автоматизации сложных многооперационных технологических и лабораторных роботизированных установок. В настоящей работе представлены основные результаты, полученные в процессе микроархитектурного проектирования платформы, а также описаны направления развития платформы.

Под платформой понимается совокупность взаимосвязанных в рамках единой архитектуры спецификаций, проектных решений и компонентов, зафиксированных для повторного использования.

Рассматриваемая платформа внедрена в коммерческом проекте, предполагающем разработку программно-аппаратных комплексов для семейства лабораторных установок, одновременно выполняющих множество экспериментов в автоматическом режиме. Одно из требований к платформе — поддержка возможности ее конфигурирования на нескольких уровнях организации вычислительного процесса (от набора физических устройств до алгоритмов многостадийных процедур управления с ограничениями жесткого реального времени), осуществляемая в несколько этапов разными категориями специалистов.

Разработана и формализована модель организации вычислительного процесса для управления многооперационными установками с несколькими уровнями исполнения прикладных заданий, позволяющая покрывать требования установок заданного класса. В нее входят форматы входных заданий, методы организации их взаимодействия, алгоритмы планирования и выполнения. Поступающие в систему от оператора прикладные задания представляют собой цепочки управляющих процедур для последовательного и параллельного выполнения. Они проходят два основных этапа: создание расписания запуска управляющих процедур и обеспечение их выполнения. Кроме этого, в целях поддержания работоспособности система может автоматически планировать и запускать служебные процедуры.

Предложен метод выделения проектных решений и компонентов целевой системы, которые должны быть отнесены к платформе. Метод позволяет обоснованно формировать структуру платформы и требования к ее компонентам. Метод основан на:

- определении соотношения между универсальностью и минимально достаточным набором возможностей для адаптации элемента платформы к применению в целевой системе специалистом соответствующего профиля;
- ограничении возможностей доступа пользователя к микроархитектурным деталям реализации компонента платформы;
- требованиях длительности жизненного цикла компонентов платформы; типичный желаемый срок
 как минимум 15–20 лет, что превышает обычный срок поддержки сторонних решений и вынуждает разрабатывать собственные решения.

Предложена классификация компонентов платформы по степени готовности к повторному использованию и требования к компонентам разных типов, что позволяет оценить состав работ и требования к квалификации исполнителей на каждом этапе конфигурирования платформы для создания целевой системы. Основные типы компонентов повторного использования: готовые реализации, конфигурируемые компоненты, абстрактные реализации, архитектурные шаблоны. После создания целевой конфигурации некоторого компонента он может пополнить библиотеку как компонент следующего уровня готовности.

Разработан архитектурный шаблон для организации и взаимодействия компонентов системы с разной степенью критичности в части требований реального времени и устойчивости к отказам. Он позволяет изолировать более критичные компоненты (например, процедуры управления оборудованием) от задержек и сбоев в работе менее критичных компонентов (например, обслуживание пользовательского интерфейса), что повышает детерминизм и функциональную безопасность системы. Архитектурный шаблон основан на однонаправленности потока команд и данных, необходимых для выполнения критичных функций: компоненты с более жесткими ограничениями обрабатывают поступающие запросы, но сами не обращаются с запросами к компонентам с более мягкими ограничениями. Также каждый компонент должен иметь безопасный вариант поведения в любом сценарии взаимодействия с внешними компонентами.

На уровне исполнения отдельных стадий управляющих процедур предложен специализированный (Domain Specific Language) мультипарадигменный язык MScript для описания необходимых операций и алгоритмов, а также архитектура среды исполнения программ на этом языке. Среда исполнения имеет некоторые свойства операционной системы: запуск, выполнение и взаимодействие прикладных процессов, организация доступа к целевым структурам данных и устройствам ввода-вывода. Использование системы программирования, разработанной с учетом требований конкретной платформы, позволяет реализовать все необходимые механизмы с применением выразительных программных абстракций, приближенных к потребностям прикладных программистов, одновременно изолируя их от системных механизмов для снижения количества ошибок программирования. Собственная система программирования обеспечивает длительность жизненного цикла созданных для нее программ за счет существенно меньшей зависимости от меняющихся технологий на рынке.

На уровне планирования времени выполнения стадий управляющих процедур предложена модель вычислений для организации параллельного выполнения взаимозависимых процессов. Она позволяет обеспечить выполнение требований реального времени и оптимизировать загрузки исполнительных устройств.

На рисунке показаны основные компоненты предложенной архитектуры программно-аппаратной платформы. В ее составе выделена информационная система, которая строится по классической схеме и включает в себя модуль программного интерфейса, пользовательский web-интерфейс, прикладную информационную систему и прикладную базу данных. Управляющая система включает в себя среду исполнения процедур управления оборудованием, инструментарий, каналы разных типов для взаимодействия управляющих процессов между собой и хранилищами данных. Каналы имеют унифицированный программный интерфейс, называемый «ячейкой». Низкоуровневая часть предназначена для взаимодействия с физическим оборудованием установки.

Особенностью платформы является реализация программных компонентов на базе аппаратного и системного программного обеспечения общего назначения. В текущем варианте платформы используются стандартные IBM PC-совместимые компьютеры, операционная система Astra Linux и Open JDK. Программное обеспечение разработано на языках Kotlin и Java. Выбор стека технологий определяется, с одной стороны, снижением затрат на реализацию и сопровождение, а с другой — достаточно высокими требованиями по производительности компонентов платформы. Отладочный стенд для компонентов платформы и целевых систем был реализован с использованием инструментальной системы JupyterLab, фреймворка Dash, интерпретатора языка Python и приложения-обертки, которое включает программные компоненты платформы. В ходе тестирования получены оценки произво-



Рисунок. Архитектура платформы.

БД — база данных; BM — виртуальные машины; API — модуль программного интерфейса.

Выделены компоненты: информационной системы (белые); управляющей системы (светло-серые); низкоуровневой части (темносерые)

Figure. Platform architecture.

БД — database; BM — virtual machines; API — application program interface module. Components highlighted: information system (white); control system (light gray); low-level components (dark gray)

дительности, подтвердившие выполнение требований реального времени в необходимом диапазоне ограничений.

На базе создаваемой платформы планируется разработка открытого проекта с близким набором функций

Литература

- Burns A., Robert I.D. Mixed Criticality Systems A Review. Department of Computer Science, University of York, York, UK, 2022. 97 p.
- Rouxel B., Altmeyer S., Grelck C. YASMIN: a real-time middleware for COTS heterogeneous platforms // Proc. of the 22nd International Middleware Conference (Middleware '21). 2021. P. 298–309. https:// doi.org/10.1145/3464298.3493402
- Schuster S., Wägemann P., Ulbrich P., Schröder-Preikscha W. Proving real-time capability of generic operating systems by system-aware timing analysis // Proc. of the 25th IEEE Real-Time and Embedded Technology and Applications Symposium (RTAS). 2019. P. 318–330. https://doi.org/10.1109/rtas.2019.00034
- Ключев А.О., Кустарев П.В., Палташев Т.Т., Платунов А.Е. Применение HLD-методологии для проектирования реконфигурируемых встраиваемых систем // Научно-технический вестник информационных технологий, механики и оптики. 2014. № 4(92). С. 74–82.
- Платунов А.Е., Пинкевич В.Ю. Перспективные автоматизированные системы управления наружным освещением на платформе LMTFusion // Информатизация и системы управления в промышленности. 2020. № 1(85). С. 135–138.
- Pinkevich V.Y., Platunov A.E., Penskoi A.V. The approach to design of problem-oriented reconfigurable hardware computational units // Proc. of the Wave Electronics and its Application in Information and Telecommunication Systems (WECONF). 2020. P. 1–6. https://doi. org/10.1109/weconf48837.2020.9131512
- Pinkevich V., Kluchev A., Kluchev V., Platunov A. Designing custom computing platforms for cyber-physical systems // Proc. of the 23rd International Multidisciplinary Scientific GeoConference Surveying Geology and Mining Ecology Management, SGEM. 2023. V. 23. N 2.1. P. 65–72. https://doi.org/10.5593/sgem2023/2.1/s07.09
- Pinkevich V., Kluchev A., Platunov A., Kluchev V. Organization of mixed-critical computing systems for control of multi-operational technological systems // Proc. of the 24th International Multidisciplinary Scientific GeoConference Surveying Geology and Mining Ecology Management, SGEM. 2024. V. 24. N 2.1. P. 65–72. https://doi. org/10.5593/sgem2024/2.1/s07.09
- Platunov A., Kluchev A., Pinkevich V., Kluchev V., Kolchurin M. Training laboratories with online access on the ITMO.cLAB platform // CEUR Workshop Proceedings. 2020. V. 2893.

Авторы

Платунов Алексей Евгеньевич — доктор технических наук, профессор, профессор, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, вс 35318291200, https://orcid.org/0000-0003-3003-3949, aeplatunov@itmo.ru

Ключев Аркадий Олегович — кандидат технических наук, доцент, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, вс 56429748500, https://orcid.org/0000-0002-3892-8424, kluchev@itmo.ru

Пинкевич Василий Юрьевич — кандидат технических наук, доцент, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, **55**6951052900, https://orcid.org/0000-0002-8635-5026, vupinkevich@itmo.ru

Ключев Владислав Аркадьевич — аспирант, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, sc 57216353396, https://orcid.org/0000-0002-3052-2200, vakluchev@itmo.ru

Статья поступила в редакцию 08.11.2024 Одобрена после рецензирования 23.01.2025 Принята к печати 21.03.2025 для расширения возможностей виртуальных лабораторий ITMO.cLAB [9]: управление аппаратно-программной инфраструктурой, полунатурное моделирование встраиваемых систем, их внешнего окружения, объектов управления («умный дом» и другие).

References

- Burns A., Robert I.D. *Mixed Criticality Systems A Review*. Department of Computer Science, University of York, York, UK, 2022, 97 p.
- Rouxel B., Altmeyer S., Grelck C. YASMIN: a real-time middleware for COTS heterogeneous platforms. *Proc. of the 22nd International Middleware Conference (Middleware '21)*, 2021, pp. 298–309. https:// doi.org/10.1145/3464298.3493402
- Schuster S., Wägemann P., Ulbrich P., Schröder-Preikscha W. Proving real-time capability of generic operating systems by system-aware timing analysis. Proc. of the 25th IEEE Real-Time and Embedded Technology and Applications Symposium (RTAS), 2019, pp. 318–330. https://doi.org/10.1109/rtas.2019.00034
- Kluchev A.O., Kustarev P.V., Paltashev T.T., Platunov A.E. HLDmethodology application for reconfigurable embedded systems design. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 2014, no. 4(92), pp. 74–82. (in Russian)
- Platunov A.E., Pinkevich V.Iu. Prospective automated control systems for outdoor lighting on the LMTFusion platform. *ISUP*, 2020, no. 1(85), pp. 135–138. (in Russian)
- Pinkevich V.Y., Platunov A.E., Penskoi A.V. The approach to design of problem-oriented reconfigurable hardware computational units. *Proc.* of the Wave Electronics and its Application in Information and Telecommunication Systems (WECONF), 2020, pp. 1–6. https://doi. org/10.1109/weconf48837.2020.9131512
- Pinkevich V., Kluchev A., Kluchev V., Platunov A. Designing custom computing platforms for cyber-physical systems. *Proc. of the 23rd International Multidisciplinary Scientific GeoConference Surveying Geology and Mining Ecology Management, SGEM*, 2023, vol. 23, no. 2.1, pp. 65–72. https://doi.org/10.5593/sgem2023/2.1/s07.09
- Pinkevich V., Kluchev A., Platunov A., Kluchev V. Organization of mixed-critical computing systems for control of multi-operational technological systems. *Proc. of the 24th International Multidisciplinary Scientific GeoConference Surveying Geology and Mining Ecology Management, SGEM*, 2024, vol. 24, no. 2.1, pp. 65–72. https://doi. org/10.5593/sgem2024/2.1/s07.09
- 9. Platunov A., Kluchev A., Pinkevich V., Kluchev V., Kolchurin M. Training laboratories with online access on the ITMO.cLAB platform. *CEUR Workshop Proceedings*, 2020, vol. 2893.

Authors

Alexey E. Platunov — D.Sc., Full Professor, ITMO University, Saint Petersburg, 197101, Russian Federation, SC 35318291200, https://orcid. org/0000-0003-3003-3949, aeplatunov@itmo.ru

Arkady O. Kluchev — PhD, Associate Professor, ITMO University, Saint Petersburg, 197101, Russian Federation, sc 56429748500, https://orcid. org/0000-0002-3892-8424, kluchev@itmo.ru

Vasiliy Yu. Pinkevich — PhD, Associate Professor, ITMO University, Saint Petersburg, 197101, Russian Federation, se 56951052900, https://orcid.org/0000-0002-8635-5026, vupinkevich@itmo.ru

Vladislav A. Kluchev — PhD Student, ITMO University, Saint Petersburg, 197101, Russian Federation, sc 57216353396, https://orcid. org/0000-0002-3052-2200, vakluchev@itmo.ru

Received 08.11.2024 Approved after reviewing 23.01.2025 Accepted 21.03.2025



Работа доступна по лицензии Creative Commons «Attribution-NonCommercial» **I/İTMO**

НАУЧНО-ТЕХНИЧЕСКИЙ ВЕСТНИК ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ, МЕХАНИКИ И ОПТИКИ март–апрель 2025 Том 25 № 2 http://ntv.ifmo.ru/ SCIENTIFIC AND TECHNICAL JOURNAL OF INFORMATION TECHNOLOGIES, MECHANICS AND OPTICS March–April 2025 Vol. 25 № 2 http://ntv.ifmo.ru/en/ ISSN 2226-1494 (print) ISSN 2500-0373 (online)

ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ, МЕХАНИКИ И ОПТИКИ

doi: 10.17586/2226-1494-2025-25-2-366-372 УДК 539.1.078

Двухосевой манипулятор для измерения профиля интенсивности атомарных и ионных пучков

Александр Анатольевич Васильев¹, Марат Евгеньевич Взнуздаев², Кузьма Александрович Ившин³, Сергей Степанович Киселев⁴, Александр Николаевич Соловьев⁵, Виктор Васильевич Фимушкин⁶

1.2 Петербургский институт ядерной физики им. Б.П. Константинова Национального исследовательского центра «Курчатовский институт», Гатчина, 188300, Российская Федерация

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

^{3,5,6} Объединенный институт ядерных исследований, Дубна, 141980, Российская Федерация

¹ vasillie@gmail.com, https://orcid.org/0009-0002-5219-7215

² vznuzdaev me@pnpi.nrcki.ru, https://orcid.org/0000-0003-4448-0539

³ ivshin@jinr.ru, https://orcid.org/0000-0001-8403-0706

⁴ sskiselev@itmo.ru^{\Box}, https://orcid.org/0009-0001-5464-4915

⁵ solovev@jinr.ru, https://orcid.org/0000-0002-5355-5996

⁶ fimushkin@jinr.ru, https://orcid.org/0009-0003-9816-2808

Аннотация

Предложена конструкция устройства для перемещения чувствительного элемента (компрессионной трубки, цилиндра Фарадея или другого датчика) относительно атомарного или ионного пучка с целью измерения профиля его интенсивности. Устройство представляет собой двухосевой манипулятор с механической регулировкой, построенный по принципу вакуумного ввода движения с деформируемым уплотнительным элементом (металлическим сильфоном). Измерительная система, основанная на манипуляторе, успешно применена для построения профилей интенсивности пучков атомарной и ионной ступеней источника поляризованных ионов (POLarized Ion Source, POLIS). Измерения проводились в рамках эксперимента PolFusion, который направлен на изучение двойного поляризованного ядерного синтеза и осуществляется в Петербургском институте ядерной физики им. Б.П. Константинова Национального исследовательского центра «Курчатовский институт» — ПИЯФ).

Ключевые слова

ввод движения в вакуум, атомарный пучок, ионный пучок, цилиндр Фарадея, компрессионная трубка, спиновая ядерная поляризация, поляризованный ядерный синтез

Ссылка для цитирования: Васильев А.А., Взнуздаев М.Е., Ившин К.А., Киселев С.С., Соловьев А.Н., Фимушкин В.В. Двухосевой манипулятор для измерения профиля интенсивности атомарных и ионных пучков // Научно-технический вестник информационных технологий, механики и оптики. 2025. Т. 25, № 2. С. 366–372. doi: 10.17586/2226-1494-2025-25-2-366-372

A two-axis manipulator for measuring the intensity profile of atomic and ionic beams

Alexander A. Vasilyev¹, Marat E. Vznuzdaev², Kuzma A. Ivshin³, Sergey S. Kiselev^{4⊠}, Aleksandr N. Solovev⁵, Victor V. Fimushkin⁶

^{1,2} Petersburg Nuclear Physics Institute named by B.P. Konstantinov of National Research Centre "Kurchatov Institute", Gatchina, 188300, Russian Federation

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

3,5,6 Joint Institute for Nuclear Research, Dubna, 141980, Russian Federation

© Васильев А.А., Взнуздаев М.Е., Ившин К.А., Киселев С.С., Соловьев А.Н., Фимушкин В.В., 2025

- ¹ vasillie@gmail.com, https://orcid.org/0009-0002-5219-7215
- ² vznuzdaev_me@pnpi.nrcki.ru, https://orcid.org/0000-0003-4448-0539
- ³ ivshin@jinr.ru, https://orcid.org/0000-0001-8403-0706
- ⁴ sskiselev@itmo.ru^{\box}, https://orcid.org/0009-0001-5464-4915
- ⁵ solovev@jinr.ru, https://orcid.org/0000-0002-5355-5996
- ⁶ fimushkin@jinr.ru, https://orcid.org/0009-0003-9816-2808

Abstract

A design for a device to move a sensitive element (compression tube, Faraday cylinder or other sensor) relative to an atomic or ionic beam to measure its intensity profile is described. The device is a two-axis manipulator with mechanical adjustment, built on the principle of vacuum motion input with a deformable sealing element (a metal bellows). The manipulator-based measurement system has successfully applied to measure the intensity profiles of beams from the atomic and ionic units of the POLIS polarized ion source. The measurements were conducted within the PolFusion experiment which is aimed at studying double-polarized nuclear fusion and is being carried out at the National Research Center "Kurchatov Institute" — PNPI.

Keywords

vacuum motion input, spin nuclear polarization, atomic beam, ion beam, Faraday cylinder, compression tube, polarized nuclear fusion

For citation: Vasilyev A.A., Vznuzdaev M.E., Ivshin K.A., Kiselev S.S., Solovev A.N., Fimushkin V.V. A two-axis manipulator for measuring the intensity profile of atomic and ionic beams. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 2025, vol. 25, no. 2, pp. 366–372 (in Russian). doi: 10.17586/2226-1494-2025-25-2-366-372

Задача оптимизации пучка неизбежно возникает во всех ядерно-физических экспериментах, использующих ускорители заряженных частиц или струйные газовые мишени. Такая задача предусматривает наличие эффективного метода измерения характеристик пучка, в том числе профиля его интенсивности. Это необходимо как для коррекции настроек ускорителя при первоначальной проводке пучка, так и в процессе экспериментального сеанса [1, С. 7].

Эксперимент PolFusion [2] проводится в НИЦ «Курчатовский институт» — ПИЯФ и имеет своей целью прямое экспериментальное измерение дифференциального сечения реакции ${}^{2}\text{H}(d, p){}^{3}\text{H}$ и ${}^{2}\text{H}(d, n){}^{3}\text{H}$ е с применением поляризованных пучка и струйной газовой мишени в диапазоне энергий до 100 кэВ. Основная задача этого эксперимента связана с одним из наиболее актуальных направлений современной ядерной физики — изучением спиновой ядерной поляризации и поляризованного ядерного синтеза [3]. Данные, полученные в эксперименте, заполнят пробел в знаниях об основных параметрах реакции поляризованного dd-синтеза и послужат основой для обоснования возможности использования поляризованного топлива в термоядерных реакторах.

В схеме эксперимента присутствуют два пересекающихся под прямым углом поляризованных пучка: ионный, состоящий из ядер дейтерия, производимый линейным источником POLarised Ion Source (POLIS) [4], и атомарный (газовая струя), выходящий из источника Atomic Beam Source (ABS). Для достижения поставленной цели эксперимента атомарный и ионный пучки должны быть проведены до центрального детектора таким образом, чтобы добиться их максимальной интенсивности в области столкновения частиц [5].

Ионный источник POLIS имеет в своем составе как атомарную ступень, в которой создается поток диссоциированного дейтерия, так и ионную, в которой происходит ионизация и ускорение частиц. Для достижения заданных параметров пучка необходима настройка атомарной и ионной ступеней, что требует построения профилей интенсивности данных пучков. Эта задача может быть решена применением чувствительных элементов — датчиков интенсивности атомарной струи и ионного пучка: компрессионной трубки [6] и цилиндра Фарадея [1, С. 13] соответственно. Оба чувствительных элемента производят сигнал, пропорциональный общему потоку частиц, попадающему в их апертуру. Следовательно, для построения профиля интенсивности необходимо иметь возможность перемещать чувствительный элемент относительно исследуемого пучка, производя измерения в разных позициях. Ранее такой подход был успешно применен для построения профиля интенсивности пучка источника атомарной струи ANKE ABS [7].

Для исследования структуры атомарных и ионных пучков в эксперименте PolFusion была создана измерительная система на основе двухосевого манипулятора с ручным приводом. Необходимость создания такой системы собственными силами была обусловлена отсутствием в продаже двухосевых (координатных) манипуляторов, имеющих подходящие массогабаритные характеристики, присоединительные размеры и прочие конструктивные особенности.

Конструкция системы предусматривает размещение на манипуляторе компрессионной трубки или цилиндра Фарадея с установкой всей сборки на одной из ступеней ионного источника POLIS таким образом, чтобы обеспечить перемещение этих чувствительных элементов в плоскости, перпендикулярной оси пучка.

В настоящей работе рассматривается конструкция и принцип действия двухосевого манипулятора, а также первые результаты использования измерительной системы, построенной на его основе, для измерения профилей интенсивности ионного и атомарного пучков.

Вакуумные вводы движения, в целом, можно разделить на три типа: с контактным уплотнением,





1 — вакуумная лампа; 2 — заглушка; 3 — крестообразный вакуумный переходник; 4 — штуцер линии натекателя; 5 — пластина третьего яруса; 6 — опора концевая SH12; 7 — пластина второго яруса; 8 — линейный подшипник в обойме; 9 — цилиндрическая направляющая W12; 10 — пластина первого яруса; 11 — фланец; 12 — маховик; 13 — приводной винт; 14 — поводок; 15 — опорная шпилька; 16 — прижимной хомут; 20 — сильфон; 21 — компрессионная трубка; 22 — кронштейн энкодера; 23 — энкодер



1 — vacuum lamp; 2 — plug; 3 — cross-shaped vacuum adapter; 4 — inlet line fitting; 5 — third-tier plate; 6 — SH12 end support;
7 — second-tier plate; 8 — linear bearing in a housing; 9 — W12 cylindrical guide; 10 — first-tier plate; 11 — flange; 12 — handwheel;
13 — drive screw; 14 — drive link; 15 — support stud; 16 — clamping collar; 20 — bellows; 21 — compression tube; 22 — encoder bracket; 23 — encoder

с бесконтактным уплотнением и с деформируемым уплотнительным элементом [8, С. 7]. Рассматриваемое устройство относится к третьему типу, принцип его действия основан на использовании гибкого тонкостенного вакуумного сильфона.

На рис. 1 показан внешний вид двухосевого манипулятора и схема его устройства с обозначениями отдельных деталей.

Манипулятор смонтирован на стандартном фланце ISO-K250 *11*, который может быть установлен на выходе атомарной или ионной ступени источника POLIS. Внешний диаметр фланца (370 мм) определяет габариты устройства.

Система перемещения манипулятора имеет три яруса (5, 7, 10); каждый ярус представляет собой фрезерованную пластину толщиной 16 мм из сплава АМГ-6 с закрепленными на ней концевыми опорами типа SH12 6 цилиндрических направляющих типа W12 9 и обоймы линейных подшипников типа LM12UU DIN ISO 10285-2019¹ 8. Ярус 10 жестко закреплен на фланце 11 с помощью четырех опорных шпилек M12 15.

Стандартный вакуумный сильфон КF-40 20 номинальной длиной 150 мм герметично соединен с соот-

ветствующими отверстиями на фланце 11 и пластине яруса 5; вакуумные соединения жестко зафиксированы прижимными хомутами 16.

При необходимости сильфон может быть заменен на аналогичный меньшего стандартного размера, KF-25 или KF-16, путем применения адаптеров или замены пластин соответствующих ярусов.

Взаимное перемещение ярусов обеспечивается вручную двумя перпендикулярными приводными винтами M8 × 1 *13*, снабженными для удобства маховиками *12*. Максимальный диапазон хода как по оси *X*, так и по оси *Y* составляет 46 мм (\pm 23 мм).

Чувствительный элемент, компрессионная трубка 21 или цилиндр Фарадея, монтируются в стандартный крестообразный вакуумный адаптер стандарта КF-40 3, который герметично крепится к пластине яруса 5.

Для измерения смещения позиции датчика интенсивности пучка по каждой из осей применяются линейные энкодеры с жидкокристаллическими индикаторами 23. Инструментальная погрешность измерения составляет ± 0,01 мм.

Измерение интенсивности пучков ионов и нейтральных частиц производится различными средствами.

Наиболее простым и точным способом определения интенсивности газовой струи является использование устройства с компрессионной трубкой [9, С. 56]. Принцип действия устройства показан на рис. 2, *а*. Оно состоит из калиброванного измерительного объе-

¹ DIN ISO 10285:2019-03. Rolling bearings — Sleeve type linear ball bearings — Boundary dimensions and tolerances (ISO 10285:2007+Amd. 1:2012). Deutsches Institut für Normung, 2019.



Рис. 2. Принципы действия устройств для измерения интенсивности ионного и атомарного пучков: компрессионная трубка для атомарного пучка (*a*); цилиндр Фарадея для ионного пучка (*b*).

РТ — Pressure Transmitter (датчик давления), D⁰ — атомарный дейтерий, D² — молекулярный дейтерий, е⁻ — электрон, d⁺ — дейтрон

Fig. 2. Operating principles of devices for measuring the intensity of ion and atomic beams: Compression tube for atomic beam (*a*); Faraday cylinder for ion beam (*b*).

PT — Pressure Transmitter, D⁰ — atomic deuterium, D² — molecular deuterium, e⁻ — electron, d⁺ — deuteron

ма и трубки, присоединенной к нему и расположенной навстречу направлению пучка. В рассматриваемом устройстве применена трубка с внутренним диаметром 5 мм и длиной 100 мм; функцию измерительного объема выполняет крестообразный адаптер стандарта KF-40.

Измерение интенсивности атомарного пучка производится по давлению в измерительном объеме. Величина этого давления определяется соотношением между потоком атомов, попадающих внутрь измерительного объема через компрессионную трубку, и потока рекомбинированных молекул, выходящего из этого объема через ту же трубку в свободномолекулярном режиме [10].

В качестве измерительного прибора используется вакуумная лампа l (рис. 1) с горячим катодом. Воспроизводимость датчика такого типа составляет приблизительно 2 %, в то время как погрешность его показаний достигает величины 15–20 %. Соответственно, итоговая точность измерений, полученных с помощью компрессионной трубки, должна быть обеспечена дополнительной калибровкой. Для выполнения калибровки устройство снабжено вентилем тонкой регулировки (натекателем) и калибровочным объемом с прецизионным датчиком давления, которые обеспечивают поступление калибровочного газа в измерительный объем с точно заданной скоростью.

Цилиндр Фарадея — устройство (рис. 2, *b*), предназначенное для измерения интенсивности пучка заряженных частиц в вакууме. Работа цилиндра Фарадея построена на кулонометрическом принципе. Основная деталь устройства — чувствительная ячейка, которая имеет вогнутую форму для обеспечения сбора вторичных электронов, выбитых измеряемым пучком. Сталкиваясь с металлической поверхностью ячейки, ионы пучка теряют свой заряд и передают его металлу. Ячейка приобретает заряд, соответствующий по знаку заряду частиц пучка. Если цилиндр Фарадея включен в электрическую цепь с проводником, находящимся под нулевым потенциалом, между цилиндром и проводником возникает электрический ток, пропорциональный интенсивности пучка. Полученный ток может быть измерен прецизионным микроамперметром.

Двухосевой манипулятор был применен для измерения профилей интенсивности атомарного и ионного пучков дейтерия, производимых основными атомарной и ионной ступенями источника поляризованных ионов POLIS. Для измерений использовались компрессионная трубка в случае атомарного пучка и цилиндр Фарадея в случае ионного. Сборка манипулятора с измерительным устройством устанавливалась на фланцах соответствующих ступеней, таким образом измерялись выходные интенсивности пучков. На рис. 3 приведены результаты измерений. Графики относятся к перемещению измерительных пучков в плоскости пучка, перпендикулярной его оси.

В результате создано устройство двухосевого перемещения (манипулятор) для компрессионной трубки, цилиндра Фарадея или другого устройства (например, чувствительный элемент квадрупольного масс-спектрометра). Манипулятор успешно применен для измерения профилей интенсивности атомарного и ионного пучков экспериментальной установки PolFusion.

Полученные экспериментальные данные использованы для настройки параметров атомарного и ионного источников.



Puc. 3. Результаты измерений интенсивностей пучков атомов (*a*, *b*) и ионов (*c*, *d*) дейтерия, произведенных посредством компрессионной трубки (*a*, *b*) и цилиндра Фарадея (*c*, *d*) при прямолинейном их перемещении через геометрический центр вакуумного тракта атомарной и ионной ступеней источника POLIS: смещения вдоль оси *X* (*a*, *c*) и *Y* (*b*, *d*).

Сплошная линия — аппроксимация функцией нормального распределения Гаусса. Нулевое смещение соответствует геометрическому центру пучкового тракта. На легенде приведены значения параметров аппроксимирующей функции: константа, среднее, стандартное отклонение

Fig. 3. Results of measurements of the deuterium atom intensities (a, b) and ion (c, d) beams produced using a compression tube (a, b) and a Faraday cup (c, d) during their rectilinear movement through the geometric center of the vacuum path of the atomic and ion stages of the POLIS source.

Displacements along the X(a, c) and Y(b, d) axes. The solid line represents the approximation by a Gaussian normal distribution function. Zero displacement corresponds to the geometric center of the beam tract. The legend provides the values of the approximating function parameters: constant, mean, standard deviation

Конструкция манипулятора основана на применении доступных стандартных элементов (сильфон, направляющие, концевые опоры), нестандартные детали предельно технологичны.

Устройство может быть адаптировано для размещения на вакуумных соединениях различного типа путем минимальных переделок. Манипулятор пригоден для применения в широком круге задач, связанных с контролем свойств ионных и атомарных пучков различной природы, а также в других приложениях, требующих прецизионного ввода движения в вакуум.

Литература

- Смалюк В.В. Диагностика пучков заряженных частиц в ускорителях. Новосибирск: Параллель, 2009. 293 с.
- Andreyanov A., Fotyev V., Ivshin K., Kochenda L., Kravchenko P., Kravtsov P., Larionov V., Mikirtychyants S., Rozhdestvensky A., Solovev A., Solovyev I., Trofimov V., Vasilyev A., Vznuzdaev M. Study of ²H(d, p)³H and ²H(d, n)³He nuclear reactions with polarized deuteron beams. PolFusion experiment // Physics of Atomic Nuclei. 2021. V. 84. N 11. P. 1895–1899. https://doi.org/10.1134/ s1063778821100033
- Schieck H.P.G. The status of «polarized fusion» // The European Physical Journal A. 2010. V 44. N 2. P. 321–354. https://doi. org/10.1140/epja/i2010-10964-4
- Kremers H.R., Drentje A.G. Performance of the polarized ion source POLIS used at the AGOR accelerator facility // AIP Conference Proceedings. 1998. V. 421. N 1. P. 507–508. https://doi. org/10.1063/1.54980
- Solovev A., Andreyanov A., Barion L., Ciullo G., Engels R., Fotyev V., Ivshin K., Kochenda L., Kravchenko P., Kravtsov P., Larionov V., Rozhdestvensky A., Sherman S., Solovyev I., Trofimov V., Vasilyev A., Vznuzdaev M. Optimization and first tests of the experimental setup to investigate the double-polarized DDfusion reactions // Journal of Instrumentation. 2020. V. 15. N 8. P. C08003. https://doi.org/10.1088/1748-0221/15/08/c08003
- Терехин С.Н., Васильев А.А., Микиртычьянц М.С., Кравцов П.А., Взнуздаев М.Е. Источник поляризованных атомов дейтерия для эксперимента PolFusion // Научно-технический вестник информационных технологий, механики и оптики. 2012. Т. 79. № 3. С. 109–114.
- Brüggemann R., Emmerich R., Engels R., Kleines H., Koptev V., Kravtsov P., Lemaître S., Ley J., Lorentz B., Lorenz S., Mikirtytchiants M., Nekipelov M., Nelyubin V., Schieck H.P.G., Rathmann F., Sarkadi J., Seyfarth H., Steffens E., Ströher H., Vassiliev A., Zwoll K. The polarized internal gas target for ANKE at COSY-Jülich // AIP Conference Proceedings. 2001. V. 570. P. 830– 834. https://doi.org/10.1063/1.1384173
- Медников М.И. Вводы движения в вакуум. М.: Машиностроение, 1974. 182 с.
- Григорьев К.Ю. Создание поляризованной водороднодейтериевой газовой мишени для эксперимента ANKE на внутреннем пучке накопительного кольца ускорителя COSY: диссертация на соискание ученой степени доктора технических наук / Российская академия наук, С.-Петербургский институт ядерной физики им. Б.П. Константинова. Гатчина, 2007. 119 с.
- Mikirtychyants M., Engels R., Grigoryev K., Kleines H., Kravtsov P., Lorenz S., Nekipelov M., Nelyubin V., Rathmann F., Sarkadi J., Schieck H.P.G., Seyfarth H., Steffens E., Ströher H., Vasilyev A. The polarized H and D atomic beam source for ANKE at COSY-Julich // Nuclear Instruments and Methods in Physics Research, Section A: Accelerators, Spectrometers, Detectors and Associated Equipment. 2013. V. 721. P. 83–98. https://doi.org/10.1016/j.nima.2013.03.043

Авторы

Васильев Александр Анатольевич — кандидат физико-математических наук, заведующий лабораторией, Петербургский институт ядерной физики им. Б.П. Константинова Национального исследовательского центра «Курчатовский институт», Гатчина, 188300, Российская Федерация, SC 57220113105, https://orcid.org/0009-0002-5219-7215, vasillie@gmail.com

Взнуздаев Марат Евгеньевич — кандидат химических наук, заведующий отделом, Петербургский институт ядерной физики им. Б.П. Константинова Национального исследовательского центра «Курчатовский институт», Гатчина, 188300, Российская Федерация; ведущий инженер, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, вс 8347828900, https://orcid.org/0000-0003-4448-0539, vznuzdaev_me@pnpi.nrcki.ru

Ившин Кузьма Александрович — кандидат физико-математических наук, старший научный сотрудник, Объединенный институт ядерных исследований, Дубна, 141980, Российская Федерация, SC 56994011900, https://orcid.org/0000-0001-8403-0706, ivshin@jinr.ru Киселев Сергей Степанович — кандидат технических наук, старший научный сотрудник, доцент, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, https://orcid.org/0009-0001-5464-4915, sskiselev@itmo.ru

References

- Smaliuk V.V. Diagnostics of Charged Particle Beams in Accelerators. Novosibirsk, Parallel' Publ., 2009, 293 p. (in Russian)
- Andreyanov A., Fotyev V., Ivshin K., Kochenda L., Kravchenko P., Kravtsov P., Larionov V., Mikirtychyants S., Rozhdestvensky A., Solovev A., Solovyev I., Trofimov V., Vasilyev A., Vznuzdaev M. Study of ²H(d, p)³H and ²H(d, n)³He nuclear reactions with polarized deuteron beams. PolFusion experiment. *Physics of Atomic Nuclei*, 2021, vol. 84, no. 11, pp. 1895–1899. https://doi.org/10.1134/ s1063778821100033
- Schieck H.P.G. The status of «polarized fusion». *The European Physical Journal A*, 2010, vol. 44, no. 2, pp. 321–354. https://doi. org/10.1140/epja/i2010-10964-4
- Kremers H.R., Drentje A.G. Performance of the polarized ion source POLIS used at the AGOR accelerator facility. *AIP Conference Proceedings*, 1998, vol. 421, no. 1, pp. 507–508. https://doi. org/10.1063/1.54980
- Solovev A., Andreyanov A., Barion L., Ciullo G., Engels R., Fotyev V., Ivshin K., Kochenda L., Kravchenko P., Kravtsov P., Larionov V., Rozhdestvensky A., Sherman S., Solovyev I., Trofimov V., Vasilyev A., Vznuzdaev M. Optimization and first tests of the experimental setup to investigate the double-polarized DDfusion reactions. *Journal of Instrumentation*, 2020, vol. 15, no. 8, pp. C08003. https://doi.org/10.1088/1748-0221/15/08/c08003
- Terekhin S.N., Vasiliev A.A., Mikirtychyants M., Kravtsov P.A., Vznuzdaev M.E. Polarized atomic beam source for PolFusion experiment. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 2012, vol. 79, no. 3, pp. 109– 114. (in Russian)
- Brüggemann R., Emmerich R., Engels R., Kleines H., Koptev V., Kravtsov P., Lemaître S., Ley J., Lorentz B., Lorenz S., Mikirtytchiants M., Nekipelov M., Nelyubin V., Schieck H.P.G., Rathmann F., Sarkadi J., Seyfarth H., Steffens E., Ströher H., Vassiliev A., Zwoll K. The polarized internal gas target for ANKE at COSY-Jülich. *AIP Conference Proceedings*, 2001, vol. 570, pp. 830– 834. https://doi.org/10.1063/1.1384173
- Mednikov M.I. Feed-throughs to Move into the Vacuum. Moscow, Mashinostroenie Publ., 1974, 182 p. (in Russian)
- Grigorev K.Iu. Creation of a polarized hydrogen-deuterium gas target for the ANKE experiment on the inner beam of the COSY accelerator storage ring. Dissertation for the degree of doctor of technical sciences. Gatchina, 2007, 119 p. (in Russian)
- Mikirtychyants M., Engels R., Grigoryev K., Kleines H., Kravtsov P., Lorenz S., Nekipelov M., Nelyubin V., Rathmann F., Sarkadi J., Schieck H.P.G., Seyfarth H., Steffens E., Ströher H., Vasilyev A. The polarized H and D atomic beam source for ANKE at COSY-Julich. *Nuclear Instruments and Methods in Physics Research, Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 2013, vol. 721, pp. 83–98. https://doi.org/10.1016/j.nima.2013.03.043

Authors

Alexander A. Vasilyev — PhD (Physics & Mathematics), Head of Laboratory, Petersburg Nuclear Physics Institute named by B.P. Konstantinov of National Research Centre "Kurchatov Institute", Gatchina, 188300, Russian Federation, Sc 57220113105, https://orcid. org/0009-0002-5219-7215, vasillie@gmail.com

Marat E. Vznuzdaev — PhD (Chemistry), Head of Department, Petersburg Nuclear Physics Institute named by B.P. Konstantinov of National Research Centre "Kurchatov Institute", Gatchina, 188300, Russian Federation; Leading Engineer, ITMO University, Saint Petersburg, 197101, Russian Federation, sc 8347828900, https://orcid. org/0000-0003-4448-0539, vznuzdaev_me@pnpi.nrcki.ru

Kuzma A. Ivshin — PhD (Physics & Mathematics), Senior Researcher, Joint Institute for Nuclear Research, Dubna, 141980, Russian Federation, sc 56994011900, https://orcid.org/0000-0001-8403-0706, ivshin@jinr.ru

Sergey S. Kiselev — PhD, Senior Researcher, Associate Professor, ITMO University, Saint Petersburg, 197101, Russian Federation, https://orcid. org/0009-0001-5464-4915, sskiselev@itmo.ru

Соловьев Александр Николаевич — старший научный сотрудник, Объединенный институт ядерных исследований, Дубна, 141980, Российская Федерация, 📧 57218669665, https://orcid.org/0000-0002-5355-5996, solovev@jim.ru

Фимушкин Виктор Васильевич — старший научный сотрудник, Объединенный институт ядерных исследований, Дубна, 141980, Российская Федерация, вс 6507118905, https://orcid.org/0009-0003-9816-2808, fimushkin@jinr.ru

Статья поступила в редакцию 10.12.2024

Одобрена после рецензирования 27.02.2025

Принята к печати 20.03.2025

Aleksandr N. Solovev — Senior Researcher, Joint Institute for Nuclear Research, Dubna, 141980, Russian Federation, sc 57218669665, https://orcid.org/0000-0002-5355-5996, solovev@jinr.ru

Victor V. Fimushkin — Senior Researcher, Joint Institute for Nuclear Research, Dubna, 141980, Russian Federation, sc 6507118905, https://orcid.org/0009-0003-9816-2808, fimushkin@jinr.ru

Received 10.12.2024 Approved after reviewing 27.02.2025 Accepted 20.03.2025



Работа доступна по лицензии Creative Commons «Attribution-NonCommercial»